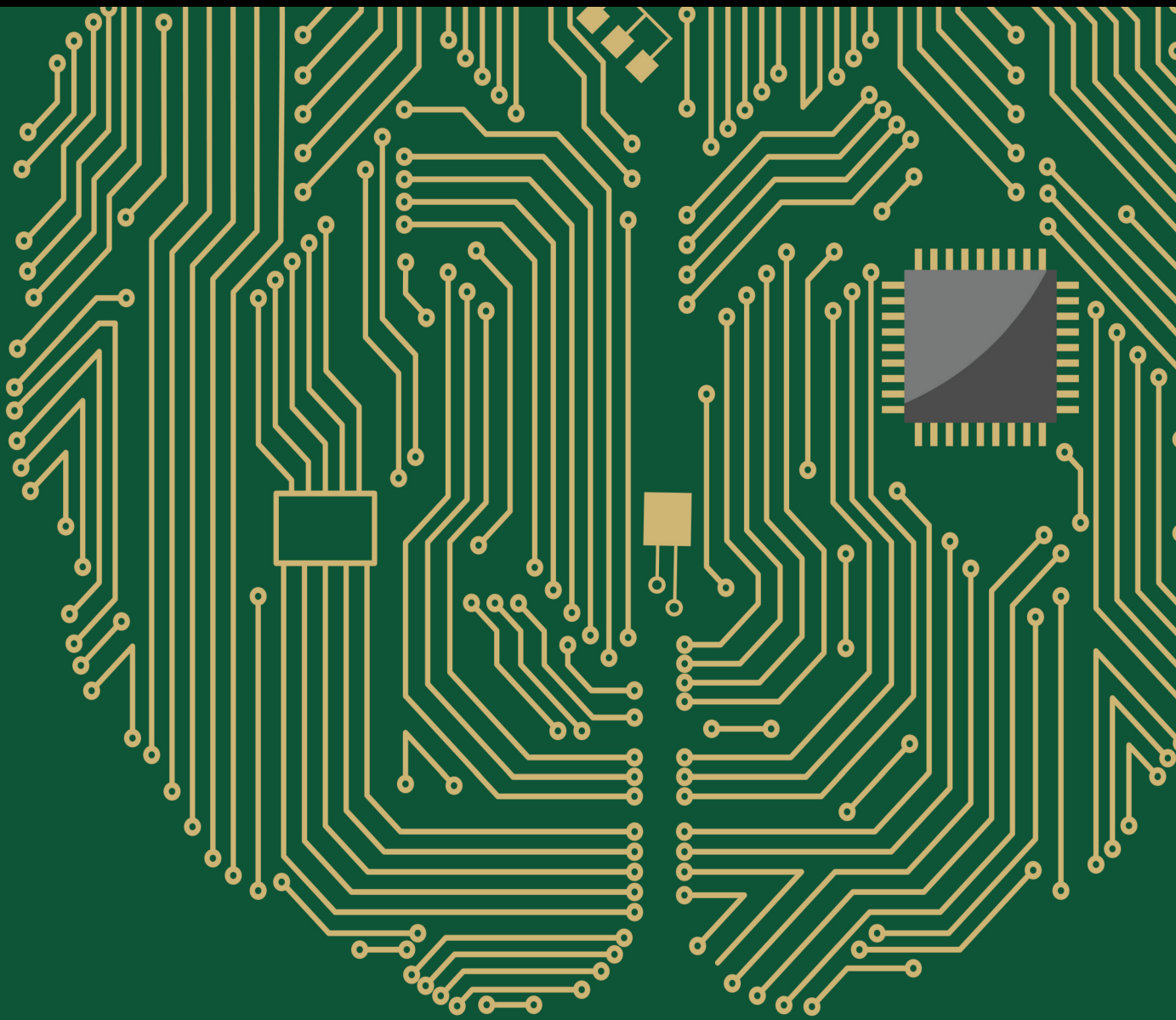# Computational Intelligence in Smart Energy Industries

Lead Guest Editor: Aboul Ella Hassanien
Guest Editors: Ashraf A. Darwish and Vaclav Snasel

Computational Intelligence in Smart Energy Industries

Computational Intelligence in Smart Energy Industries

Lead Guest Editor: Aboul Ella Hassanien
Guest Editors: Ashraf A. Darwish and Vaclav Snasel

Contents

Retraction

Retracted: Experimental Study on Hydraulic Pulsation Features of Intelligent Variable Valve System for Auto Energy Saving

Computational Intelligence and Neuroscience

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Y. Xu, J. Chen, and Z. Liu, "Experimental Study on Hydraulic Pulsation Features of Intelligent Variable Valve System for Auto Energy Saving," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 5679520, 9 pages, 2022.

Retraction

Retracted: Problems and Countermeasures of China's Greenway Economic Development from the Perspective of Computer Internet

Computational Intelligence and Neuroscience

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] S. Zhang, "Problems and Countermeasures of China's Greenway Economic Development from the Perspective of Computer Internet," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 6286833, 11 pages, 2022.

Research Article

Dynamic Prediction of Internet Financial Market Based on Deep Learning

Zixuan Zhang [ID],[1] **Xiaojun Jia** [ID],[2,3] **Shan Chen** [ID],[2] **Menggang Li** [ID],[2,3,4] **and Fang Wang** [ID][2,4]

[1]*Business School, The University of Hong Kong, Pok Fu Lam, Hong Kong*
[2]*National Academy of Economic Security, Beijing Jiaotong University, Beijing 100044, China*
[3]*Beijing Center for Industrial Security and Development Research, Beijing Jiaotong University, Beijing 100044, China*
[4]*Beijing Laboratory of National Economic Security Early-warning Engineering, Beijing Jiaotong University,*
 Beijing 100044, China

Correspondence should be addressed to Xiaojun Jia; xjjia@bjtu.edu.cn

P2P lending is an important part of Internet finance, which is popular among users because of its efficiency, low cost, wide range, and ease of operation. The problem of predicting loan defaults is affected by many factors, such as the linear and nonlinear nature of the data itself and time dependence and multiple external factors, which have not been well captured in the previous work. In this paper, we propose a multiattention mechanism to capture the different effects of various time slices and various external factors on the results, introduce ARIMA and LSTM to capture the linear and nonlinear characteristics of the lending data respectively, and establish a Time Series Multiattention Prediction Model (MAT-ALSTM) based on LSTM and ARIMA. This paper uses the Lending Club dataset from the United States to prove that our model is superior to ANN, SVM, LSTM, GRU, and ARIMA models in the prediction effect of MAE, RMSE, and DA.

1. Introduction

Internet finance [1] refers to a new financial business model in which traditional financial institutions and Internet companies use Internet technology information and communication technology to achieve financing, payment, investment, and information intermediary services. Compared with traditional finance, Internet finance has the advantages of high efficiency, low cost, wide scope, and convenient operation, but it is also accompanied by higher risks. At present, e-banking, online banking, and mobile banking are widely promoted by commercial banks, all belonging to the category of Internet [2] finance. However, the model of Internet finance has, to a certain extent, caused problems such as low threshold in the financial market, enhanced market liquidity, and difficulties in supervision.

In 2017, China has surpassed the United States to become the largest online lending transaction time in the world. The comprehensive transaction volume of online lending platforms

reached 1163.98 billion yuan. According to the World Bank's forecast, the global crowdfunding market will reach \$300 billion in 2025. Behind the rapid development of Internet finance, it bears many risks [3], such as high credit risk, high network security risk, policy stay, and lack of supervision, which has led to chaos in the industry and unscrupulous elements engaging in illegal fund raising and fraud. Since 2020, P2P online lending platforms such as "Taojindi," "Youyi.com," and "Antai Zhuoyue" have been exposed to "runaway" incidents.

As a basic part of the P2P market, P2P online lending platforms accept social loan requests and provide people with investment opportunities. Compared to traditional bank loans, P2P loan origination relies heavily on the borrower's credit as collateral, and P2P loans are funded by thousands of active lenders on the platform. These characteristics mean that P2P loans are unsecured, and the potential for property loss is high.

Therefore, tracking and predicting the lending market dynamics of Internet finance can predict and grasp the

system risk of the platform promptly, and developing appropriate default risk management methods is significant to the system, platform, and users. However, the dynamic tracking and prediction of lending data is difficult due to the high liquidity, uncertainty, and volatility of the lending market. In addition, the number of influential variables in the online environment has increased, and the relationships of various time series have become more complex [4].

The prediction of P2P lending data is a typical multifactor time series prediction problem [5]. This problem differs from traditional forecasting problems in many ways. Specifically, it is manifested in the following points. (1) The data have obvious time sequence and strong time correlation with linear and nonlinear characteristics. (2) Historical data of different periods have different influences on the data to be predicted, manifesting a greater influence between series with closer time and a smaller influence between series with farther time. (3) Various factors have different effects on time series data. For example, if a lender has a higher annual income and a higher credit rating, the loan may be smaller and the default rate on the loan is lower. However, the impact is different, so the prediction process also needs to consider the impact of different factors on the results.

Most traditional time series forecasting algorithms model financial P2P lending data as time variable series. Among them, the ARIMA model is a classic time series forecasting method, which can better reflect the linear characteristics of time series data. However, time series data in finance in general consists of two components: linear and nonlinear [6]. A single ARIMA model can hardly handle the nonlinear variation of financial lending data completely and effectively, and it needs to be combined with other algorithms. In the deep learning algorithm, due to its special network structure, the LSTM model is faster and easier to converge to the optimal solution than the traditional neural network when dealing with time series problems. It is very suitable for processing time series data such as financial loan data.

Based on the above analysis of the P2P data prediction problem, this paper proposes a new MAT-ALSTM (Time Series Multiattention Prediction Model Based on LSTM and ARIMA) model, which can be effectively applied in multifactor finance time series forecasting, and the contributions of this paper can be summarized as follows:

(1) Modeling loan data as multifactor time series data, fully considering the time correlation of P2P loan markets, and the linear and nonlinear dependencies between time series data model the linear and nonlinear properties of the lending data with ARIMA and LSTM, respectively.

(2) This paper proposes a multiattention mechanism to integrate input variables, where according to the different influences of historical data in different periods on the prediction period, a time series attention mechanism is proposed. Aiming at the different influences of varying time sequence features of input variables on the target value, a sequence feature attention mechanism is proposed.

(3) Given the characteristics of P2P loan market data, this paper uses long and short-term memory neural networks to capture the time correlation of loan data and proposes a time series P2P market loan default prediction model MAT-ALSTM based on a multiattention mechanism.

(4) This paper conducts a large number of experiments on a real dataset, which proves the superiority of the MAT-LSTM model proposed in this paper in terms of RMSE, MAE, and DA compared with LSTM, GRU, and other models.

The rest of this paper is described as follows. In Section 2, we review the traditional time series forecasting methods and the application of deep learning to time series problems. In Section 3, we introduce the problem of loan data time series prediction, long short-term memory neural network, and attention mechanism. In Section 4, we propose a time series financial market dynamic prediction model MAT-ALSTM based on a multiattention mechanism and introduce the internal structure of the model in detail. In Section 5, we conduct prediction experiments and analyze the results. Section 6 summarizes the work of this paper.

2. Literature Review

In this section, we review the previous work on modeling and prediction of financial lending data in terms of (1) traditional machine learning-based prediction work and (2) deep learning network-based prediction work.

Some research works are based on traditional statistical methods and machine learning. Traditional statistical models analyze and predict default risk by finding the optimal linear combination of input variables, mainly including logistic regression, discriminant analysis, risk index models, and probabilistic models. For example, Serrano-Cinca and Gutiérrez-Nieto [7] demonstrated the advantages of multivariate discriminant analysis and logistic regression over neural networks in medium-term bankruptcy prediction of "Tunisian" firms. Chen et al. [8] used the partial least squares discriminant analysis (PLS-DA) model to predict the US banking crisis and obtained results close to SVM. Malekipirbazari and Aksakalli [9] used the borrower data of the P2P lending platform "Paipaidai," used the information gain technology to filter the features, and constructed a logistic regression model to evaluate the borrower's credit risk. Ma [10] proposed a random forest-based classification method to predict the status of borrowers, which outperformed the FICO credit score. Niu et al. [11] used the LightGBM algorithm to predict the default rate, which performed better than the XGBoost algorithm and also found that the most important factor affecting the borrower's repayment was the loan details.

There are different kinds of deep learning models: deep multilayer perceptron, RNN, LSTM, CNN, restricted Boltzmann machines, and autoencoder [12]. Deep learning analyzes historical data of online lending and explores the inner logic behind the data, which in turn enables the assessment and prediction of credit risk. For example, Wang

and Ni [13] proposed a deep dense convolutional network for social lending repayment prediction. Many works are based on LSTM and other models for financial time series forecasting. For example, Chen et al. [14] used the long short-term memory (LSTM) model for the first time to analyze P2P. For sequence data, this model has greater potential than traditional time series models. Cao et al [15] used the LSTM model to improve the accuracy of stock return forecasts by 12.9%. Molina et al [16] showed better performance in stock market price prediction by combining the LSTM model and empirical mode decomposition. In the field of e-commerce, LSTM models can effectively predict the future behavior of customers [17]. Yan et al. [18] improved the genetic algorithm by using the long short-term memory network LSTM and constructed the fitness function in a new way. The algorithm can effectively avoid the local optimal trap of the genetic algorithm and has better global search performance, and it promotes the combination of intelligent scheduling with intelligent manufacturing and e-commerce. Compared with BP neural network, traditional RNN, and RNN-improved LSTM deep neural network, LSTM can effectively predict stock market time series and have higher prediction accuracy [19]. In addition, the attention mechanism improves the generalization ability of the network and improves the average prediction accuracy of systemic risk indicators in China's financial market [20]. Meanwhile, hybrid models are used in some papers. Yuan et al. [21] combined CNN and LSTM, CNN is used to make the quantitative stock selection strategy for judging stock trends, and then LSTM is used to make the quantitative timing strategy for improving the profits. Batres-Estrada [22] proposed a novel depth and breadth neural network, which is a combination of RNN and CNN. Hochreiter and Schmidhuber [23] combined DBN and MLP to construct a stock portfolio by predicting the monthly log returns of each stock and selecting only those stocks that are expected to perform better than the median. In addition, new methods have been used in some studies.

From the analysis of existing work, we can draw the following conclusions:

(1) Traditional statistical methods and machine learning methods only consider several features of a single lending data without considering the temporal coupling relationship existing in the data itself. Besides, many machine learning methods cannot capture the nonlinear relationships among time series data well, resulting in poor model coupling.

(2) Some time series deep learning models can handle time series data better; however, many models do not consider the interactions between adjacent one-step or multistep time slices and fail to capture the time dependence well.

(3) Few works consider the importance of different sequence features on the prediction results. Machine learning and deep learning models are able to handle multifeature data; however, different features have different impacts on lending default rates, yet many models fail to capture them well.

Based on the above analysis, this paper proposes a time series prediction model based on a multiattention mechanism, using ARIMA and LSTM algorithms to deal with both linear and nonlinear features of financial loan data to fully consider the time series and the different importance of sequence features to the results.

3. Method

3.1. Financial Lending Time Series Forecasting Problem.

A time series is a sequence of observations recorded in chronological order over a fixed time interval. The prediction problem is to fit a model to predict the future values of the series, taking into account the past values. Let $X = \{X_1, X_2, \ldots, X_T,\}$ be the historical data of a time series, H be the desired prediction range, and the task is to predict the next value of the series $\{X_{T+1}, X_{T+2}, \ldots, X_{T+H},\}$. $\widehat{X} = \{\widehat{X_1}, \widehat{X_2}, \ldots, \widehat{X_T}\}$ is the vector of predicted values, and the goal is to minimize the prediction error.

Time series forecasting refers to using historical data in the past period to predict data in the future, including continuous forecasting (numerical forecasting and range estimation) and discrete forecasting (event forecasting and data classification). The financial lending time series forecasting problem can be summarized as follows.

Given a period τ, each financial time series includes F eigenvalues, and then, all the financial time series features in the τ The financial period can be shown as $X^\tau = (X_1^\tau, X_2^\tau, \ldots, X_N^\tau) \in R^{N \times F}$, where $X_i^\tau \in R$ represents all eigenvalues of the i-th time series, N is the total number of time series, and we set $y_i^\tau \in R$ to represent the characteristics of financial variables in the future τ time period. The financial prediction model can be expressed, using the past financial time series features X^τ to predict the financial variable features $y_i^\tau \in R$ in the future τ time period.

3.2. LSTM.

LSTM (long short-term memory) [24] is an improved RNN (recurrent neural network) model. The basic unit of the LSTM model is a memory block, which includes a memory cell and three control memories. The gate structure of the cell state include the forget gate, the input gate, and the output gate. The forget gate decides to forget the useless historical information from the memory unit state, the input gate decides the influence of the current input data on the memory unit state, and the output gate decides the output information, thus selectively forgetting the unimportant information and reinforcing the important information of the previous sequence.

The structure of the LSTM memory module is shown in Figure 1. Suppose x_t represents the input vector at time t, h_{t-1} represents the output at time t-1, $W_f, W_i, W_C, W_o, U_f, U_i, U_C, U_o$ and W_f, W_i, W_C, W_o, respectively, represent the weight matrix and the bias vector, and the process of the memory module for state update and information output is as follows.

First, the forget gate forgets useless historical information:

$$f_t = \sigma\left(W_f x_t + U_f h_{t-1} + b_f\right). \tag{1}$$
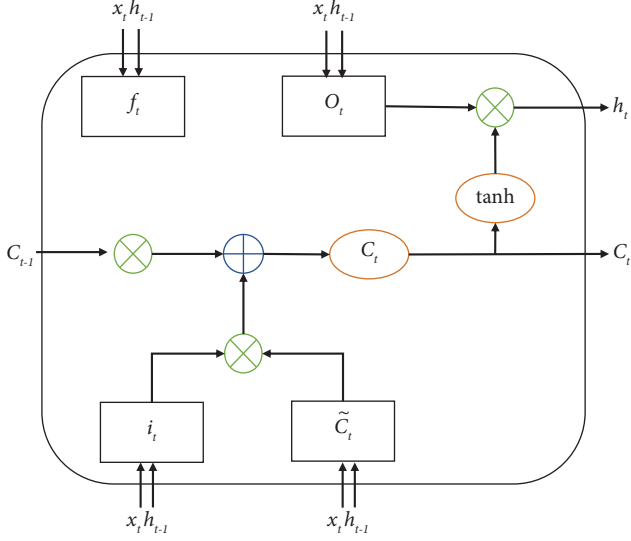
FIGURE 1: Structure diagram of LSTM memory module. f_t: the forget gate, i_t: the input gate, o_t: the output gate, C_t: the state of the memory cell at time t, h_{t-1}: the output at time $t-1$.

Then, the input gate performs state updates based on the input data and historical information:

$$i_t = \sigma\left(W_i x_t + U_i h_{t-1} + b_i\right), \tag{2}$$

$$\tilde{C}_t = \tanh\left(W_C x_t + U_C h_{t-1} + b_C\right), \tag{3}$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t. \tag{4}$$

Finally, the output gate outputs the information at the current moment:

$$o_t = \sigma\left(W_o x_t + U_o h_{t-1} + b_o\right), \tag{5}$$

$$h_t = o_t * \tanh\left(C_t\right), \tag{6}$$

where σ is the logistic sigmoid function, and f_t, i_t, and o_t, respectively, represent the output state of the forget gate, the input gate, and the output gate at time t, and C_t represents the state of the memory cell at time t.

3.3. Attention Mechanism.

The attention mechanism in deep learning [25] is a resource allocation scheme that allocates computational resources to more important tasks while solving the information overload problem when computational power is limited. In neural network learning, the more parameters a model has, the more expressive the model is, and more information is stored in the model, but this will bring about the problem of information overload. Then, by introducing an attention mechanism, focusing on the information that is more critical to the current task among the input data, reducing the attention to other information, or even filtering out irrelevant information, the information

overload problem can be solved and the efficiency and accuracy of task processing can be improved.

The attention mechanism in the time series problem divides the learning model into two parts. First, the encoder composed of a single-layer or multilayer RNN encodes the input sequence according to the time relationship, which is used to learn the predependency and postdependency relationship and the current state representation of the known sequence, and generates the state representation of the current time. During the process of cyclic encoding, we get the state of the last moment and keep it, and this state retains the dynamic information of the input sequence and the current sequence state, which is denoted as vector C. Secondly, a decoder is composed of neural network units of similar structure, which converts the encoding vector E into time series information with a prediction length of τ. $(y_1, y_2, y_3, \ldots, y_{j-1})$ is a vector obtained by common mapping, and the output value at time j is the predicted value at the corresponding time, namely,

$$E = F\left(x_1, x_2, x_3, \ldots, x_\tau\right), \tag{7}$$

$$y_i = G\left(E, y_1, y_2, y_3, \ldots, y_{j-1}\right). \tag{8}$$

In the traditional model, the context vector E used at each moment of decoding is fixed. This structure does not incorporate the principle of different information at different times into the model [19]. Therefore, this paper designs the attention mechanism and combines the structure of the encoder and decoder, and a method of sequential attention mechanism is proposed. For each prediction time j, the encoder obtains a dynamic context vector E_j with different attention information so that the decoding process can pay more attention to the prediction content of the current time and important historical information specifically by

$$E_j = F\left(x_1, x_2, x_3, \ldots, x_\tau, h, h'_{j-1}\right), \tag{9}$$

$$y_i = G\left(E_j, y_1, y_2, y_3, \ldots, y_{j-1}\right), \tag{10}$$

where F represents the process of combining the attention mechanism with the encoder part, h'_{j-1} is the hidden state of the previous step at time j of the decoder, and h is the hidden state set of the encoder.

4. MAT-ALSTM

In this section, on the basis of III, a multiattention mechanism-based time series financial market dynamic prediction model MAT-ALSTM is proposed. The model structure is shown in Figure 2:

The MAT-LSTM model consists of the following modules, including a data preprocessing module, a time series modeling module, and a prediction module. Among them, the data preprocessing module processes the raw financial loan time series data for missing values and outliers
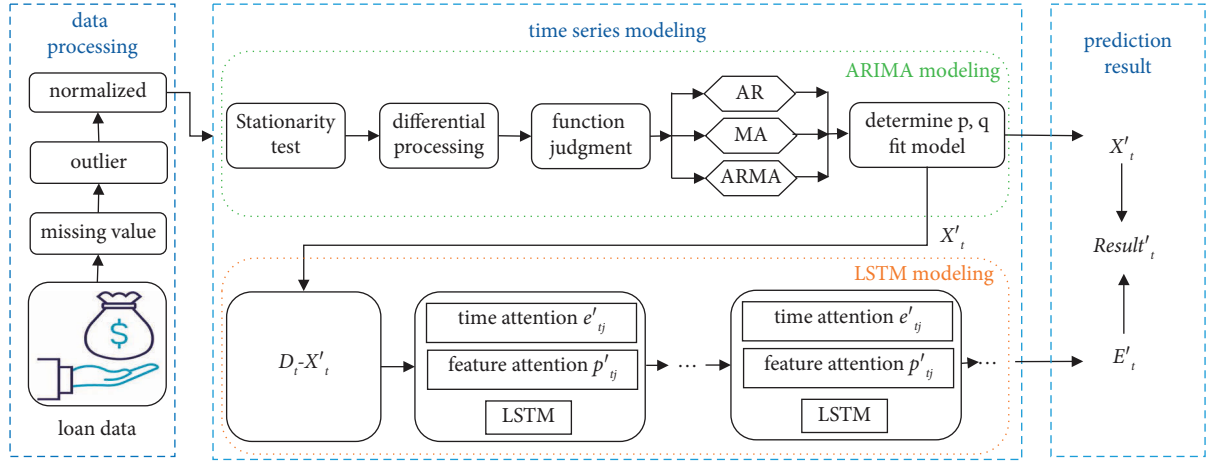
FIGURE 2: MAT-ALSTM model architecture.

and normalizes the data to meet the needs of model training. The time series modeling module consists of two parts: ARIMA modeling and LSTM modeling. Among them, ARIMA modeling mainly extracts the linear part in the loan time series data, and LSTM modeling mainly extracts the nonlinear part in the loan time series data. In the LSTM module, we also introduce a multiattention mechanism to distinguish the different effects of different time series and sequence features on the results. Finally, the prediction module mainly combines the modeling results of ARIMA and the modeling results of LSTM to obtain the final prediction results.

4.1. Data Preprocessing.
This work uses lending data disclosed by the US P2P lending platform Lending Club as the data source. The data range is from June 2008 to December 2015. A total of 2,260,701 data are available in the raw data. The data preprocessing process mainly deals with the data in the original data. For missing values and outliers, first, outlier detection is performed on the data, outliers are regarded as missing values, and the missing values are filled by using Lagrangian interpolation. Secondly, we calculate the average default rate of new loans per month as the dependent variable, and the formula is as follows:

$$\text{Default rate of monthly new loan} = \frac{\text{Defaults of monthly new loans}}{\text{Total monthly new loans}}. \tag{11}$$

In order to make the data more suitable for the model training process, we standardized the data by max-min on columns. The normalization process is as follows:

$$X_{norm} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}. \tag{12}$$

Financial loan data have both linear and nonlinear features. Denote the loan data as D_t, which can be decomposed into a linear part l_t and a nonlinear part n_t, which are recorded as

$$D_t = l_t + n_t. \tag{13}$$

In the preprocessing process, the data are first processed for missing values. For the original dataset of 144 columns of features, there are 60 columns of features containing missing values and 43 columns of features with a percentage of true values above 30%, so we first remove the features with missing values above 30%. The digitized features are filled using the mean or median of other samples, and the non-digitized features are predicted using Sklearn. For outliers, we take the approach of using the mean rather than the outlier samples.

4.2. ARIMA Modeling.
The ARIMA model [26] is a differential autoregressive moving average model, which includes two processes: AR autoregression (p) and MA moving average regression (q). The general form of this model is given as

$$y_t = \mu + \sum_{i=1}^{p} \gamma_i y_{t-i} + \epsilon_t + \sum_{i=1}^{q} \theta_i \epsilon_{t-i}, \tag{14}$$

where p is the lag order of the autoregressive process, q is the lag order of the moving average process, μ is the constant term coefficient, and ϵ_t is the random disturbance term sequence, which is shown as the white noise sequence $\epsilon_t \sim WriteNoise(0, \sigma_t^2)$. The ARIMA model requires that the event sequence must be a stationary sequence during modeling. For nonstationary time series data, the d-order difference should be performed before modeling. Therefore, the complete difference autoregressive moving average process $ARIMA(p, d, q)$ is introduced in the introduction. The model after the lag factor is given as

$$\left(1 - \sum_{i=1}^{p} \varphi_i L^i\right)(1 - L)^d y_t = c + \left(1 + \sum_{i=1}^{q} \theta_i L^i\right)\varepsilon_t, \tag{15}$$

where L represents the lag factor, which is defined as $L^n y_t = y_{t-n}$, and $(1 - L)^d$ is the d-order difference operation.

4.3. Time Series Modeling.

First, we use the ARIMA model to model the data, then fit the training data, and make predictions, and the output prediction result is recorded as X_t', and the residual error can be obtained as

$$R_t = D_t - X_t'. \qquad (16)$$

In view of the different influences of historical data of different time periods on the prediction period and the different influences of different sequence features of input variables on the target value, we propose a multiattention mechanism.

The hidden state at different times has different degrees of attention. In the time dimension, an attention mechanism based on the historical state of the lending market is constructed as follows:

$$e_{tj} = V_a^T \tanh \left(W_a h_{j-1}' + U_a h_t \right), \qquad (17)$$

$$e_{tj}' = \text{softmax}\left(e_{tj}\right), \qquad (18)$$

where h_{j-1}' is the hidden state of the model training in the previous stage; h_t is the loan market state at time t in the model training; V_a^T, W_a, and U_a are all learnable parameter matrices. e_{tj} represents the degree of influence of the state of the loan market at time t in the encoder on the output of the state at the current predicted time j. Finally, the softmax function is used to normalize e_{tj} to obtain the weight factor of the current prediction of the state of the loan market at each historical time. e_{tj}' is the attention value in the time dimension.

Different sequence features have different degrees of attention. In the sequence feature dimension, an attention mechanism based on the time series features of the lending market is constructed as follows:

$$p_{tj} = V_p^T \tanh \left(h_{j-1}' W_p \right) + b_p, \qquad (19)$$

$$p_{tj}' = \text{softmax}\left(p_{tj}\right), \qquad (20)$$

where h_{j-1}' is the hidden state of the model training in the previous stage; V_p^T, W_p, and b_p are all learnable parameter matrices; p_{tj} represents the sequence feature of the lending market at time t in the encoder for the influence degree of the state output of j at the current prediction time; and finally, normalize p_{tj} through the softmax function so as to obtain the weight factor p_{tj}' of different sequence features to the current prediction, that is, the attention value in the sequence feature dimension.

The nonlinear components in the loan data are hidden in the residual sequence R_t. Next, the LSTM model is used to train and predict the residual sequence, and the predicted value E_t' of the residual sequence is obtained.

4.4. Prediction Result.

After preprocessing, the model will be trained by the ARIMA model, and the training result X_t' will be used as the output of the linear feature capture module, followed by using the original loop data D_t to do residual processing with X_t', which is noted as R_t which is the
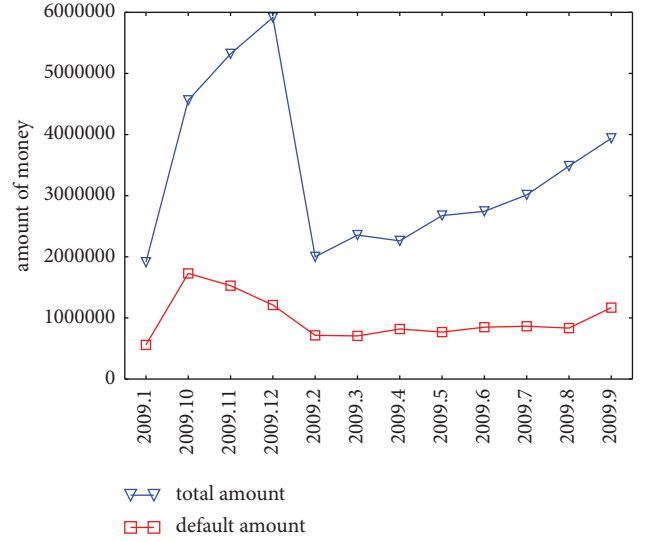
Figure 3: Total borrowings and defaults from January 2009 to December 2009.

nonlinear part of the borrowing data, followed by adding a temporal attention mechanism to the temporal dimension of the data and a multifactor attention mechanism to the feature dimension, followed by using LSTM for the capture of that part of the features. Finally, the results are linearly summed to obtain the final prediction results.

Combining the linear prediction value X_t' of the ARIMA time series model and the nonlinear prediction value E_t' of the LSTM model, the final prediction value $Result_t'$ is

$$Result_t' = X_t' + E_t'. \qquad (21)$$

5. Experiment

5.1. Lending Club Dataset.

In the experiment, the desensitized data published on the official website of Lending Club were selected for the experiment. The time range is from 2008 to 2015. The data include personal information, loan information, and credit information. The loan status variable is the main focus of this paper. It has seven states: fully paid, charge off, default, current, in grace period, late (16–30 days), and late (31–120 days). In this paper, we argue that all statuses should be considered as default except for the status of full payment.

As shown in Figure 3, we count the borrowings and defaults in the raw data from January 2009 to December 2009.

5.2. Model Evaluation.

Models built on time series data must make accurate forecasts in two dimensions of metrics: prediction accuracy and prediction trend accuracy. Therefore, we choose root mean square error (RMSE), mean absolute error (MAE), and orientation accuracy (DA) as the evaluation metrics in this paper. Among them, RMSE and MAE are used to evaluate the prediction accuracy of model

TABLE 1: Experimental set-up of LSTM and GRU models.

Hyperparameters	LSTM	GRU
Hidden layers	5	5
Optimization algorithm	Adam	Adam
Learning rate	0.001	0.001
Epoch	Early stopping	Early stopping
Mini-batch	15	15
Dropout rate	0.005	0.005

TABLE 2: Experimental setup of ARIMA, SVM, and ANN models.

Parameters	ARIMA	Hyperparameters	SVM	Hyperparameters	ANN
AR	1	Kernel	Linear	Nodes	32
Difference	0	Gamma	0.001	Learning rate	0.001
MA	1	Cost	25	Optimizers	Adam
—	—	—	—	Activation	Relu

TABLE 3: Comparison between MAT-ALSTM and the baseline model.

Evaluation	RMSE	MAE	DA
ARIMA	0.113	0.104	0.455
SVM	0.104	0.09	0.547
ANN	0.073	0.059	0.636
LSTM	0.055	0.05	0.636
GRU	0.055	0.047	0.7
MAT-ALSTM	**0.049**	**0.043**	**0.7**

predictions, and DA is used for evaluating the forecast trend accuracy of models' predictions. Formulas of the three statistical indices are as follows:

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \widehat{y}_i)^2}, \quad (22)$$

$$MAE = \frac{1}{m}\sum_{i=1}^{m}\left|(y_i - \widehat{y}_i)\right|, \quad (23)$$

$$DA = \frac{1}{m}\sum_{i=1}^{m}a_i, a_i = \begin{cases} 1, & (y_{i+1} - y_i)(\widehat{y}_{i+1} - y_i) > 0 \\ 0, & (y_{i+1} - y_i)(\widehat{y}_{i+1} - y_i) \leq 0, \end{cases} \quad (24)$$

where $\widehat{y}_i$ is the prediction result of the model, y_i is the true value of the origin sample data, and m is the number of samples.

5.3. Experimental Set-Up.

In the following, we explain the experimental set-up for the models involved in this paper, as shown in Tables 1 and 2.

5.4. Results and Analysis.

We use ARIMA [26], SVM [27], ANN [28], LSTM [24], and GRU [29] models for comparison, and Table 3 shows the performance of this model on the loan default rate from January 2014 to December 2015, and the optimal result has been bolded. Figure 4shows the numerical prediction of the loan default rate of each model from January 2014 to December 2015.

Here, we will briefly introduce the baseline model. ARIMA [26] is an autoregressive integrated moving average model, which analyzes and forecasts time series data with perfection and accuracy. SVM [27] is a class of generalized linear classifier that performs binary classification of data in a supervised learning manner. ANN [28] simulates the neuronal activity with a mathematical model and is an information processing system based on mimicking the structure and function of neural networks in the brain. LSTM [24] and GRU [29] are variants of recurrent neural networks, which can better capture the nonlinear temporal data properties.

From Table 3, we know that the indices of MAT-ALSTM model on the three metrics are 0.049, 0.043, and 0.7, respectively, which achieved optimal results compared to ARIMA and GRU models and achieved a good enhancement effect, as shown in Table 4.

In order to compare the prediction accuracy of each model more conveniently, we draw the following model comparison chart, as shown in Figure 5. It can be observed from the image that the MAT-ALSTM model curve obtains the lowest value in the RMSE and MAE indicator and the maximum value in the DA metrics.

Next, we show the prediction results of the six models at some time points with the real data, as shown in Figure 4, and it can be observed approximately that the prediction results of the MAT-ALSTM model are closer to the real situation compared to the other models.

(a)

(b)
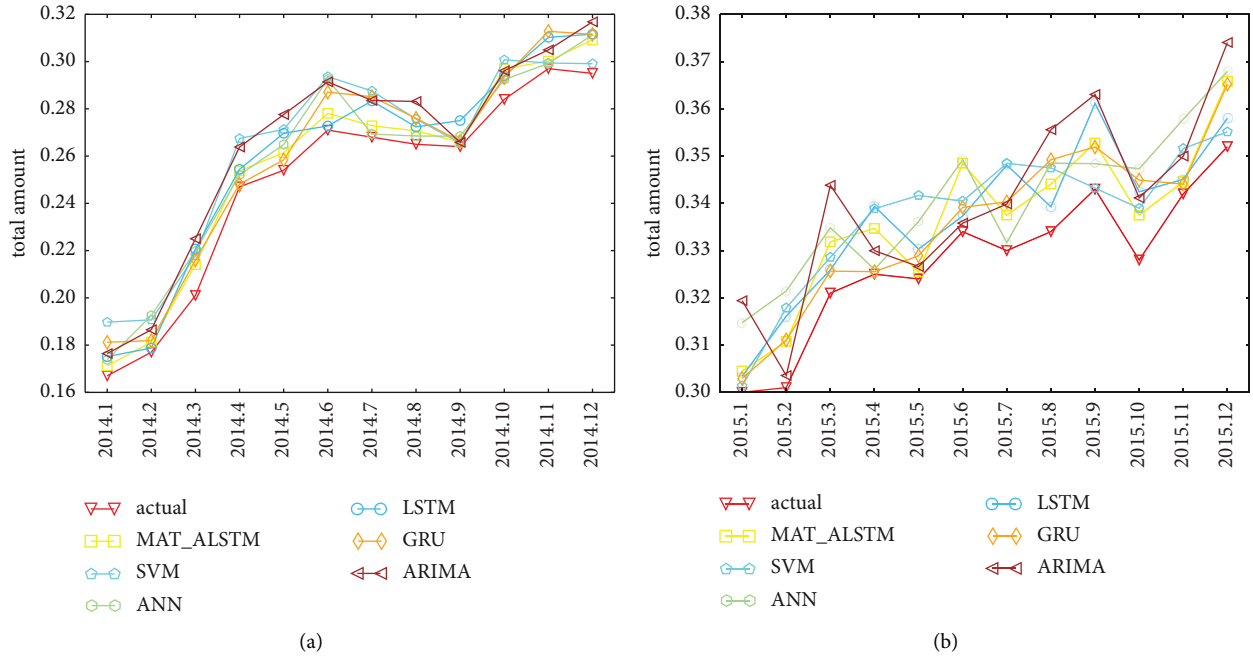
FIGURE 4: Predictions of loan default rates from January 2014 to December 2015. (a) The prediction result from January 2014 to December 2014. (b) The prediction result from January 2015 to December 2015.

TABLE 4: Performance improvement of MAT-ALSTM compared to other models.

	RMSE (%)	MAE	DA (%)
Compare with ARIMA	130.6	141.85	53.8
Compare with SVM	112.2	109.3%	27.9
Compare with ANN	48.9	37.2%	10
Compare with LSTM	12.2	16.2%	10
Compare with GRU	12.2	9.3%	0
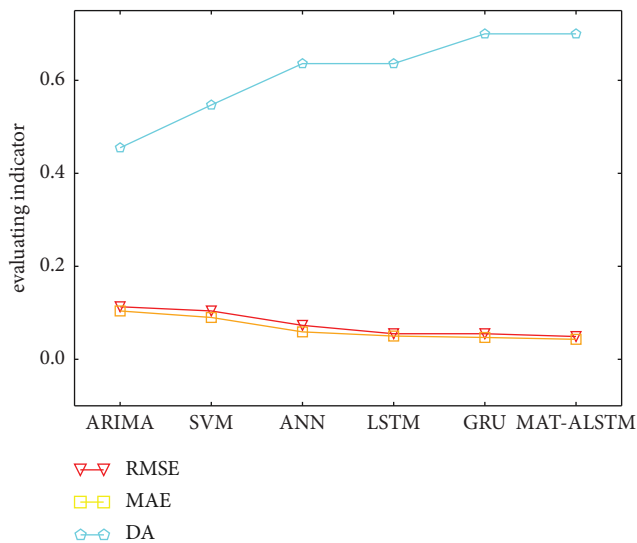
FIGURE 5: Comparison of prediction accuracy of each model.

As shown in Table 1 and Figures 4 and 5, the results of this paper are analyzed as follows.

(1) The MAT-MLSTM model proposed in this paper has better prediction effect than the LSTM and GRU models

(2) The SVM model and ANN model do not take into account the time series characteristics of loan data, so the prediction effect is poor

(3) Compared with the ARIMA model and LSTM model, the MAT-ALSTM model has significantly better prediction results because the MAT-ALSTM model captures the linear and nonlinear characteristics of the loan data, respectively

(4) Although the LSTM model and GRU model consider the time series characteristics of loan data, they do not consider the important difference between time series and series characteristics, while the MAT-ALSTM model fully considers the above characteristics and introduces many attention mechanisms, so the prediction effect is better

6. Conclusion

P2P lending is one of the important components of Internet finance. Accurate lending data prediction is of great significance for applications such as platform construction, user evaluation, and system upgrades. This paper takes the loan default rate of users as the research object, fully considers the linear and nonlinear characteristics of the loan data, and proposes a multiattention model MAT-ALSTM based on LSTM and ARIMA, which can handle the time series in loan data well and fully capture the importance of

different time series and sequence features to the prediction results. The experimental results show that the MAT-ALSTM model in this paper has a better prediction effect than GRU, LSTM, and other models.

Theoretically, our model provides a new perspective for processing and prediction of time series data, i.e., starting from the linear and nonlinear characteristics of time series data itself and using the corresponding mechanisms to capture and merge the results to complete the prediction.

In practice, our model has better prediction results compared with LSTM and GRU models, and the model complexity is not too high, and it can be used as a base model to compose components of deeper models, which are more accurate compared with simple RNN and its variants.

In the future, we will try to combine the model with other domains, such as transportation and retail to verify its usability in other domains. Besides, in this paper, we use a self-attentive mechanism to describe the importance of different time slices and features, and we will also try the attention mechanism design approach. Finally, the interpretability problem of the model has been a bottleneck that hinders the application of deep learning, so we will try to improve the interpretability of the model.

Data Availability

All data used in our study could be accessed by the request to the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] J. Xu, "China's Internet finance: a critical review," *China and World Economy*, vol. 25, no. 4, pp. 78–92, 2017.

[2] P. Xie, C. Zou, and H. Liu, *Internet Finance in China: Introduction and Practical approaches[M]*, Routledge, 2015.

[3] R. Xu, C. Mi, R. Mierzwiak, and R. Meng, "Complex network construction of Internet finance risk," *Physica A: Statistical Mechanics and Its Applications*, vol. 540, Article ID 122930, 2020.

[4] A. Goyal and R. Kaur, "A survey on ensemble model for loan prediction[J]," *International Journal of Engineering Trends and Applications (IJETA)*, vol. 3, no. 1, pp. 32–37, 2016.

[5] S. Siami-Namini, N. Tavakoli, and A. S. Namin, "A comparison of ARIMA and LSTM in forecasting time series[C]," in *Proceedings of the 2018 17th IEEE international conference on machine learning and applications (ICMLA)*, pp. 1394–1401, IEEE, Orlando, FL, USA, 17-20 December 2018.

[6] M. Kouki and A. Elkhaldi, "Toward a predicting model of firm bankruptcy: evidence from the Tunisian context[J]," *Middle Eastern Finance and Economics*, vol. 14, pp. 26–43, 2011.

[7] C. Serrano-Cinca and B. Gutiérrez-Nieto, "Partial least square discriminant analysis for bankruptcy prediction," *Decision Support Systems*, vol. 54, no. 3, pp. 1245–1255, 2013.

[8] Y. Chen, Q. Q. Su, and Q. S. Liu, "Effects of quinestrol on the vocal behavior of mice during courtship interactions," *Physiology and Behavior*, vol. 173, pp. 216–222, 2017.

[9] M. Malekipirbazari and V. Aksakalli, "Risk assessment in social lending via random forests," *Expert Systems with Applications*, vol. 42, no. 10, pp. 4621–4631, 2015.

[10] X. Ma, J. Sha, D. Wang, Y. Yu, Q. Yang, and X. Niu, "Study on a prediction of P2P network loan default based on the machine learning LightGBM and XGboost algorithms according to different high dimensional data cleaning," *Electronic Commerce Research and Applications*, vol. 31, pp. 24–39, 2018.

[11] Z. Niu, G. Zhong, and H. Yu, "A review on the attention mechanism of deep learning," *Neurocomputing*, vol. 452, pp. 48–62, 2021.

[12] J. Y. Kim and S. B. Cho, "Predicting repayment of borrows in peer-to-peer social lending with deep dense convolutional network," *Expert Systems*, vol. 36, no. 4, Article ID e12403, 2019.

[13] Y. Wang and X. S. Ni, "Risk prediction of peer-to-peer lending market by a LSTM model with macroeconomic factor[C]," in *Proceedings of the 2020 ACM Southeast Conference*, pp. 181–187, 2020.

[14] K. Chen, Y. Zhou, and F. Dai, "A LSTM-based method for stock returns prediction: a case study of China stock market [C]," in *Proceedings of the 2015 IEEE international conference on big data (big data)*, pp. 2823-2824, IEEE, 2015.

[15] J. Cao, Z. Li, and J. Li, "Financial time series forecasting model based on CEEMDAN and LSTM," *Physica A: Statistical Mechanics and Its Applications*, vol. 519, pp. 127–139, 2019.

[16] L. E. Molina, S. Bhulai, and V. S. Reader, *Understanding User Behavior in E-Commerce with Long Short-Term Memory (LSTM) and Autoencoders[D]*, Vrije Universiteit Amsterdam, Masters thesis, 2018.

[17] P. J. Wu and D. Yang, "E-commerce workshop scheduling based on deep learning and genetic algorithm," *International Journal of Simulation Modelling*, vol. 20, no. 1, pp. 192–200, 2021.

[18] X. Yan, W. Weihan, and M. Chang, "Research on financial assets transaction prediction model based on LSTM neural network," *Neural Computing & Applications*, vol. 33, no. 1, pp. 257–270, 2021.

[19] Z. S Ouyang, X. T Yang, and Y Lai, "Systemic financial risk early warning of financial market in China using Attention-LSTM model," *The North American Journal of Economics and Finance*, vol. 56, Article ID 101383, 2021.

[20] S. Liu, C. Zhang, and J. Ma, "CNN-LSTM Neural network model for quantitative strategy analysis in stock markets," in *Neural Information Processing*, pp. 198–206, Springer International Publishing, 2017.

[21] Z. Yuan, R. Zhang, and X. Shao, "Deep and wide neural networks on multiple sets of temporal data with correlation," in *Proceedings of the 2018 International Conference on Computing and Data Engineering - ICCDE 2018*, ACM Press, 2018.

[22] B. Batres-Estrada, *Deep Learning for Multivariate Financial Time Series (Master's Thesis)*, KTH, Mathematical Statistics, 2015.

[23] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[24] J. K. Chorowski, D. Bahdanau, and D. Serdyuk, "Attention-based models for speech recognition[J]," *Advances in Neural Information Processing Systems*, vol. 28, 2015.

[25] G. E. P. Box and D. A. Pierce, "Distribution of residual autocorrelations in autoregressive-integrated moving average time series models," *Journal of the American Statistical Association*, vol. 65, no. 332, pp. 1509–1526, 1970.

[26] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[27] S. C. Wang, "Artificial neural network[M]," *Interdisciplinary Computing in Java Programming*, pp. 81–100, Springer, Boston, MA, 2003.

[28] J. Chung, C. Gulcehre, and K. H. Cho, "Empirical evaluation of gated recurrent neural networks on sequence modeling[J]," 2014, http://arXiv.org/abs/1412.3555.

[29] K. Chengeta and E. R. Mabika, "Peer to peer social lending default prediction with convolutional neural networks[C]," in *Proceedings of the 2021 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD)*, pp. 1–10, IEEE, Durban, South Africa, 05-06 August 2021.

Research Article

Rational Uniform Consensus with General Omission Failures

Yansong Zhang,[1,2] Bo Shen,[1,2] and Yingsi Zhao [3]

[1]*School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing, China*
[2]*Key Laboratory of Communication and Information Systems, Beijing Municipal Commission of Education, Beijing, China*
[3]*School of Economics and Management, Beijing Jiaotong University, Beijing, China*

Correspondence should be addressed to Yingsi Zhao; yszhao@bjtu.edu.cn

Generally, system failures, such as crash failures, Byzantine failures, and so on, are considered as common reasons for the inconsistencies of distributed consensus and have been extensively studied. In fact, strategic manipulations by rational agents are not ignored for reaching consensus in a distributed system. In this paper, we extend the game-theoretic analysis of consensus and design an algorithm of rational uniform consensus with general omission failures under the assumption that processes are controlled by rational agents and prefer consensus. Different from crashing one, agent with omission failures may crash or omit to send or receive messages when it should, which leads to difficulty of detecting faulty agents. By combining the possible failures of agents at the both ends of a link, we convert omission failure model into link state model to make faulty detection possible. Through analyzing message passing mechanism in the distributed system with n agents, among which t agents may commit omission failures, we provide the upper bound on message passing time for reaching consensus on a state among nonfaulty agents and message chain mechanism for validating messages. Then, we prove that our rational uniform consensus is a Nash equilibrium when $n > 2t + 1$, and failure patterns and initial preferences are blind (an assumption of randomness). Thus, agents have no motivation to deviate the consensus, which could provide interpretable stability for the algorithm in multiagent systems such as distributed energy systems. Our research strengthens the reliability of consensus with omission failures from the perspective of game theory.

1. Introduction

How to reach consensus despite failures is a fundamental problem in distributed computing. In consensus, each process proposes an initial value and then executes a unique consensus algorithm independently. Eventually all processes need to agree on a same decision chosen from the set of initial values even if there may be some system failures, such as crash failures, omission failures, and Byzantine failures [1]. In the crash model, processes can get into failure state by stopping executing the remaining protocol. In the omission model, processes can get into failure state by omitting to send or receive messages. Also, in the Byzantine model, processes can fail by exhibiting arbitrary behavior. Extensive studies have been conducted on fault-tolerant consensus.

Moreover, two kinds of consensus problems are usually distinguished. One is non-uniform version (usually called "consensus" directly) where no two nonfaulty processes decide differently. The other is uniform version (called "uniform consensus") where no two processes (whether correct or not) decide on different values. We believe that consensus protocols cannot simply replace uniform consensus protocols because the condition of non-uniform consensus is inadequate for many applications [2]. From [3], uniform consensus is harder than consensus because one additional round is needed to decide. Also, uniform consensus is meaningless with Byzantine failures.

Game theory provides interpretable equilibrium by analyzing the game among intelligent players. We argue that its incentive mechanism and punishment mechanism can be effectively applied in distributed systems. Recently, there is an increasing interest on distributed game theory especially in several fields such as peer-to-peer network, biological system, cryptocurrency, and e-commerce, in which

processes are selfish called rational agents (or intelligent agents). Combining distributed computing with algorithmic game theory is an interesting research area enriching the theory of fault-tolerant distributed computing. In this framework, agents may deviate from protocols with any behaviors in order to increase their own profits according to utility functions, which could be regarded as general Artificial Intelligence. In [4], this kind of deviation is referred to as strategic manipulation of distributed protocol. This research is necessary in some practical scenarios, in which each process has selfish incentives. Also, we argue that the fairness of algorithms must be promoted by game theory. Clearly, the goal of distributed computing in the context of game theory is to design algorithms for reaching Nash equilibrium, in which all agents have no incentive to deviate from the algorithms. Perhaps, this framework has been investigated and formalized for the first time in the context of secret sharing and multiparty computation [5–8]. More recently, some fundamental tasks in distributed computing such as leader election and consensus have been studied from the perspective of game theory [9–17].

Following this new line of research, we combine fault-tolerant consensus with rational agents and study the rational uniform consensus problem in synchronous round-based system, where every agent has its own preference on consensus decisions. Thus, an algorithm of rational uniform consensus needs to be constructed. Also, for each agent, its utility is not less with following the consensus algorithm than with deviating from the algorithm. That achieves a Nash equilibrium. It is easy to see that standard consensus algorithms cannot reach equilibrium and they can be easily manipulated by even a single rational agent. Several research studies on rational consensus have been conducted [4, 12–16, 18, 19], but none of them consider the uniform property. Also, most studies on rational consensus only support that there are crash failures or no system failures. We argue that omission failures, which are more subtle and complicated than crashing one, cannot be ignored for reaching uniform consensus. In this paper, we pay attention to a distributed system with n agents, among which t agents may experience omission failures. In this setting, we extend the game-theoretic analysis of consensus. Specifically, our contributions in this paper include the following:

(i) We utilize a punishment mechanism to convert omission failure model into link state model, which makes faulty detection more direct. In the link state model, faulty links never recover whether or not omission failures recover. Therefore, it can provide an idea to simplify the problem of faulty recovery in distributed computing.

(ii) An almost complete mechanism analysis is given for message passing in the distributed system with general omission failures. Then, we provide the upper bound $x + 1$ on message passing time for reaching consensus on a link state. The upper bound determines the round complexity of our algorithm. Next, a message chain mechanism is introduced for validating messages.

(iii) An algorithm of rational uniform consensus with agent omission failures is presented for any $n > 2t + 1$. We give a complete formal proof of correctness of our algorithm. The proof shows that our consensus is a Nash equilibrium.

The rest of the paper is organized as follows. Section 2 introduces the related work. Section 3 describes the model that we are working on. Section 4 presents the algorithm of rational uniform consensus for achieving Nash equilibrium and proves it correct. Section 5 concludes the paper.

2. Related Work

From the view point of modeling methods about agents, the research framework for distributed game theory in the literature may be divided into three categories. In the first category, all of the agents in distributed system are controlled by rational agents preferring consensus and some of them may randomly fail by system failures. Bei et al. [4] studied distributed consensus tolerating both unexpected crash failures and strategic manipulations by rational agents. They considered agents that may fail by crashing. However, the correctness of their protocols needs a strong requirement that it must achieve agreement even if agents deviate. Afek et al. [18] proposed two basic rational building blocks for distributed system and presented several fundamental distributed algorithms by using these building blocks. However, their protocol is not robust against even crash failures. Halpern and Vilaça [12] presented a rational fair consensus with rational agents and crash failures. They used failure pattern to describe the random crash failures of agents. Clementi et al. [13] studied the problem of rational consensus with crash failures in the synchronous gossip communication model. The protocols of Halpern et al. and Clementi et al. do not tolerate omission failure, but we think the consideration to it is necessary. Harel et al. [15] studied the equilibria of consensus resilient to coalitions of $n - 1$ and $n - 2$ agents. They gave a separation between binary and multi-valued consensus. However, they assumed that there are no faulty agents.

The second category is named rational adversary. Groce et al. [19] studied the problem of Byzantine agreement with a rational adversary. Rather than the first model, they assumed that there are two kinds of processes: one is honest and follows the protocol without question; the other is a rational adversary and prefers disagreement. Amoussou-Guenou et al. [14] studied Byzantine fault-tolerant consensus from the game theory point. They modeled processes as rational players or Byzantine players and consensus as a committee coordination game. In [14], the Byzantine players have utility functions and strategies, which can be regarded as rational adversaries similar to [19]. In our opinion, this framework limits the scope of the Byzantine problem.

Finally, the BAR framework (Byzantine, Altruistic, and Rational) was proposed in [20]. In [16], Ranchal-Pedrosa and Gramoli studied the gap between rational agreements that are robust against Byzantine failures and rational agreements that are robust against crash failures. Their

model consists of four different types of players: correct, rational, crash, or Byzantine, which is similar to the BAR model. They consider that rational players prefer to cause a disagreement than to satisfy agreement, which we view as a bit limited because only referring rational players as rational adversaries is one of the questions in the Byzantine model. Moreover, no protocols are proposed in [16].

3. Model

We consider a synchronous system with n agents and each of agent has a unique and commonly known identify in $N = \{1, \ldots, n\}$. Execution time is divided into a sequence of rounds. Each round is identified by the consecutive integer starting from 1. There are three successive phases in a round: a send phase in which each agent sends messages to other agents in system, a receive phase in which each agent receives messages that are sent by other agents in the send phase of the same round, and a computation phase where each agent verifies and updates the value of local variables and executes local computation based on the messages sent and received in that round. We assume that every pair of agents i and j in N is connected by a reliable communication link denoted by $link_{ij}$. For an agent i, all links in the system can be divided into two types: direct link $link_{ij}$ where $j \in N$, and indirect link $link_{kp}$ where neither k nor p is equal to i.

3.1. Failure Model. Here the general omission failures [21], which occur in agents and not in communication links [22], are considered. That is, an agent crashes or experiences either send omissions or receive omissions. Also, send omission means that the agent omits sending messages that it is supposed to send. Receive omission means that the agent omits receiving messages that it should receive. We define that agent omission failures never recover. We argue that our protocol also works even if failures could recover, but proving this seems more complicated. It is easy to see that crash failure can be converted to omission failure because if an agent crashes, it must omit to send and receive messages with all other agents after it has crashed. We assume that there are t agents undergoing general omission failures.

Based on the failure model, we divide the agents in the system into three types:

(i) *Good Agent.* Good agents do not have omission failures.

(ii) *Risk Agent.* Risk agents experience omission failures but we temporarily consider them as correct agents in our protocol.

(iii) *Faulty Agent.* Faulty agents have omission failures with more than t agents.

It is easy to see that t is the sum of the number of risk and faulty agents. We treat good agents and risk agents as nonfaulty agents uniformly. Send omission and receive omission are symmetrical. For example, the cases that i omits to send messages with j and that j omits to receive messages with i have the same view for i and j. Therefore, we may not be able to directly detect the states of some agents

with omission failures. Thus, we call them risk agents and consider them as correct agents. For an agent that has omission failures with more than t agents, it must have omission failures with at least one good agent and then clearly we can know it is a faulty agent.

Due to the symmetry of agent omission failures, we model the agent omission failures as the link state problem by a punishment mechanism. Specifically, in our protocol, if an agent i receives no messages from j in a round, then in the following rounds, i sends no messages to j and does not receive messages from j [23]. Thus, both send omission or receive omission will cause the link interruption. So in a round, we divide each link $link_{ij}$ into three types: correct link, where neither i nor j experiences omission failures in this round, faulty link, where at least one of i and j has omission failures with the other one in the round, and unknown – state link, where the state of $link_{ij}$ in this round is unknown to another agent k. It is easy to see that we can determine the type of an agent by the number of correct direct links of it, which is the fault detection method in our protocol. Similarly, faulty links never recover under this punishment mechanism whether or not omission failures recover.

3.2. Consensus. In the consensus problem, we assume that every agent i has an initial preference v_i in a fixed value set V (we follow the concept of initial preference in [12]). We are interested in uniform consensus in this paper. A protocol solving uniform consensus must satisfy the following properties [3]:

(i) *Termination.* Every correct agent eventually decides.

(ii) *Validity.* If an agent decides v, then v was the initial value of some agent.

(iii) *Uniform Agreement.* No two agents (whether correct or not) decide on different values.

To solve uniform consensus in presence of agent omission failures, we assume that $n > 2t + 1$ and $n \geq 3$.

In uniform consensus, an agent's final decision must be one of the following formalized types:

(i) $\perp$: it means that there is no consensus. $\perp$ is a punishment for inconsistency.

(ii) $\parallel$: it means no decision. Deciding $\parallel$ is not ambiguous with validity, as $\parallel$ cannot be proposed [23]. It does not affect the final consensus outcome.

(iii) $v \in V$: it satisfies the property of validity, which must be the initial preference of some agent.

3.3. Rational Agent. We consider that distributed processes act as rational agents according to the definition in game theory. Each agent i has a utility function u_i. We assume that agents have solution preference [18], and an agent's utility depends only on the consensus value achieved. Thus, for each agent i, there are three values of u_i based on the consensus value achieved: (i) β_0 is i's utility if i's initial preference v_i is decided; (ii) β_1 is i's utility if there is a

consensus value which is not equal to i's initial preference; (iii) β_2 is i's utility if there is no consensus. It is easy to see that $\beta_0 > \beta_1 > \beta_2$, and our results can easily be extended to deal with independent utility function for each agent.

The strategy of an agent i is a local protocol σ_i satisfying the system constrains. i takes actions according to the protocol σ_i in each round. That is, σ_i is a function from the set of messages received to actions. Each agent chooses the protocol in order to maximize its expected utility. Thus, there are n local protocols chosen by every agent, which is called strategy profile $\overrightarrow{\sigma}$ in game theory. The equilibrium is a strategy profile, where each agent cannot increase its utility by deviating if the other agents fix their strategies. For each agent i, if the local protocol σ_i is our consensus algorithm when reaching an equilibrium, then we say that consensus is a Nash equilibrium and the consensus reaching a Nash equilibrium is called rational consensus. Formally, if a strategy profile (or consensus) $\overrightarrow{\sigma}$ is a Nash equilibrium, then for all agents i and all strategies σ_i' for i, it must have $u_i(\sigma_i', \overrightarrow{\sigma}_{-i}) \le u_i(\overrightarrow{\sigma})$.

3.4. Notation Description. The main notations used in following sections are summarized in Table 1.

4. Rational Uniform Consensus with General Omission Failures

4.1. A Rational Uniform Consensus in Synchronous Systems. In order to reach rational uniform consensus that can tolerate omission failures, our protocol adopts a simple idea from an early consensus protocol [23]: An agent does not send or receive any messages to those agents that did not send messages to it previously. Then, we convert the omission failure model which cannot be detected into the link state model which can be detected by agents in each round. However, the presence of rational agents makes protocol more complicated. It requires the protocol to prevent the manipulation of rational agents. Hence, the security of the algorithm needs to be improved from three aspects. The first is interacting with the latest network link states and message sources in each round. The update process of the latest link states within each agent depends on complete message chains, and we can obtain a unified decision round and decision set from message passing mechanism in omission failure environment. The second is using secret sharing for agents' initial preferences [24]. It encrypts the initial preferences so as to prevent an agent knowing the values of other agents in advance. The third is signing each message with a random number and marking faulty links by faulty random numbers [4]. This can improve the difficulty of a rational agent to do evil.

The protocol is described in Algorithm 1. In more detail, we proceed as follows.

Initially, each agent i generates a random number proposal$_i$ which is used for consensus election later (line 2). Then i computes two random 1-degree polynomials q_i and b_i with $q_i(0) = v_i$ and $b_i(0) = $ proposal$_i$, respectively (line 3). They satisfy $(2, n)$ threshold which means that an agent $j \ne i$ can

TABLE 1: Notation description.

Variable	Description
link$_{ij}$	The link between i and j
link$_{ij}^r$	The link between i and j in round r
X-random$_i^r$	Faulty random number of i in round r
random$_i^r$	Message random number for i in round r
NS_i^r	The latest states known to i in round r
HS_i	Historical link states in agent i
v_i	Initial value of i
proposal$_i$	The number of i for computing consensus
t_A	The first tuple of $NS_i^r[l_{kp}]$
t_B	The second tuple of $NS_i^r[l_{kp}]$
State$_i[l_{ij}^r]$	Detection result of agent i on the state of l_{ij} in round r
$C_i^j(m_1, m_2)$	Agent chain (or message propagation path) from agent i to j.
Nf^r	Set of nonfaulty agents in round r
F^r	Set of faulty agents in round r
$F^{\Delta r}$	Set of faulty agents newly detected in round r
x_r	The number of risk agents in round r

restore v_i or proposal$_i$ if it knows more than two pieces of q_i or b_i. Then i initialize set lost$_i$, NS_i^0, HS_i, decision$_i$ and consensus$_i$ (line 4); we discuss these in more detail below. Then i generates the faulty random number X-random$_i^1[k][l_{ij}]$ for each agent $k \ne i$ and each direct link l_{ij} that is the abbreviation of link$_{ij}$ for the link between i and j (lines 5–7; l_{ij}^r represents the link$_{ij}$ in round r). And the message random number random$_i^1$ for round 1 is randomly chosen from $\{0, \ldots, n-1\}$ (line 8). For each link, i generates $n-1$ faulty random numbers and then sends them to other agents, respectively, in round 1. So, we can get that X-random$_i^1$ contains $(n-1)^2$ faulty random numbers in total. Then i puts X-random$_i^1$ and random$_i^1$ into X-RANDOM and RANDOM, respectively (lines 9 and 10), where X-RANDOM is a function storing all faulty random numbers known to i and RANDOM stores all message random numbers. Agents can invoke these two functions to verify random numbers. Specifically, input the id, link, and round to invoke X-RANDOM and input id and round to invoke X-RANDOM.

There are $t+4$ rounds in total and each round has three phases. In phase 1 of round r, $1 \le r \le t+4$, i only sends messages to each agent j who does not belong to lost$_i$ that is a set of agents that have omission failures with i detected by i (line 15). If $1 \le r \le t+3$, i sends random$_i^r$ and NS_i^{r-1} to j. And if $1 \le r \le t+2$, i also sends X-random$_i^r[j]$ which contains $n-1$ random numbers (lines 16 and 17). If $r=1$, i also sends the piece of q_i, $q_i(j)$ and the piece of b_i, $b_i(j)$ to j (line 16). If $r=t+3$, i also sends all the secret shares $q_l(i)$ and $b_l(i)(l \ne j)$ that it has received from other agents (line 18). It is easy to see that the piece $q_l(i)$ and $b_l(i)$ must be in pairs. That is, if i restores v_j, then it can also restore proposal$_j$. Finally, if $r=t+4$, i only sends consensus$_i$ to j (line 19). For each agent i, consensus$_i$ is the set of all consensus values calculated and received by i. Hence, if the algorithm is executed validly, |consensus$_i$| must be equal to 1.

In phase 2 of round r, $1 \le r \le t+4$, i only receives messages from agents that are not in set lost$_i$ (line 22). And if there are no messages received from an agent j, $j \notin $ lost$_i$, i

(1) **function** CONSENSUS(v_i)
(2)   proposal$_i$← a random number
(3) q_i, b_i←random 1 degree polynomials with $q_i(0) = v_i$ and $b_i(0) = $ proposal$_i$ ▷$(2, n)$ threshold
secret sharing
(4)   lost$_i$←∅; NS_i^0←∅; HS_i←∅; decision$_i$←none; consensus$_i$←∅
(5)   **for all** $j ≠ i$ **do**
(6)   **for all** $k ≠ i$ **do**
(7) $X-$random$_i^1 [k][l_{ij}]$←a random bit
(8)   random$_i^1$←random value in $\{0, \ldots, n-1\}$
(9)   puts $X-$random$_i^1$ into X-RANDOM
(10)   puts random$_i^1$ into RANDOM
(11)
(12)   **for** round $r = 1 → t + 4$ **do**
(13) $\{NS^{r-1}\}$←∅; $\{$random$^r\}$←∅
(14)   Phase 1: send phase
(15)   **for all** $j ∉$ lost$_i$ **and** $j ≠ i$ **do**
(16)   **if** $r = 1$ **then** Send $\langle q_i(j), b_i(j), NS_i^{r-1}, X-$random$_i^r[j],$ random$_i^r\rangle$ to j
(17)   **if** $2 ≤ r ≤ t + 2$ **then** Send$\langle NS_i^{r-1}, X-$random$_i^r[j],$ random$_i^r\rangle$ to j
(18)   **if** $r = t + 3$ **then** Send $\langle NS_i^{r-1}, (q_l(i))_{l≠j}, (b_l(i))_{l≠j},$ random$_i^r\rangle$ to j
(19)   **if** $r = t + 4$ **then** Send consensus$_i$ to j
(20)
(21)   Phase 2: receive phase
(22)   **for all** $j ∉$ lost$_i$ **and** $j ≠ i$ **do**
(23)   **if** new message has received from j **then**
(24)   puts $X-$random$_j^r[i]$ into X-RANDOM   ▷round 1 to $t + 2$
(25)   puts random$_j^r$ into RANDOM   ▷round 1 to $t + 3$
(26) $\{NS^{r-1}\}$←$\{NS^{r-1}\} ∪ NS_j^{r-1}$; $\{$random$^r\}$←$\{$random$^r\} ∪$ random$_j^r$ ▷round 1 to $t + 3$
(27)   save $(q_l(j))_{l≠i}$ and $(b_l(j))_{l≠i}$ ▷round $t + 3$
(28)   consensus$_i$←consensus$_i ∪$ consensus$_j$ ▷round $t + 4$
(29)   **else** lost$_i$←lost$_i ∪ j$
(30)   **if** $|$lost$_i| > t$ **then**   ▷Faulty agent
(31)   **Decide** $(\|)$ ▷No decision
(32)
(33)   Phase 3: computation phase
(34)   **if** $r ≤ t + 3$ **then**
(35) NS_i^r←NS_i^{r-1}
(36) NS_i^r, HS_i← VERIFYANDUPDATE$(r, i, NS_i^r, HS_i, \{NS^{r-1}\}, \{$random$^r\})$ ▷Punishment if an inconsistency is detected
(37)   **if** $r ≤ t + 2$ **then**
(38)   random$_i^{r+1}$←random value in $\{0, \ldots, n-1\}$
(39)   **for all** $j ≠ i$ **do**
(40)   **for all** $k ≠ i$ **do**
(41) $X-$random$_i^{r+1}[k][l_{ij}]$←a random bit
(42)   puts $X-$random$_i^{r+1}$ into X-RANDOM
(43)   puts random$_i^{r+1}$ into RANDOM
(44)   **else if** $r = t + 3$ **then**
(45)   LASTUPDATE (HS_i, NS_i^{t+3}) ▷Update the $Msg - X$ of HS_i
(46)   **for** $m = 1 → t + 2$ **do**
(47)   **if** m is the first reliable round in HS_i **then**
(48) m^*←m; break   ▷decision round
(49) D←the set of nonfaulty agents in $HS_i^{m^*}$ ▷decision set
(50)   **for all** $j ∈ D$ **do**
(51)   **if** the number of $q_j(l) < 2$ and $j ≠ i$ **then** break
(52) q_j, b_j← restored by $q_j(l)$ and $b_j(l)$ received
(53) $v_j,$ proposal$_j$←$q_j(0), b_j(0)$
(54)   **if** all values are known in D **then**
(55) C←the set of agents with the second max proposal in D
(56)   **if** $|C| = 1$ **then** consensus$_i$←consensus$_i ∪ v_{C.\text{element}}$
(57)   **else if** $|C| = 0$ **then**
(58) S←the same proposal mod $|D|$ ▷the proposals are the same for all agents in D
(59)   consensus$_i$←consensus$_i ∪ v_j$, where j is the $(S + 1)$ st highest id in D
(60)   **else**   ▷$|C| > 1$

ALGORITHM 1: Continued.

```
(61)                    pr←the second max proposal in D
(62)                    S←pr mod |C|
(63)                    consensus_i←consensus_i ∪ v_j, where j is the (S + 1) st highest id in C
(64)           else                              ▷round t + 4
(65)              if|consensus_i| = 1 then
(66)                 decision_i←consensus_i.element
(67)                 Decide(decision_i)
(68)              else                            ▷Inconsistency
(69)                 Decide(⊥)
```

ALGORITHM 1: Agent i's uniform consensus protocol with initial value v_i $(n > 2t + 1)$.

adds j to $lost_i$ (line 29). Otherwise, i stores the received information (lines 24–28). $\{NS^{r-1}\}$ and $\{random^r\}$ are the sets of all new link states NS_j^{r-1} and message random numbers $random_j^r$, respectively, received by i from each agents $j \notin lost_i$ in round r (line 26). Correspondingly, the elements in $\{NS^{r-1}\}$ and $\{random^r\}$ are one-to-one correspondence. Specially, if $|lost_i| > t$, i knows that it becomes a faulty agent and then i must decide $\parallel$ directly and no longer run in later rounds (line 31). And we say that $\parallel$ means agent i does not decide in the end, which has no influence on the solution.

In phase 3 of round $r \leq t + 3$, i firstly uses NS_i^{r-1} to update NS_i^r which is useful for the update and verification of link states (line 35). Then i invokes the function VER-IFYANDUPDATE to verify and update NS_i^r and HS_i by $\{NS^{r-1}\}$ and $\{random^r\}$ (line 36; see Algorithm 2 for details). NS_i^r is the latest state known to i of all links in the system in round r. HS_i is the historical link state including $t + 3$ rounds in total. If $r \leq t + 2$, i generates the message random number $random_i^{r+1}$ and the faulty random numbers $X\text{-}random_i^{r+1}$ for round $r + 1$, which will be sent to other agents in round $r + 1$, and then puts these random numbers into X-RAN-DOM and RANDOM, respectively (lines 37–43). Then if $r = t + 3$, i last updates HS_i by NS_i^{t+3} (line 45). Specifically, if a link l_{kp} is faulty in round m in NS_i^{t+3}, then change the state of l_{kp} into fault from round m to round $t + 3$ in HS_i. This is the last time modifying HS_i. And following that, i utilizes HS_i to find the decision round m^* from round 1 to round $t + 2$, which is the first reliable round in HS_i (lines 46–48). We follow the concept of clean round in [12]. The number of faulty agents does not increase in clean round and the previous round of clean round is reliable round. Specially, we say that reliable round cannot be round 0, so that the first reliable round is the previous round of the second clean round if the first clean round is round 1. In HS_i^r, if less than $n - t - 1$ links to agent j are correct, then j is a faulty agent. Otherwise, j is a nonfaulty agent. We define that it must remove the explicitly faulty agents when computing the state of j in round r by HS_i. Then i computes the decision set D that is the set of nonfaulty agents in $HS_i^{m^*}$ (line 49). And then i uses all the secret shares it has received in round $t + 3$ to try to restore the initial preference and proposal of each agent $j \in D$ (lines 50–53). If i can reconstruct the values of all agents $\in D$, then i must know all proposals of these agents. Then i computes the consensus proposal (lines 54–63). Firstly, i sorts all the proposals in D

and finds the set C of agents with the second max proposal value (line 55). Then if there is only one agent in C, i puts the initial preference of this agent into $consensus_i$ (line 56). If there are more than one agents in C, that is, more than one agents have the same second max proposal value pr in D (we say that the probability is extremely low), then i uses the pr to mod the agent number of C and gets S (lines 60–62). In this case, i finally puts v_j into its $consensus_i$ where j is the $(S + 1)$ st highest id in C (line 63). Finally, if there is no agent in C, then the proposals must be the same for all agents in D and the second max proposal value does not exist. Thus, i uses the same proposal to mod the agent number of decision set D and gets S (line 58). Similarly, i elects the initial preference of the agent with the $(S + 1)$ st highest id in D (line 59). But if i cannot restore all the values of agents $\in D$, it does nothing and keeps $consensus_i$ as the empty set. Finally, if $r = t + 4$, if $consensus_i$ contains only one value, then i makes a decision (lines 65–67). Otherwise, an inconsistency is detected and i decides $\perp$ (line 69).

The detailed implementation of the verification and update protocol in phase 3 is given in Algorithm 2.

Basically, for each link l_{kp}, $NS_i^r[l_{kp}]$ is a tuple containing two tuples, t_A and t_B. The first tuple t_A represents the state of l_{kp}, which contains three types: $Msg\text{-}R$, $Msg\text{-}X$ and $Msg\text{-}O$, representing correct link, faulty link, and unknown-state link, respectively. The format of type $Msg\text{-}R$ is $(m, k, random_p^m)$, where m is the round of the link state, k is the agent reporting the link state, and $random_p^m$ is the message random number sent by p in round m. It is easy to see that if k reports the state of its direct link l_{kp} is correct ($Msg\text{-}R$) in round m, then it must know $random_p^m$. The format of type $Msg\text{-}X$ is $(X, m, k, X\text{-}random_k^m[l_{kp}])$, where m and k are the same as those in $Msg\text{-}R$, X is an identifier, and $X\text{-}random_k^m[l_{kp}]$ is the sorted set of faulty random numbers on l_{kp} which is generated by k in round m-1. Specifically, the set is sorted by the ids of agents from small to large. The format of type $Msg\text{-}O$ is $\varnothing$ because the state of l_{kp} is unknown for i. The second tuple t_B describes the source of t_A and has the form (j, m), where j is the agent sending the link state t_A to i, and m is the round when j sends it to i. Specially, for direct link, when i first updates the state in round r, t_B is ϕ meaning that the message source is i itself. $HS_i^r[l_{kp}]$ denotes the state of link l_{kp}^r known to i and r could range from 1 to $t + 3$. It contains at most two different tuples because the state of l_{kp} in round r can only be detected and

Require: $r \leftarrow$ round, $i \leftarrow id$, NS_i^r, HS_i, $\{NS^{r-1}\}$, $\{$random$^r\}$
Ensure: (NS_i^r, HS_i) **or** decided
(1)   **function** VERIFYANDUPDATE$(r, i, NS_i^r, HS_i, \{NS^{r-1}\}, \{$random$^r\})$
(2) $T \leftarrow$ IDs$\{NS^{r-1}\}$ ▷ The function IDs returns ids from NS set
(3) $S \leftarrow N - T - \{i\}$
(4)
(5)   Phase 1: update the state of direct links
(6)   **for** $j \in T$ **do**
(7) $NS_i^r[l_{ij}] \leftarrow ((r, i, $random$_j^r), \phi)$ ▷ Message source verification is not
    required if ϕ
(8)   APPENDHS $(HS_i, l_{ij}, (r, i, $random$_j^r))$ ▷ append the state into HS or
    decide $\perp$
(9)   **for** $j \in S$ **do**
(10)   **if** Type $NS_i^r[l_{ij}] = Msg - X$ **then**
(11)   continue
(12)   **else**
(13) $NS_i^r[l_{ij}] \leftarrow ((X, r, i, X-$random$_i^r[l_{ij}]), \phi)$
(14)   APPENDHS $(HS_i, l_{ij}, (X, r, i, X-$random$_i^r[l_{ij}]))$
(15)
(16)   Phase 2: verify message chain
(17)   **for** $j \in T$ **do**
(18)   vERIFYMSGCHAIN(NS_j^{r-1}) ▷ Message chain verification or decide $\perp$
(19)
(20)   Phase 3: verify and update
(21)   **for** $j \in T$ **do**
(22)   **for** $k = 1 \rightarrow n - 1$ **do**
(23)   **for** $p = k + 1 \rightarrow n$ **do**
(24)   recvState $\leftarrow NS_j^{r-1}[l_{kp}]$.state
(25)   localState $\leftarrow NS_i^r[l_{kp}]$.state
(26)   **if** an inconsistency is detected **then**
(27)   **Decide** $(\perp)$ ▷ Punishment
(28)   **if** Type $($recvState$) = Msg - O$ **then**   ▷Case 10
(29)   Continue
(30)   **else if** Type $($localState$) = Msg - O$ **then**   ▷ Case 11
(31) $NS_i^r[l_{kp}] \leftarrow ($recvState$, (j, r))$
(32)   APPENDHS$(HS_i, l_{kp}, $recvState$)$
(33)   **else**
(34) $lr \leftarrow$ localState.round
(35) $li \leftarrow$ localState.id
(36) $rr \leftarrow$ recvState.round
(37) $ri \leftarrow$ recvState.id
(38)   **if** $(k = i$ **or** $p = i)$ **then**   ▷ Direct link
(39)   **if** TYPE $($localState$) = Msg - R$ **and** TYPE $($recvState$) = Msg - R$
    then   ▷ Case 1
(40)   APPENDHS$(HS_i, l_{kp}, $recvState$)$
(41)   **else if** TYPE $($localState$) = Msg - R$ **and** TYPE $($recvState$) = Msg - X$
    then   ▷ Case 2
(42)   **Decide** $(\perp)$
(43)   **else if** TYPE $($localState$) = Msg - X$ **and** TYPE $($recvState$) = Msg - R$
    then   ▷ Case 3
(44)   APPENDHS$(HS_i, l_{kp}, $recvState$)$
(45)   **else if** TYPE $($localState$) = Msg - X$ **and** TYPE $($recvState$) = Msg - X$
    then   ▷ Case 4 and 5
(46)   **if** $li = i$ **then**   ▷ Case 4
(47)   **if** $rr = lr$ **or** $rr = lr + 1$ **then**
(48)   APPENDHS$(HS_i, l_{kp}, $recvState$)$
(49)   **else if** $rr = lr - 1$ **then**
(50) $NS_i^r[l_{kp}] \leftarrow ($recvState$, (j, r))$
(51)   APPENDHS$(HS_i, l_{kp}, $recvState$)$
(52)   **else**   ▷ Indirect link

ALGORITHM 2: Continued.

```
(53)                    if TYPE(localState) = Msg − R and TYPE(recvState) = Msg − R
       then                                        ▷ Case 6
(54)                        if rr ≤ lr then
(55)                            APPENDHS(HS_i, l_{kp}, recvState)
(56)                        else if rr > lr then
(57)                            NS_i^r[l_{kp}] ← (recvState, (j, r))
(58)                            APPENDHS (HS_i, l_{kp}, recvState)
(59)                    else if TYPE(localState) = Msg − R and TYPE(recvState) = Msg − X
       then                                        ▷ Case 7
(60)                        NS_i^r[l_{kp}] ← (recvState, (j, r))
(61)                        APPENDHS (HS_i, l_{kp}, recvState)
(62)                    else if TYPE(localState) = Msg − X and TYPE(recvState) = Msg − R
       then                                        ▷ Case 8
(63)                        APPENDHS (HS_i, l_{kp}, recvState)
(64)                    else if TYPE(localState) = Msg − X and TYPE(recvState) = Msg − X
       then                                        ▷ Case 9
(65)                        if ri ≠ li then
(66)                            if lr ≤ rr then
(67)                                APPENDHS (HS_i, l_{kp}, recvState)
(68)                            else
(69)                                NS_i^r[l_{kp}] ← (recvState, (j, r))
(70)                                APPENDHS (HS_i, l_{kp}, recvState)
(71)    return NS_i^r, HS_i
```

ALGORITHM 2: Agent i verifies and updates link state in round r.

reported by k and p. The form of each tuple is similar to t_A but the round in t_A must be r. And the agents in the two tuples must be different and be k and p, respectively. Specially, if the types of two tuples are Msg-R and Msg-X, then we think the state of l_{kp}^r is faulty. And if Msg-O and Msg-O, then l_{kp}^r is a unknown-state link which is regarded as a correct link when computing decision round and decision set in round $t + 3$.

The pseudocode in Algorithm 2 is explained in detail as follows.

i initially generates T from $\{NS^{r-1}\}$, which is easy to see that i must receive the messages sent by agent $j \in T$ in round r (line 2). And i also computes set S that is equal to $lost_i$ (line 3).

Firstly, in phase 1, i updates the states of direct links in round r. For each agent $j \in T$, i has received the messages from it in round r so that i updates the t_A of $NS_i^r[l_{ij}]$ to Type Msg-R (line 7). And i must be able to obtain the message random number $random_j^r$ from $\{random^r\}$. Then i invokes APPENDHS to append the state $(r, i, random_j^r)$ into $HS_i^r[l_{ij}]$ (line 8). We stipulate APPENDHS must guarantee that the inputting state satisfies the properties of HS which we have discussed above. For example, each link l_{kp} has at most two different tuples in each round, and they come from different agents, k and p. If a state violated the properties of HS, APPENDHS would decide $\bot$ and terminate the protocol early. Then for each agent $j \in S$, it has omission failures detected by i because i does not receive a message from it. If the type of link l_{ij} is already faulty in NS_i^r inherited NS_i^{r-1}, i does nothing because for a link, NS only records the earliest round when the link has failures (lines 10–11). Otherwise, i updates the t_A to Type Msg-X and appends the new state into HS_i (lines 13–14).

Then in phase 2, i utilizes message chain mechanism to verify the correctness of messages $\{NS^{r-1}\}$ received in receive phase (lines 17–18).

Message Chain Mechanism. For each agent j, its message NS_j^m has the following properties.

Suppose S_j^m is the set of agents that disconnected from j in or before round m and T_j^m is the set of agents that are still connected to j in round m. Suppose $X(r)$ represents the $Msg - X$ tuple where the round number is equal to r.

Claim 1. For link l_{kp} in NS_j^m, where $k = j$ and $p \in S_j^m \cup T_j^m$, its state in round m must be known and the number of correct links in $\{l_{jp}\}$ is greater than or equal to $n - t - 1$.

Claim 2. For link l_{kp} in NS_j^m, where $k = j$ and $p \in S_j^m \cup T_j^m$, its state in round $m + 1$ and later must be unknown.

Claim 3. For link l_{kp} in NS_j^m, where $k \in S_j^m$ and $p \in S_j^m$, its state in round $m - 1$ and later must be unknown.

Claim 4. For link l_{kp} in NS_j^m, where $k \in S_j^m$ and $p \in S_j^m$, if the state of $NS_j^m[l_{jk}]$ is $X(m_1)$ and the state of $NS_j^m[l_{jp}]$ is $X(m_2)$, suppose $m \geq m_1 \geq m_2$, then the state of $l_{kp}^{m_1-2}$ must be known.

Claim 5. For link l_{kp} in NS_j^m, where $k \in T_j^m$ and $p \in S_j^m \cup T_j^m$, its state in round $m - 1$ must be known and that in round m and later must be unknown.

Claim 6. For link l_{kp} in NS_j^m, where $k \in T_j^m$ and $p \in S_j^m$, if the state of l_{kp}^{m-1} is $Msg - R$, then the state of link l_{pt} in round

$m-2$, where $t \in S_j^m$, must be known and its state in round $m-1$ and later must be unknown.

Claim 7. For link l_{kp} in NS_j^m, where $k \in T_j^m$ and $p \in S_j^m$, if the state of l_{kp}^{m-1} is equal to $X(m')$, where $m' \le m-1$, then the state of $l_{pt}^{m'-2}$, $t \in S_j^m$, must be known.

Explicitly, we say that the function VERIFYMSGCHAIN is to verify whether a message $NS_j^{r-1} \in \{NS^{r-1}\}$ violates the above claims. If not, then continue to the phase 3. Otherwise, it decides $\perp$ and terminates the protocol early.

Finally, in phase 3, i updates NS_i^r and HS_i by the states in $\{NS^{r-1}\}$. For a link, i compares its state in NS_i^r with the state in $NS_j^{r-1} \in \{NS^{r-1}\}$, so as to implement update according to different cases.

Claim 8. For each link l_{kp} ($k, p \in N$), the agent of t_A must be k or p in all $NS[l_{kp}]$ and $HS[l_{kp}]$.

Case 1. For the direct link l_{ij} of i, if the t_A of $NS_i^r[l_{ij}]$ is (r, i, random) and the t_A of $NS_j^{r-1}[l_{ij}]$ is (r', k, random'), then i only needs to append the new state into HS_i (lines 39–40).

Claim 9. In Case 1, there must be $r' < r$.

Case 2. For the direct link l_{ij} of i, if the t_A of $NS_i^r[l_{ij}]$ is (r, i, random) and the t_A of $NS_j^{r-1}[l_{ij}]$ is $(X, r', k, X\text{-random}_k^{r'}[l_{ij}])$, then i detects an inconsistency and decides $\perp$ (lines 41–42).

Case 3. For the direct link l_{ij} of i, if the t_A of $NS_i^r[l_{ij}]$ is $(X, r', k, X\text{-random}_k^{r'}[l_{ij}])$ and the t_A of $NS_j^{r-1}[l_{ij}]$ is (r'', p, random), then i only needs to append the new state into HS_i (lines 43–44).

Claim 10. In Case 3, there must be $r'' \le r'$.

Case 4. For the direct link l_{ij} of i, if the t_A of $NS_i^r[l_{ij}]$ is $(X, r', i, X\text{-random}_i^{r'}[l_{ij}])$ and the t_A of $NS_j^{r-1}[l_{ij}]$ is $(X, r'', k, X\text{-random}_k^{r'}[l_{ij}])$, then i only needs to append the new state into HS_i when $r'' = r'$ or $r'' = r' + 1$, and i must update NS_i^r and HS_i when $r'' = r'$-1 (lines 45–51). When updating NS_i^r, the t_B of $NS_i^r[l_{ij}]$ must be (j, r) because the new state is obtained from NS_j^{r-1} and updated in round r.

Claim 11. In Case 4, if $k = i$, the t_A of $NS_j^{r-1}[l_{ij}]$ must be the same as the t_A of $NS_i^r[l_{ij}]$, and if $k = j$, it must have $0 \le |r' - r''| \le 1$.

Case 5. For the direct link l_{ij} of i, if the t_A of $NS_i^r[l_{ij}]$ is $(X, r', j, X\text{-random}_j^{r'}[l_{ij}])$ and the t_A of $NS_j^{r-1}[l_{ij}]$ is $(X, r'', k, X\text{-random}_k^{r'}[l_{ij}])$, then i does nothing.

Claim 12. In Case 5, if $k = j$, the t_A of $NS_j^{r-1}[l_{ij}]$ must be the same as the t_A of $NS_i^r[l_{ij}]$, and if $k = i$, it must have $r'' = r' + 1$.

Case 6. For the indirect link l_{kp} of i, if the t_A of $NS_i^r[l_{kp}]$ is (r'', y, random') and the t_A of $NS_j^{r-1}[l_{kp}]$ is (r'', z, random'), then i only needs to append the new state into HS_i when $r'' \le r'$, and i must update NS_i^r and HS_i when $r'' > r'$ (lines 53–58).

Case 7. For the indirect link l_{kp} of i, if the t_A of $NS_i^r[l_{kp}]$ is (r', y, random) and the t_A of $NS_j^{r-1}[l_{kp}]$ is $(X, r'', z, X\text{-random}_z^{r''}[l_{kp}])$, then i needs to update NS_i^r and append the new state into HS_i (lines 59–60).

Claim 13. In Case 7, if $z = y$, it must have $r' < r''$, and if $z \ne y$, it must have $r' \le r''$.

Case 8. For the indirect link l_{kp} of i, if the t_A of $NS_i^r[l_{kp}]$ is $(X, r', y, X\text{-random}_y^{r'}[l_{kp}])$ and the t_A of $NS_j^{r-1}[l_{kp}]$ is (r'', z, random), then i only needs to append the new state into HS_i (lines 62–63).

Claim 14. In Case 8, if $z = y$, it must have $r' > r''$, and if $z \ne y$, it must have $r' \ge r''$.

Case 9. For the indirect link l_{kp} of i, if the t_A of $NS_i^r[l_{kp}]$ is $(X, r', y, X\text{-random}_y^{r'}[l_{kp}])$ and the t_A of $NS_j^{r-1}[l_{kp}]$ is $(X, r'', z, X\text{-random}_z^{r''}[l_{kp}])$, then i does nothing when $z = y$, and i appends the new state into HS_i when $z \ne y$. Specially, i also updates NS_i^r using the new state received if $r' > r''$ (lines 64–70).

Claim 15. In Case 9, if $z = y$, the t_A of $NS_j^{r-1}[l_{kp}]$ must be the same as the t_A of $NS_i^r[l_{kp}]$, and if $z \ne y$, it must have $0 \le |r' - r''| \le 1$.

Case 10. If the t_A of $NS_j^{r-1}[l_{kp}]$ is $\varnothing$, then i does nothing (lines 28–29).

Case 11. For the indirect link l_{kp} of i, if the t_A of $NS_i^r[l_{kp}]$ is $\varnothing$ and the t_A of $NS_j^{r-1}[l_{kp}]$ is (r', z, random) or $(X, r', z, X\text{-random}_z^{r'}[l_{kp}])$, then i needs to update NS_i^r and append the new state into HS_i. (lines 30–32).

Claim 16. If in round r, agent i receives a message in which t_B is (j, m) and $j \in T_i^m$ or $j = i$, then t_A of the message must already be in HS_i when in round r.

Claim 17. If in round r, agent i receives a message in which t_B is $(j, r-1)$ from k, then $\text{Type}(NS_k^{r-1}[l_{kj}^{r-1}])$ must be $Msg\text{-}R$.

In phase 3, for a link l_{kp}, i needs to detect whether there is an inconsistency firstly (line 26). An inconsistency detected in phase 3 may be because

(1) (message format verification). The format of $NS_j^{r-1}[l_{kp}]$ is incorrect;

(2) (message source verification). $NS_j^{r-1}[l_{kp}]$ violates Claim 16 or Claim 17;

(3) (random number verification). If the type of $NS_j^{r-1}[l_{kp}]$ is $Msg\text{-}R$, the message random number

in $NS_j^{r-1}[l_{kp}]$ is different from that in RANDOM, or if Msg-X, the faulty random numbers in X-RANDOM are different from the random numbers at the corresponding indexes of the sorted set in $NS_j^{r-1}[l_{kp}]$;

(4) (round number verification). $NS_j^{r-1}[l_{kp}]$ violates one of the claims from Claim 8 to Claim 15.

If i detects an inconsistency, then it decides $\perp$ (line 27). If not, i updates the states as previously discussed.

4.2. Proof of the Protocol. The proof assumes $n > 2t + 1$. Some variables are defined as follows.

Definition 1. $\text{State}_i[l_{ij}^r]$ denotes the detection result of agent i on the state of direct link l_{ij} in round r. The type of $\text{State}_i[l_{ij}^r]$ must be Msg-R or Msg-X.

Definition 2. $C_i^j(m_1, m_2)$ denotes the agent chain (or we can call it message propagation path) from agent i to j. i detects a direct link state in round m_1 and sends it to agent $k \neq i, j$ in round $m_1 + 1$. Then k also sends the state to another agent in round $m_1 + 2$. Finally, j receives the state in round m_2.

Definition 3. Nf^r denotes the set of nonfaulty agents in round r. F^r denotes the set of faulty agents in round r. $F^{\Delta r}$ denotes the set of faulty agents newly detected in round r. x_r denotes the number of risk agents in round r.

We first prove the upper bound of message passing time and give the round complexity of the algorithm.

Theorem 1. *(message passing mechanism). If $i, j \in Nf^{r+t+1}$, all link states in round r can be reached a consensus between i and j at the latest in round $r + t + 1$.*

Proof. Consider the state of link_{kp} in round r, where $k, p \in N$. Specially, we can consider the messages sent by k and p to be independent of each other and this does not affect the final consensus outcome. For example, $\text{Type}(\text{State}_k[l_{kp}^r]) = Msg$-$X$ and it is received by all nonfaulty agents in round $m_1 (< r + t + 1)$, and $\text{Type}(\text{State}_p[l_{kp}^r]) = Msg$-$R$ and it is received by all nonfaulty agents in round $m_2 (m_1 < m_2 < r + t + 1)$. Even if the detection result of p may no longer be forwarded after round m_1, we still have the correct consensus state in round $r + t + 1$ when we consider two detection results independently. We have following cases:

(i) *Case 1.* k and p are good agents in round r. In round $r + 1$, k and p send their detection results to all good agents. So if $t = 0$, all agents reach a consensus on the state of l_{kp} in round $r + 1$. If $t > 0$, all nonfaulty agents reach a consensus in round $r + 2$. Therefore, all link states of round r among good agents can reach a consensus in round $r + t + 1$.

(ii) *Case 2.* k is a risk agent or faulty agent and p is not equal to k. Generally, since a receive omission can be converted to a send omission, then each risk agent and faulty agent has at most the following 3 choices when sending messages in each round:

(1) It has sending omissions with all other agents.
(2) It does not have sending omissions with at least a good agent.
(3) It has sending omissions with all good agents and no sending omissions with some risk agents or faulty agents.

Hence, k has three choices in round $r + 1$.

(1) *Case 2.1.* k chooses 1. Then all agents do not know $\text{State}_k[l_{kp}^r]$. All nonfaulty agents agree on the "unknown-state."

(2) *Case 2.2.* k chooses 2. Then there must be some good agents knowing $\text{State}_k[l_{kp}^r]$ in round $r + 1$. And all good agents and k know the state in round $r + 2$. If $t = 1$, k is the only risk agent, then there is a consensus on the state in round $r + 2$. But if $t > 1$, all nonfaulty agents receive $\text{State}_k[l_{kp}^r]$ in round $r + 3$ because all good agents must send it to all nonfaulty agents in this round. Thus, the lemma holds.

(3) *Case 2.3.* k chooses 3. So no good agents know $\text{State}_k[l_{kp}^r]$ in round $r + 1$ and k is detected faulty in round $r + 1$. Suppose that there is only a risk agent receiving the state. Since agents are independent of each other, it is easy to scale the number of the agents from one to many. We can also divide this case into two cases.

(a) *Case 2.3.1.* k only sends messages to p in round $r + 1$. It means that $\text{Type}(\text{State}_k[l_{kp}^r]) = Msg$-$R$. If $\text{Type}(\text{State}_p[l_{kp}^r]) = Msg$-$X$, p does not receive messages from k in round $r + 1$. Then the result is the same as that in case 2.1. But if $\text{Type}(\text{State}_p[l_{kp}^r]) = Msg$-$R$, k has no influence on the final result and the consensus result of l_{kp}^r depends on the choice of p. If p also chooses Case 2.3.1, then two states of l_{kp}^r only exist in k and p. The final result is also the same as that in case 2.1.

(b) *Case 2.3.2.* k has no sending omissions with some risk agents or faulty agents other than p. Then in round $r + 2$, the risk agents and faulty agents that have received messages from k also have 3 choices. Take one of the risk agents l as an example. If l chooses 1 or 2 in round $r + 2$, the results are the same as those in case 2.1 and case 2.2 where the lemma holds. And if l chooses 3, no good agents know $\text{State}_k[l_{kp}^r]$ in round $r + 2$. Suppose that when risk and faulty agents choose 3, they must send $\text{State}_k[l_{kp}^r]$ to risk agents or faulty agents other than the source agent of the state because if they only send the state back to the source agent, the final results depend only on the source agent, not on themselves. Then until round $r + t$, if from round $r + 1$ to round $r + t - 1$, all risk agents and faulty agents that have received $\text{State}_k[l_{kp}^r]$ choose 3, then the risk (or faulty) agent z in round $r + t$ must be the last risk (or faulty) agent in system. At this time, z

has only 2 choices: 1 and 2. And it is easy to get that $State_k[l_{kp}^r]$ must be consensus in round $r + t + 1$. But if from round $r + 1$ to round $r + t - 1$, some risk agents or faulty agents that have received $State_k[l_{kp}^r]$ choose 1 or 2, then the results are the same as those in case 2.1 and case 2.2.

In summary, the lemma holds.   □

Corollary 1. *If $i, j \in N f^{r+x_r+1}$, all link states in round r can reach a consensus between i and j in round $r + x_r + 1$.*

Proof. There are $t - x_r$ faulty agents in round r. Since the faulty agents before round r do not send any messages in round $r + 2$, it is equivalent to case 2.1 that k sends messages to these faulty agents in round $r + 1$. Then the total number of risk and faulty agents in case 2.3 can be reduced to x_r. Therefore, in case 2.3.2, if keeping choosing 3, there are no risk or faulty agents anymore up to round $r + x_r$ and all link states in round r can be reached a consensus between i and j in round $r + x_r + 1$.   □

Lemma 1. *(round complexity). The link states HS of the second clean round and all previous rounds can reach a consensus among all nonfaulty agents at the latest in round $t + 3$.*

Proof. By Theorem 1, the smaller the round r, the smaller the supremum of the round in which the link states in round r can reach a consensus. Hence, we directly consider the second clean round. Suppose the second clean round is y. Then there are already at least $y - 2$ faulty agents in round y. That is, $x_y \leq t - y + 2$. By Corollary 1, the link states in round y can reach consensus in round $y + x_y + 1$. Since $y + x_y + 1 \leq t + 3$, the lemma holds.

Then it is proved that the algorithm satisfies all the properties of uniform consensus with general omission failures.   □

Lemma 2. *If i is a nonfaulty agent, then $Type(HS_i^{r \sim t+3}[l_{kp}]) = Msg\text{-}X$ when $Type(HS_i^r[l_{kp}]) = Msg\text{-}X(k, p \in N)$.*

Proof. Link l_{kp} cannot recover after a fault occurs. So if l_{kp} is a faulty link in round r, then its state must also be $Msg\text{-}X$ in subsequent rounds. Moreover, HS also expands all $Msg\text{-}X$ states backwards in LASTUPDATE.   □

Lemma 3. *If i is a nonfaulty agent, then $Type(HS_i^{1 \sim r-1}[l_{kp}]) = Msg\text{-}R$ when $Type(HS_i^r[l_{kp}]) = Msg\text{-}R(k, p \in N)$.*

Proof. Since $Type(HS_i^r[l_{kp}]) \neq Msg\text{-}O$, there must be an agent in k or p (supposing p) that has reported $State_p[l_{kp}^r]$ in round $r + 1$, and finally the state has been transmitted to i. We suppose that i receives the state in round r'. Then we have $C_p^i(r, r')$. Since link omission is irreversible, $C_p^i(r, r')$ must be nonfaulty from round 1 to round $r - 1$.

Hence, $State_p[l_{kp}^{1 \sim r-1}]$ must eventually be received by i. That means $Type(HS_i^{1 \sim r-1}[l_{kp}]) \neq Msg\text{-}O$. Combining Lemma 2, it is easy to get $Type(HS_i^{1 \sim r-1}[l_{kp}]) \neq Msg\text{-}X$. Thus, the lemma holds.   □

Lemma 4. *If i is a nonfaulty agent, then $Type(HS_i^{r \sim t+3}[l_{kp}]) = Msg\text{-}O$ when $Type(HS_i^r[l_{kp}]) = Msg\text{-}O(k, p \in N)$.*

Proof. For a contradiction, let $Type(HS_i^{m_1}[l_{kp}]) \neq Msg\text{-}O(m_1 > r)$ when $Type(HS_i^r[l_{kp}]) = Msg\text{-}O$. Suppose that i receives the state of $link_{kp}^{m_1}$ in round $m_2 (m_2 > m_1)$. Let us suppose i receives it from k. Then we must have $C_k^i(m_1, m_2)$. Since link omission is irreversible, the message propagation path is also correct for round r, so that i must receive $State_k[l_{kp}^r]$ and then $Type(HS_i^r[l_{kp}]) \neq Msg\text{-}O$. Therefore, we have a contradiction here and the lemma holds.   □

Lemma 5. *A nonfaulty agent must have correct links with at least $n - t - 1$ agents other than itself in a round.*

Proof. We can know that for a nonfaulty agent i, $|lost_i|$ must be less than or equal to t. So it is easy to get that i have correct links with at least $n - t - 1$ agents.   □

Lemma 6. *A nonfaulty agent must have correct links with at least 2 good agents other than itself in a round.*

Proof. We analyze the nonfaulty agent i from two aspects of good agent and risk agent.

   (i) *Case 1.* i is a good agent. Then i must have correct links with all other $n - t - 1$ good agents. Since $n > 2t + 1$ and $n \geq 3$, there must be $n - t - 1 \geq 2$.

   (ii) *Case 2.* i is a risk agent. Suppose that i is nonfaulty in round r and there are f faulty agents in this round. Because it must remove faulty agents when computing the state of i, combining Lemma 5, the risk agent i needs to have correct links with at least $(n - t - 1) - (t - f - 1) = n - 2t + f$ good agents in round r. Since $n - 2t > 1$, $n - 2t + f > f + 1$. Then $n - 2t + f \geq 2$ always holds.

Therefore, the lemma holds.   □

Remark 1. For a faulty agent f in round r, since it has faulty links with more than t agents in round r, then it does not send messages to any agents after at most 2 rounds. Hence, we claim that in round r and later, the faulty agent f needs to be removed when computing the number of connections of other agents.

Lemma 7. *Suppose that the direct link state information of j in round r can be agreed by all nonfaulty agents in round m. If i is a nonfaulty agent in round m and agent j is considered to be a uncertain agent in HS_i^r, then j must be a faulty agent in HS_i^{r+1}.*

Proof. The proof argument is by contradiction. Assume that j is considered to be a nonfaulty agent or a uncertain agent in HS_i^{r+1}.

    (i) *Case 1.* j is a nonfaulty agent in HS_i^{r+1} when it is considered to be a uncertain agent in HS_i^r. j must send NS_j^r to at least 2 good agents in round $r + 1$ (Lemma 6). Then these good agents send the direct link states of j to all nonfaulty agents. Hence, j must be a certain agent in HS_i^r. A contradiction.

    (ii) *Case 2.* j is a uncertain agent in HS_i^{r+1} when it is considered to be a uncertain agent in HS_i^r. It is easy to see that the link states between j and good agents cannot be unknown-state in round r and $r + 1$. Since the number of good agents $n - t$ must be greater than t, j cannot have faulty links with all good agents. Then it must send NS_j^r to some good agents in round $r + 1$. Equally, j must be a certain agent in HS_i^r. A contradiction.

Thus, we reach contradictions in all cases, which proves the lemma. □

Lemma 8. *If round r is a clean round, the state of l_{ij}^{r-1} can reach a consensus by all nonfaulty agents in round $r + 2$, where $i, j \in Nf^{r-1}$ and $j \neq i$.*

Proof. We can pay attention to the state of l_{ij}^{r-1}. Consider the following cases:

    (i) *Case 1.* i and j are good agents. By Lemma 6, a risk agent must have correct links with some good agents. Hence, i sends $State_i[l_{ij}^{r-1}]$ to all good agents and risk agents having correct links with i in round r. And j also does this. In round r all good agents have two detection results of l_{ij}^{r-1}. Then after updating, they send the uniform state to all risk agents that have faulty links with i and j in round $r + 1$.

    (ii) *Case 2.* i and j are risk agents. i and j send their detection results of l_{ij}^{r-1} to some good agents (denoted by U) and risk agents in round r. Then two results are sent to all good agents by the agents in U in round $r + 1$. So every good agent knows the uniform state of l_{ij}^{r-1} in round $r + 1$. Therefore, all nonfaulty agents reach a consensus on the state in round $r + 2$.

    (iii) *Case 3.* i is a good agent and j is a risk agent. Similarly, it is easy to get that all good agents have the uniform state of l_{ij}^{r-1} in round $r + 1$ by case 1 and case 2. So this is what we want. Thus, the lemma holds. □

Lemma 9. *If round r is a clean round, then in the HS_i^{r-1} ($i \in Nf^{r+2}$), the state of link l_{kp} ($k \in Nf^{r-1}$ and $p \in N$) cannot be $Msg - O$.*

Proof. Assume, without influence, that the messages of p have no effect on the state of l_{kp}. Since k is a nonfaulty agent,

by Lemma 8, $State_k[l_{kp}^{r-1}]$ is received by all nonfaulty agents in round $r + 2$. Hence, i must know the state of l_{kp}^{r-1}. The lemma holds. □

Corollary 2. *If round r is a reliable round and the total number of rounds is greater than $r + 3$, there are not uncertain agents in round r.*

Proof. For a contradiction, let i be a uncertain agent in round r, by Lemma 7, we have two cases. For both case 1 and case 2, by Lemma 8, there are contradictions to the assumption. Hence, i must be a faulty agent in HS^{r+1}. The unknown-state link is regarded as correct link so that i is regarded as a nonfaulty agent in HS^r. Then it is a contradiction to the assumption that r is a reliable round. Thus, the lemma holds. □

Lemma 10. *There is at least one clean round in $t + 1$ rounds.*

Proof. Suppose, for a contradiction, that there is no clean round in $t + 1$ rounds. Then there must be new faulty agents added in each round. So there are at least $t + 1$ faulty agents in $t + 1$ rounds. This contradicts the assumption that there are at most t faulty agents. □

Corollary 3. *There are at least two clean rounds in $t + 2$ rounds.*

Corollary 4. *In $t + 2$ rounds, there must be one reliable round r in which at most one new faulty agent is detected.*

Proof. Suppose that there are a clean rounds in $t + 2$ rounds. We prove the lemma from two cases:

    (i) *Case 1.* All clean rounds are greater than round 1. And for a contradiction, two new faulty agents are detected in each reliable round. Then there are $2a$ faulty agents and there are still $t - 2a$ faulty agents remaining in $t + 2 - 2a$ rounds. It is easy to see that $t + 2 - 2a > t - 2a$. Therefore, there must be clean rounds in the remaining $t + 2 - 2a$ rounds. This contradicts the assumption that there are a clean rounds in $t + 2$ rounds.

    (ii) *Case 2.* Round 1 is a clean round. And for a contradiction, two new faulty agents are detected in each reliable round. Then there are $2(a - 1)$ faulty agents and there are still $t - 2a + 2$ faulty agents remaining in $t + 2 - 2a + 1$ rounds. Since $t + 2 - 2a + 1 > t - 2a + 2$, there must be clean rounds in the remaining $t + 2 - 2a + 1$ rounds, a contradiction.

Thus, we reach a contradiction in every case, which proves the lemma. □

Lemma 11. *In round $t + 3$, if a faulty agent $i \in F^{\Delta t+3}$ can receive messages from at least one good agent, the link states of the second clean round and all previous rounds can also reach a consensus among i and all nonfaulty agents.*

Proof. By Corollary 4, from the second clean round y to round $t + 3$, there must be a reliable round (suppose the first is r) in which at most one new faulty agent is detected because the total number of risk and faulty agents is $t - y + 2$ and the total number of rounds is $t - y + 4$. Suppose $r + 1 = y'$ which is a clean round. Since $F^{\Delta t+3} \neq \emptyset$, $y < y' < t + 3$. Since no new faulty agents are detected in round y', risk agents can only choose 2 in round y' (see details in Theorem 1). Hence, in round $y' + 1$, all good agents reach a consensus on the link states HS of round y and before rounds. We divide y' into two cases to prove as follows:

(i) *Case 1.* $y < y' < t + 2$. Then we have $y' + 2 \leq t + 3$. In round $y' + 2$, all good agents send the latest and uniform link states of round y and before rounds to all agents. Thus, i must reach a consensus.

(ii) *Case 2.* $y' = t + 2$. We assume that for the reliable rounds in which two or more faulty agents are detected, the faulty agents can be averaged to the next round and then the clean round can also be regarded as a normal round. Then it can be seen that the number of faulty agents keeps increasing in each round from round $y + 1$ to round $t + 1$. Thus, at least $t + 1 - y$ faulty agents have been added until round y'. Since there are $x_y (\leq t - y + 2)$ risk agents in round y, then at most one risk agent remains in round y' and it must be i. Then it is easy to get see i must reach a consensus in round $t + 3$.

Thus, the lemma holds. □

Theorem 2. *Consensus solves uniform consensus if at most t agents omit to send or receive messages, $n > 2t + 1$, and suppose that all agents are honest.*

Proof. Since $n > 2t + 1$, it is easy to see that no inconsistency is detected.

Termination. From Algorithm 1, nonfaulty agents must decide in round $t + 4$ and faulty agents decide before round $t + 4$.

Validity. Since no inconsistency is detected, all agents make decisions different from $\perp$. For agent i, if i decides a value decision$_i$, decision$_i$ must be the initial preference of an agent in decision set D. Since D depends on $HS_i^{m^*}$, it must have $D \subseteq N$. Therefore, decision$_i$ satisfies the validity property. If i decides $\parallel$, i has no decision and $\parallel$ does not affect the final consensus outcome. Thus, it also conforms to the validity.

Uniform Agreement. We prove this from the following cases:

(i) *Case 1.* Agents i and $j \in Nf^{t+4}$. By Corollary 3, there must be a decision round m^* in $t + 3$ rounds. And by Lemma 1, we have $HS_i^{1 \sim m^*+1} = HS_j^{1 \sim m^*+1}$. Then $D_i = D_j = D^*$. Since the pieces of preferences and proposals of all the agents in D^* must be saved by at least 2 good agents (Lemma 6), all good agents and some risk agents can restore all initial values and proposals of the agents in D^* in round $t + 3$. We

denote these agents by Nf_1^{t+4} and $Nf_2^{t+4} = Nf^{t+4} \setminus Nf_1^{t+4}$. Thus, all agents in Nf_1^{t+4} have the same set C and give a unified *consensus* set containing one value. That is, if agent u and $v \in Nf_1^{t+4}$, then there must be consensus$_u$ = consensus$_v$ = {cons} and |consensus$_u$| = |consensus$_v$| = 1 in round $t + 3$. And if agent $w \in Nf_2^{t+4}$, consensus$_w$ = $\emptyset$ in round $t + 3$.

(ii) *Case 2.* We analyze the agents in F^{t+4}. It is easy to see that $F^{t+4} = F^{t+2} \cup F^{\Delta t+3} \cup F^{\Delta t+4}$.

(1) *Case 2.1.* $i \in F^{t+2}$. i must decide $\parallel$ at the latest in the receive phase of round $t + 3$.

(2) *Case 2.2.* $i \in F^{\Delta t+3}$. Suppose that $F_1^{\Delta t+3}$ denotes the set of faulty agents that can receive the messages from some good agents in round $t + 3$ and send messages in round $t + 4$. Then $F_2^{\Delta t+3} = F^{\Delta t+3} \setminus F_1^{\Delta t+3}$. It is easy to see that the agents in $F_2^{\Delta t+3}$ definitely do not send messages in round $t + 4$ and they must decide $\parallel$ in round $t + 3$. If $i \in F_1^{\Delta t+3}$, by Lemma 11, D_i must be the same as D^* of good agents in case 1. Thus, if i can restore all initial preferences and proposals in D_i, it must have consensus$_i$ = {cons} and |consensus$_i$| = 1 in round $t + 3$. Otherwise, consensus$_i$ = $\emptyset$ in round $t + 3$.

(3) *Case 2.3.* $i \in F^{\Delta t+4}$. Since i is a nonfaulty agent in round $t + 3$, consensus$_i$ has the same two possible states in round $t + 3$ as in case 1. Suppose $F^{\Delta t+4} = F_1^{\Delta t+4} \cup F_2^{\Delta t+4}$. $F_1^{\Delta t+4}$ denotes the set of faulty agents that cannot detect faulty by itself in the receive phase of round $t + 4$. $F_2^{\Delta t+4}$ denotes the set of faulty agents that can detect that they become faulty agents in the receive phase of round $t + 4$. Therefore, the agents in $F_1^{\Delta t+4}$ decide *cons* by consensus$_i$ and the agents in $F_2^{\Delta t+4}$ decide $\parallel$ in round $t + 4$.

In summary, consensus set only has two types in round $t + 3$ and round $t + 4$: {cons} and $\emptyset$. The agents in $Nf_1^{t+4} \cup Nf_2^{t+4} \cup F_1^{\Delta t+4}$ decide cons in round $t + 4$, the agents in $F^{t+2} \cup F^{\Delta t+3}$ decide $\parallel$ before round $t + 4$, and the agents in $F_2^{\Delta t+4}$ decide $\parallel$ in round $t + 4$. Thus, uniform agreement holds.

To achieve the Nash equilibrium, we make some appropriate assumptions about initial preferences and failure patterns. Failure pattern represents a set of failures that occur during an execution of the consensus protocol [12]. Specifically, we assume that initial preferences and failure patterns are blind. □

Definition 4. The blind initial values mean that each agent cannot guess the preferences of other agents and the probability of its own preference becoming the consensus cannot be improved by trusting others.

By Definition 4, we can get that if an agent wants to improve its own utility, it can only rely on itself, for example, increasing the probability of entering the decision set and reducing the number of agents in the decision set and so on.

Definition 5. The blind failure patterns mean that before t faulty agents appear, an agent cannot guess the link states in the following rounds. Then we have that

  (i) If agent i does not know the link states of round m in round r and j is a nonfaulty agent in round r, then$P(i \in F^{\Delta m}|\text{link}_{ij}$ is faulty$) = P(j \in F^{\Delta m}|\text{link}_{ij}$ is faulty$) \leq \alpha$.

  (ii) For round m_1 and m_2, if $m_1 < m_2$ and i does not know the link states of round m_2, then $P(v_i \text{becomesconsensus}|m_1 \text{isthedecisionround})$ $=P(v_i \text{becomesconsensus}|m_2$ is the decision round$)$.

  (iii) For $t + 1$ rounds, if the link states of each round in $t + 1$ rounds are unknown to agent i, then for a round r in $t + 1$ rounds,$P(r$ is a clean round$) \geq 1/(t + 1)$.

Theorem 3. *If $n > 2t + 1$, at most t agents have omission failures at the same time, agents prefer consensus, and failure patterns and initial preferences are blind, then $\overrightarrow{\sigma}^{CONSENSUS}$ is a Nash equilibrium.*

Proof. To prove Nash equilibrium, we need to show that it is impossible for each agent $i \in N$ to increase its utility u_i with all possible deviations σ_i. That is, proving that for each agent i, there must be

$$u_i\left(\sigma_i^{CONSENSUS}, \sigma_{-i}^{CONSENSUS}\right) \geq u_i\left(\sigma_i, \sigma_{-i}^{CONSENSUS}\right). \quad (1)$$

We use the same deduction method as in [12]. Consider all the ways that i can deviate from the protocol to affect the outcome as follows:

  (1) i generates a different value $v_i' \neq v_i$ (or proposal$_i' \neq$ proposal$_i$) and sends $q_i'(j)$ (or $b_i'(j)$) to some agents $j \neq i$.

  (2) $q_i(j)$ (or $b_i(j)$) sent by i cannot restore v_i (or proposal$_i$).

  (3) i does not choose random$_i$ or X-random$_i$ or proposal$_i$ appropriately, such as not randomly.

  (4) i sends an incorrectly formatted message to $j \neq i$ in round m.

  (5) If $|\text{lost}_i| > t$ in round m, i does not decide $\parallel$, but continues to execute the following protocol.

  (6) i lies about the state of l_{kp} in round m; that is, in round m, i sends a state of l_{kp} which is different from $NS_i^{m-1}[l_{kp}]$.

  (7) i sends an incorrect random or X-random of l_{kp} to j in round m.

  (8) i sends an incorrect $q_l(i)$ (or $b_l(i)$) to $j \neq l$ different from the $q_l(i)$ (or $b_l(i)$) that i has received from l in round 1.

  (9) i sends an incorrect consensus$_i$ to $j \neq i$ in round $t + 4$.

  (10) i pretends to crash in round m.

We consider these deviations one by one and prove that i does not gain by any of deviations. That is, equation (1) holds if i deviates from the protocol by these deviations on the list above.

  (i) *Type 1.* (i) If i sends $q_i'(j)$ to some agents, then either an inconsistency is detected because of secret restoring error, or i does not gain. Specifically, if i is the agent whose value is chosen, then i is worse off if it lies than it does not, since some agents cannot restore v_i, but they can restore it when following the protocol. Then if i is not the agent whose preference is chosen, then it does not affect the outcome. (ii) i sends $b_i'(j)$. Then an inconsistency is detected if restoring polynomial error or generating different consensus values in the system. And if no inconsistency is detected, then either all agents that receive b_i or b_i' are faulty or both b_i and b_i' do not affect the final outcome. Since changing the proposal cannot increase i's utility, i does not gain. Therefore, both (i) and (ii), i does at least as well if i uses the strategy $\sigma_i^{CONSENSUS}$, as it deviates from the protocol according to type 1. So, equation (1) holds.

  (ii) *Type 2.* It is easy to see that either an inconsistency is detected or no benefit because there is no increase in the probability that v_i becomes the consensus. Thus, equation (1) holds.

  (iii) *Type 3.* (i) Since other agents follow the protocol, it does not affect the final outcome because the two kinds of random numbers are only used for verification. (ii) Since i does not know the proposals of other agents in round 1, then using different proposals cannot improve the probability that v_i becomes the consensus. Thus, equation (1) holds.

  (iv) *Type 4.* If i sends an incorrectly formatted message to j, then either an inconsistency is detected by j or it does not affect the outcome since j omits to receive messages from i in round m. Thus, i does not gain, so equation (1) holds.

  (v) *Type 5.* Since $|\text{lost}_i| > t$, i does not receive messages from at least $t + 1$ agents in round m, that is, i has receiving omission failures with at least two good agents.

    (1) *Case 1.* i does not guess message random numbers in round $m + 1$. Then by Claim 1, an inconsistency is detected.

    (2) *Case 2.* i guesses the message random number in round $m + 1$ and has correct links with the remaining $n - t - 2$ agents, and these agents are all nonfaulty agents. Then by the Claim 1, i can successfully send messages in round $m + 1$ iff i can guess a message random number random$_j^m$ from a nonfaulty agent j and i has no sending omission failures with j. That is, the random guessing does not change the detection result of the state of i by other agents in

round m. Clearly the probability that i can guess a random number is $1/n$.

(a) *Case 2.1.* If i guesses some random numbers from agents j and has sending omission failures with each agent j, then it does not affect the outcome even if the random numbers are correct guesses.

(b) *Case 2.2.* Suppose that i guesses only one random number from the good agent j in a round and i does not have sending omission failures with j. If i only guesses the message random number in round $m + 1$, then we have that

$$u_i\left(\sigma_i, \sigma_{-i}^{\mathrm{CONSENSUS}}\right) \leq \frac{1}{n} P \left(\text{the decision round is in } m + 1 \text{ rounds}\right) \frac{1}{|D|} \beta_0 + \gamma_1 \beta_1 + \gamma_2 \beta_2. \tag{2}$$

Thus,

$$u_i\left(\sigma_i, \sigma_{-i}^{\mathrm{CONSENSUS}}\right) \leq \frac{1}{n} \times \frac{1}{n - t + 1} \beta_0 + \gamma_1 \beta_1 + \gamma_2 \beta_2, \tag{3}$$

which means that i only guesses one random number in round $m + 1$ and then i must be in decision set D ($|D| \geq |G| = n - t$). Since $i \in D$, there are at least $n - t + 1$ agents in D. It is easy to see that if i guesses random numbers in multiple rounds, the utility must be less than (3). If i follows the protocol, then i must decide $\parallel$ in round m. Since i has receiving omission failures in round m, the link states after round $m - 1$ must be unknown to i. Thus, by Definition 1 and the assumptions about omission failures, we can get that

$$P \left(\text{round } m - 1 \text{ or } m \text{ is a clean round}\right) \geq \frac{1}{t + 1}. \tag{4}$$

Then

$$u_i\left(\sigma^{\mathrm{CONSENSUS}}\right) \geq \frac{1}{t + 1} \times \frac{1}{|D|_{m-2}} \beta_0 + \gamma^c \beta_1. \tag{5}$$

Since $n > 2t + 1$, (1) must hold.

(c) *Case 2.3.* If i guesses more than one random numbers in round $m + 1$, then the utility of i must be less than (3). And i does at least well by following the protocol as the deviation because the guessing work does not affect the state of i in round m. Specifically, either if i is a nonfaulty agent detected by other agents, then (5) holds, or if i is a faulty agent, then it does not affect the outcome even if i guesses the random correctly. Thus, (1) holds.

(3) *Case 3.* If i has sending omission failures with the remaining $n - t - 2$ agents, then either no benefit since i is faulty in round m detected by other agents, or i does not gain if i is nonfaulty

because the utility of deviating from the protocol must be less than (3).

In summary, either an inconsistency is detected by Claim 1 if i does not guess the message random numbers, or no benefit from guessing the message random numbers. Thus, yet again, (1) holds.

(vi) *Type 6.* By the proof of Type 5, it is easy to see that i must be a nonfaulty agent detected by i itself in round $m - 1$. Since there is more than one state of a link, we partition this deviation into eight cases and show that i does at least well by using $\sigma_i^{\mathrm{CONSENSUS}}$ as it deviates from the protocol by these eight deviations.

(1) *Case 1.* k or $p = i$, such as $k = i$, and Type $(NS_i^{m-1}[l_{kp}^r]) = Msg\text{-}R$ where $r \leq m - 1$, and i pretends Type$(\mathrm{State}_i[l_{kp}^r]) = Msg\text{-}X$ in round m.

(a) *Case 1.1.* $r < m - 1$. Then $\mathrm{State}_i[l_{kp}^r]$ must be sent in round $r + 1$ in order to enable message chain mechanism to succeed. Since $r + 1 < m$, an inconsistency must be detected in round m.

(b) *Case 1.2.* $r = m - 1$ and r is the decision round m^*. Since i does not know the link states of round $m - 1$ in the sending phase of round m, by Definition 1, if p is a nonfaulty agent in round m, then $P(i \in F^{\Delta m - 1} | i \text{ pretends link}_{ip} \text{ is faulty}) = P(p \in F^{\Delta m - 1} | i \text{ pretends link}_{ip} \text{ is faulty}) \leq \alpha$. Suppose that the decision set in round r is D when following the protocol. We can see that $|D| \geq 3$. (i) If all agents become faulty due to the deviation of i, then there is no consensus in the system and i does not gain. (ii) If the deviation does not cause no solution and p is a nonfaulty agent in round $m - 1$, then we have that&ecmath;

$$u_i\left(\sigma_i, \sigma_{-i}^{\text{CONSENSUS}}\right) \le \alpha(1-\alpha)\frac{1}{|D|-1}\beta_0 + \alpha(1-\alpha)\frac{|D|-2}{|D|-1}\beta_1 + (1-\alpha)^2\frac{1}{|D|}\beta_0 + (1-\alpha)^2\frac{|D|-1}{|D|}\beta_1 + \alpha\beta_1.$$

$$(6)$$

That is,

$$u_i\left(\sigma_i, \sigma_{-i}^{\text{CONSENSUS}}\right) \le \frac{(1-\alpha)(|D|-1+\alpha)}{(|D|-1)|D|}\beta_0 + \left[\alpha(1-\alpha)\frac{|D|-2}{|D|-1} + (1-\alpha)^2\frac{|D|-1}{|D|} + \alpha\right]\beta_1.$$

$$(7)$$

It is easy to get that

$$u_i\left(\sigma^{\text{CONSENSUS}}\right) = \frac{1}{|D|}\beta_0 + \frac{|D|-1}{|D|}\beta_1. \qquad (8)$$

If $\alpha = 0$, then $(1-\alpha)(|D|-1+\alpha)$ in (7) takes its supremum $|D|-1$. Hence, there must be equation (1). (iii) If p is a faulty agent in round $m-1$, then the deviation does not affect the outcome. Thus, cases (i),(ii), and (iii) hold equation (1).

(c) *Case 1.3.* $r = m-1$ and $r = m^* + 1$. Since the link states of next reliable round are unknown to i, either there is no solution since all agents become faulty; the deviation of i does not affect the final outcome; or by Definition 1, i does at least well by using $\sigma_i^{\text{CONSENSUS}}$ as it deviates from the protocol because the probability that v_i becomes consensus does not increase. Thus, equation (1) holds.

(d) *Case 1.4.* $r = m-1$ and $r \ne m^*$ and $r \ne m^* + 1$. Either there is no solution since all agents become faulty or there is no benefit because it does not affect the decision round.

In summary, all cases in case 1 cannot make i gain.

(2) *Case 2.* k or $p = i$, such as $k = i$, and Type$(NS_i^{m-1}[l_{kp}^r]) = Msg$-$X$ where $r \le m-1$, and i pretends Type$(\text{State}_i[l_{kp}^r]) = Msg$-$R$ in round m. If r < m-1, i does not gain, which is the same as that in case 1.1. If $r \le m-1$, i does not gain, which is the same as that in case 1.1. If $r = m-1$, since i is nonfaulty in round $m-1$ and i does not know the link states of round $m-1$, then by Definition 1, there is no benefit in guessing the message random number with the probability $1/n$. So equation (1) holds.

(3) *Case 3.* k and $p \ne i$, and Type$(NS_i^{m-1}[l_{kp}^r]) = Msg$-$R$ or $Msg-O$, and i pretends Type$(\text{State}[l_{kp}^r]) = Msg$-$X$ in round m. Suppose i lies about the detection result $\text{State}_k[l_{kp}^r]$ of k. (i) If k is faulty in round r, it does not affect the final outcome even if no inconsistency is detected. (ii) If k is nonfaulty in round r, then k must send the faulty random numbers

X-random$_k^r$ to at least two good agents (Lemma 6). And i guesses a faulty random number with the probability $1/2$. If it does not affect the states of k and p, then i does not gain even if guessing the random correctly. Otherwise, if the decision round is changed, then

$$u_i\left(\sigma_i, \sigma_{-i}^{\text{CONSENSUS}}\right) \le \frac{1}{2} \times \frac{1}{n-t}\beta_0 + \frac{1}{2} \times \frac{n-t-1}{n-t}\beta_1 + \frac{1}{2}\beta_2.$$

$$(9)$$

And since i is nonfaulty in round $m-1$, then

$$u_i\left(\sigma^{\text{CONSENSUS}}\right) \ge \frac{1}{n}\beta_0 + \frac{n-1}{n}\beta_1. \qquad (10)$$

By the definition of $n > 2t + 1$, equation (1) holds. So equation (1) holds both in (i) and (ii).

(4) *Case 4.* k and $p \ne i$, and Type$(NS_i^{m-1}[l_{kp}^r]) = Msg$-$X$ or Msg-O, and i pretends Type$(\text{State}[l_{kp}^r]) = Msg$-$R$ in round m. We also suppose that i lies about the detection result $\text{State}[l_{kp}^r]$ of k. (i) If Type$(NS_i^{m-1}[l_{kp}^r]) = Msg$-$O$, then it does not affect the final outcome even if i guesses the random correctly, since Msg-R and Msg-O have the same meaning when computing the state of an agent. (ii) If Type$(NS_i^{m-1}[l_{kp}^r]) = Msg$-$X$, then either an inconsistency is detected by message random number verification or link state conflict; it does not affect the outcome if the states of k and p are unchanged after deviating; or it makes round r a clean round. Then if there is already a decision round r^* and $r^* \le r-1$, it does not affect the outcome because we need the first reliable round finally. And if $r^* > r-1$, then the utility of i decreases because decision round is advanced compared to following the protocol. If there is no decision round, by Definition 1, i does at least well by using $\sigma_i^{\text{CONSENSUS}}$ as it deviates from the protocol. Hence, equation (1) holds again.

(5) *Case 5.* k or $p = i$, and Type$(NS_i^{m-1}[l_{kp}^r]) = Msg$-$R$ or Msg-X where $r \le m-1$, and i pretends Type$(\text{State}_i[l_{kp}^r]) = Msg-O$ in round m. By Claim 1, an inconsistency is detected. Thus, equation (1) holds.

(6) *Case 6.* k and $p \neq i$, and $\text{Type}(NS_i^{m-1}[l_{kp}^r]) = Msg - X$ or $Msg\text{-}R$, and i pretends $\text{Type}(\text{State}[l_{kp}^r]) = Msg\text{-}O$ in round m. Since $Msg - R$ and $Msg\text{-}O$ have the same meaning when computing the state of an agent, there is no benefit, which is the same as that in case 4.

(7) *Case 7.* k and $p \neq i$, and $\text{Type}(NS_i^{m-1}[l_{kp}^r]) = X(r)$, and i pretends $\text{Type}(\text{State}[l_{kp}^r]) = X(r - 1)$ in round m. We can turn this case into case 3 because the type of $NS_i^{m-1}[l_{kp}^{r-1}]$ must be $Msg - R$. So it has the same result as that in case 3. Thus, yet again, equation (1) holds.

(8) *Case 8.* k or $p = i$, such as $k = i$, and $\text{Type}(NS_i^{m-1}[l_{kp}^r]) = Msg\text{-}R$ where $r \leq m - 3$, and i receives $\text{Type}(\text{State}_p[l_{kp}^r]) = Msg\text{-}X$, and i pretends $\text{Type}(\text{State}_p[l_{kp}^r]) = Msg\text{-}O$ or $Msg\text{-}R$ in round m. (i) If p is a nonfaulty agent in round $r + 1$, then it must send $\text{State}_p[l_{kp}^r]$ to at least two good agents except i in round $r + 1$. Thus, all nonfaulty agents must know $\text{State}_p[l_{kp}^r]$ in round m, so that i does not gain. (ii) If p is faulty in round $r + 1$, then since i is nonfaulty in round $m - 1$, i is also a nonfaulty agent in round r even if the link between i and p is faulty. Thus, the results are the same as those in case 4 and case 6. Therefore, equation (1) holds both (i) and (ii).

In summary, in any case, i's utility is at least as high with $\sigma_i^{\text{CONSENSUS}}(\sigma_i, \sigma_{-i}^{\text{CONSENSUS}})$ as with $u_i(\sigma_i, \sigma_{-i}^{\text{CONSENSUS}})$.

(vii) *Type 7.* Since the random numbers are only used in inconsistency detection, either j detects an inconsistency and decides $\perp$ or it does not affect the final outcome if no inconsistency is detected. Thus, equation (1) holds.

(viii) *Type 8.* It is easy to see that either an inconsistency is detected due to consensus difference or restoring secret faulty; it does not affect the outcome if $l \notin D$ or l is not the agent whose preference is chosen; or i does not gain due to the blind initial preferences by Definition 1 and the random agents' proposals.

(ix) *Type 9.* Clearly it does not affect the outcome if j decides $\parallel$ in the receiving phase of round $t + 4$ or j does not receive the messages from i. Otherwise, j must receive the messages from at least two good agents in last round. Then j detects an inconsistency and decides $\perp$. So equation (1) holds.

(x) *Type 10.* We divide this type into two cases to prove.

(1) *Case 1.* There is no consensus in the system. This case happens when either all agents become faulty due to the deviation, or restoring secret faulty in round $t + 3$ for all good agents because of the missing pieces of i. Thus, i's utility with pretending to crash is

lower than with following the protocol in case 1.

(2) *Case 2.* There is a consensus finally. (i) $m \leq m^*$. Since i does not send messages to any agents before decision round, i cannot exist in decision set D. Thus, i's utility also decreases when pretending the protocol. (ii) $m > m^*$. It does not affect the outcome. Therefore, (1) holds both in cases 1 and 2.

Finally, concluding the proof. $\qquad\square$

5. Conclusion

In this paper, we introduce game theory as an interpretable method for studying the algorithms in multiagent system and provide an algorithm for uniform consensus that is resilient to both omission failures and strategic manipulations. We prove that our uniform consensus is a Nash equilibrium as long as $n > 2t + 1$, and failure patterns and initial preferences are blind. Additionally, we present the theory of message passing in presence of process omission failures. We argue that our research enriches the theory of fault-tolerant distributed computing and strengthens the interpretable reliability of consensus with omission failures from the perspective of game theory. And our contribution provides a theoretical basis for the combination of distributed computing and strategic manipulations in omission failure environments, which we think is an interesting research area.

In our opinion, there are many interesting open problems and research directions which are not covered in this paper. We list a few here: (a) whether an algorithm for rational uniform consensus exists if coalitions are allowed; (b) the study of rational consensus with more general types of failures, such as Byzantine failures, is important; (c) with the problem setting of this paper, whether the rational consensus exists if we relax the constraint $n > 2t + 1$; (d) studying the rational consensus in asynchronous system, which seems significantly more complicated; and (e) introducing the assumption of agent bounded rationality may be useful in practical scenarios.

Data Availability

The data and proof used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] L. Lamport and N. Lynch, "Distributed computing: models and methods," *Formal Models and Semantics*, Elsevier, Amsterdam, Netherlands, 1990.

[2] V. Hadzilacos and S. Toueg, "A modular approach to fault-tolerant broadcasts and related problems," Report, Cornell University, Ithaca, NY, USA, 1994.

[3] B. Charron-Bost and A. Schiper, "Uniform consensus is harder than consensus," *Journal of Algorithms*, vol. 51, no. 1, pp. 15–37, 2004.

[4] X. Bei, W. Chen, and J. Zhang, "Distributed consensus resilient to both crash failures and strategic manipulations," 2012, https://arxiv.org/abs/1203.4324.

[5] J. Halpern and V. Teague, "Rational secret sharing and multiparty computation," in *Proceedings of the 36th annual ACM symposium on Theory of computing (STOC)*, pp. 623–632, 2004.

[6] A. Lysyanskaya and N. Triandopoulos, "Rationality and adversarial behavior in multi-party computation," in *Annual International Cryptology Conference (CRYPTO)*Springer, Berlin, Germany, 2006.

[7] G. Fuchsbauer, J. Katz, and D. Naccache, "Efficient rational secret sharing in standard communication networks," in *Theory of Cryptography Conference (TCC)*, pp. 419–436, Springer, Berlin, Germany, 2010.

[8] V. Dani, M. Movahedi, Y. Rodriguez, and J. Saia, "Scalable rational secret sharing," in *Proceedings of the 30th annual ACM SIGACT-SIGOPS symposium on Principles of distributed computing (PODC)*, pp. 187–196, San Jose, CA, USA, January, 2011.

[9] A. Yifrach and Y. Mansour, "Fair leader election for rational agents in asynchronous rings and networks," in *Proceedings of the 2018 ACM Symposium on Principles of Distributed Computing (PODC)*, pp. 217–226, 2018.

[10] I. Abraham, D. Dolev, and J. Y. Halpern, "Distributed protocols for leader election," *ACM Transactions on Economics and Computation*, vol. 7, no. 1, pp. 1–26, 2019.

[11] K.-M. Chung, T.-H. H. Chan, T. Wen, and E. Shi, "Game-theoretic fairness meets multi-party protocols: the case of leader election," in *Proceedings of the Annual International Cryptology Conference (CRYPTO)*Berlin, Germany, Springer, August 2021.

[12] J. Y. Halpern and X. Vilaça, "Rational consensus," in *Proceedings of the 2016 ACM Symposium on Principles of Distributed Computing (PODC)*, pp. 137–146, 2016.

[13] A. Clementi, L. Gualà, G. Proietti, and G. Scornavacca, "Rational fair consensus in the gossip model," in *Proceedings of the IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, pp. 163–171, IEEE, Orlando, FL, USA, May 2017.

[14] Y. Amoussou-Guenou, B. Biais, M. Potop-butucaru, and S. Tucci-Piergiovanni, "Rational vs byzantine players in consensus-based blockchains," in *Proceedings of the 19th International Conference on Autonomous Agents and Multi-Agent Systems*, pp. 43–51, 2020.

[15] I. Harel, A. Jacob-Fanani, M. Sulamy, and Y. Afek, "Consensus in equilibrium: can one against all decide fairly?" in *Proceedings of the 23rd International Conference on Principles of Distributed Systems (OPODIS 2019)*, Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2020.

[16] A. Ranchal-Pedrosa and V. Gramoli, "Rational agreement in the presence of crash faults," in *Proceedings of the IEEE International Conference on Blockchain (Blockchain)*, pp. 470–475, IEEE, Melbourne, Australia, December 2021.

[17] L. Solodkin and R. Oshman, "Truthful information dissemination in general asynchronous networks," in *Proceedings of the 35th International Symposium on Distributed Computing (DISC 2021)*, Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2021.

[18] Y. Afek, Y. Ginzberg, S. Landau Feibish, and M. Sulamy, "Distributed computing building blocks for rational agents," in *Proceedings of the 2014 ACM symposium on Principles of distributed computing (PODC)*, pp. 406–415, 2014.

[19] A. Groce, J. Katz, A. Thiruvengadam, and V. Zikas, "Byzantine agreement with a rational adversary," in *International Colloquium on Automata, Languages, and Programming (ICALP)* Springer, Berlin, Germany, 2012.

[20] A. S. Aiyer, L. Alvisi, A. Clement, M. Dahlin, J.-P. Martin, and C. Porth, "Bar fault tolerance for cooperative services, ACM SIGOPS Operating Systems Review," in *Proceedings of the 20th ACM symposium on Operating systems principles (SOSP)*, vol. 39, no. 5, pp. 45–58, 2005.

[21] C. Delporte-Gallet, H. Fauconnier, and F. C. Freiling, "Revisiting failure detection and consensus in omission failure environments," in *International Colloquium on Theoretical Aspects of Computing*, pp. 394–408, Springer, Berlin, Germany, 2005.

[22] C. R. F. Campusano, "Distributed eventual leader election in the crash-recovery and general omission failure models," Thesis, upv/ehu, Biscay, Spain, 2020.

[23] P. R. Parvédy and M. Raynal, "Optimal early stopping uniform consensus in synchronous systems with process omission failures," in *Proceedings of the 16th annual ACM symposium on Parallelism in algorithms and architectures (SPAA)*, pp. 302–310, 2004.

[24] A. Shamir, "How to share a secret," *Communications of the ACM*, vol. 22, no. 11, pp. 612-613, 1979.

Retraction

Retracted: Problems and Countermeasures of China's Greenway Economic Development from the Perspective of Computer Internet

Computational Intelligence and Neuroscience

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] S. Zhang, "Problems and Countermeasures of China's Greenway Economic Development from the Perspective of Computer Internet," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 6286833, 11 pages, 2022.

Research Article

Problems and Countermeasures of China's Greenway Economic Development from the Perspective of Computer Internet

Shuping Zhang [ID]

Hotel Management Major, Department of Tourism and Foreign Languages, Ma'anshan Teacher's College, Ma'anshan, Anhui 243041, China

Correspondence should be addressed to Shuping Zhang; 10119@massz.edu.cn

Given the advantages of urban greenways (UGCs) to urban populations' health and well-being, the growing usage of UGCs has gotten a lot of attention. However, most studies on UGCs have been undertaken in Western nations, but the understanding of greenway usage in underdeveloped countries is scarce. China's current environmental change in urban enhancement has been emphasized by a rush of urban green projects. The greenway is a kind of green infrastructure that has lately been the focus of many municipal governments. The federal government's new, ecologically friendly approach to urbanization is the subject of this article, which offers a preliminary assessment of the role played by greenways in that strategy. UGCs are green landscape routes designed for walkers and bicycles in cities, according to landscape design. A growing number of regions and cities in China are participating in the development of UGC, with some projects meeting their goals on an annual basis. UGCs are being built in China; however, a number of challenges exist, including a lack of legal basis, the public's incorrect understanding of UGCs, and insufficient scientific investigation. According to research on UGC construction and analysis of the peculiarities of various urban developments, the article highlighted current challenges in China's UGC construction and provided corresponding remedies in order to enhance China's UGC construction. In this survey, we performed statistical analysis using the chi-square test as well as the ANOVA test.

1. Introduction

China's economy has grown rapidly during the last few decades. By 2017, China's GDP had grown from 149.6 billion to 12.25 trillion USD, making it one of the world's wealthiest nations, with an additional 800 million people gaining access to better living conditions [1]. Increasing levels of environmental degradation, energy utilization, and industrial productivity have been associated with this growth [2]. Environmental deterioration, biodiversity loss, and soil erosion are all consequences of rapid economic growth. Figure 1 illustrates the effect of environmental degradation by minimizing the environmental expenses of the nation. China's central government is more worried about the economic and social consequences of pollution. As a result, China's environmental issues are becoming worse due to escalating tensions between the country's need to protect the environment and its desire to grow economically.

China is progressively shifting from a focus on quick economic growth to a focus on high-quality economic development. In today's world, the global economy must undergo a green transition. Entrepreneurship in China's cities is greening because of a desire to use the environment to promote economic growth, according to recent research. Sustainable green growth is an inextricable aspect of sustainable development and a quality-development method that enables nations to overcome old technologies in a variety of areas through the utilization of energy-efficient and clean technology. For that aim, China has to take action to establish enabling conditions for the development of energy-efficient and clean technology, including laws and practices in its own nation and abroad, as well as technological transfer. Innovation in technology, environmental legislation, government involvement, industrial cooperation, utilization of renewable energy, and green total factor
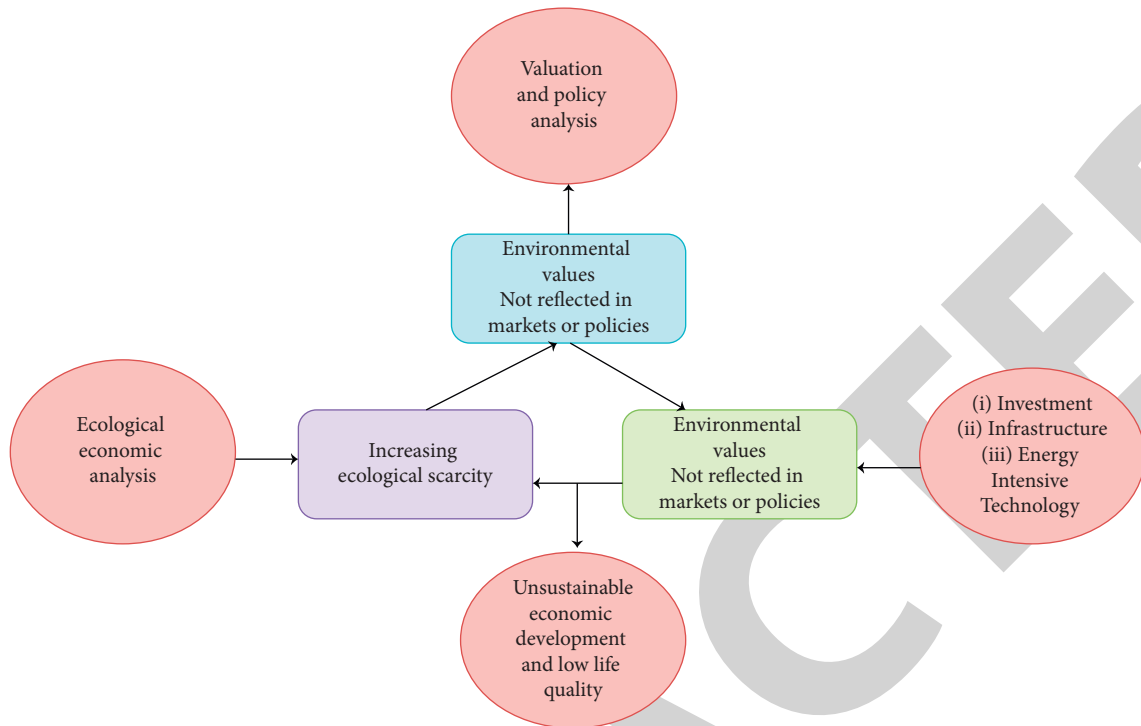
FIGURE 1: Relationship between environmental degradation and economic development.

productivity are all covered in the context of "China's manufacturing and logistics business."

Reducing pollution and the negative impact of heavy industry's dispersion on public perception are two critical components of economic reform's long-term success. Urban green infrastructure serves to emphasize the value of nature in urban settings. It has emerged in urban studies as a focus on the economics of financing urban green infrastructure programmes, such as park construction, urban stream restoration, and urban farm creation. Many of these programmes are aimed at attracting new economic investments and providing a green premium for real estate development by promoting a cleaner, greener image for communities, among other things. A natural, green trail built on protected linear corridors would enhance the environmental aspects and provide opportunities for outdoor recreation. With footpaths and/or cycle ways that are surrounded by trees, streams, and other natural features that frequently pass between urbanized areas and rural areas in China, this notion is realized.

Greenways have been reproduced around the country after receiving very good public reaction since their prototype development in 2010. However, greenway development is associated with some problems. A comprehensive survey on problems in greenway development and countermeasures for green economic development using Internet technologies is less investigated. As a result, this study analyses the issues and solutions related to China's greenway economic development from the perspective of the Internet and computers. The advantages of greenway in economic development and social well-being of Chinese people are also presented in this review. In addition, applications of smart computer Internet technologies in enhancing greenway planning and green economic development of China are provided in this paper.

2. Need for Greenway Development

A greenway is a kind of open, green area that runs the length of a city or town. Natural and manmade corridors like valleys, rivers, mountain ranges, and landscaped highways are common places for this kind of development to take root. Walking and cycling paths may be found throughout the park. Major parks and natural reserves are linked with historic sites and urban and rural communities. Because of their linear, connecting design, greenways are distinct from parks and other alternate open spaces observed in cities. A national programme may sensibly encompass green infrastructure and activities such as tree planting on highways, permeable vegetation on the surface, drainage basins, and green corridors inside these protected areas. To achieve sustainable development, it is necessary to properly harness natural resources as well as human resources, capital generation, and technology use [3]. The benefits of greenway development are explained in the following section and also depicted in Figure 2.

2.1. Tourism and Quality Economic Development. Greenways may be linked to tourism, serve as new tourist sites, and contribute to the economic growth of the areas they pass through. As a result, a tourism-related concept of greenways might offer the opportunity of generating new permanent tourist attractions, lowering seasonality in visitor
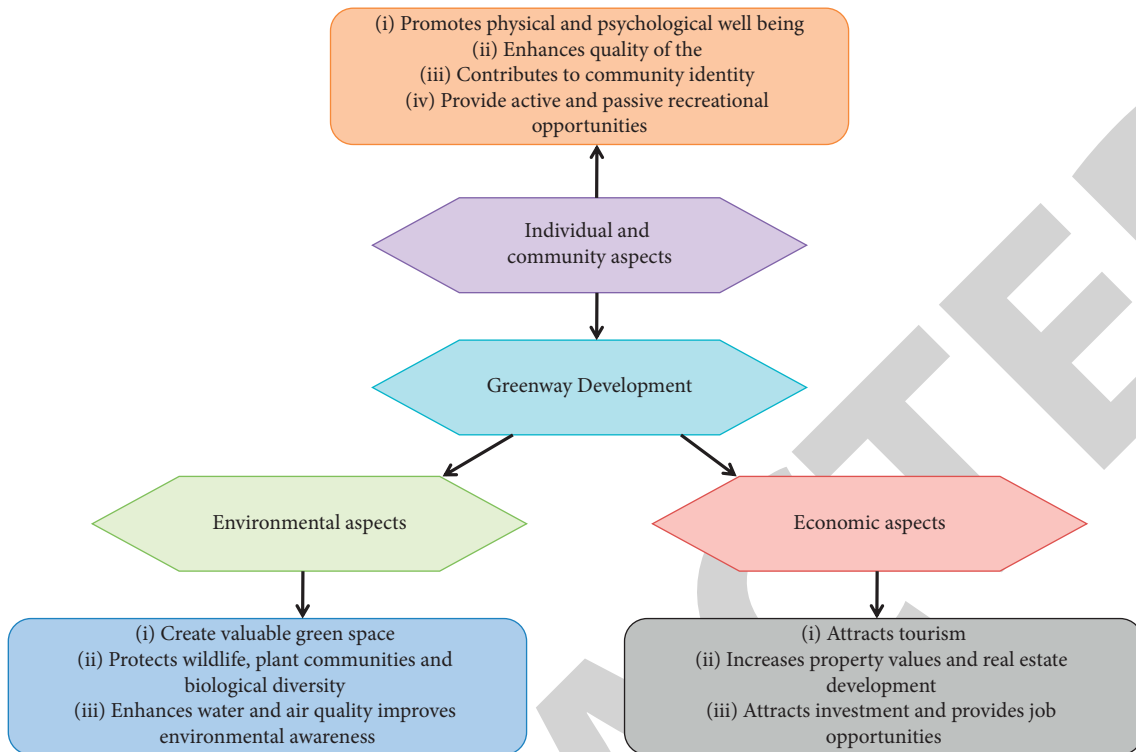
FIGURE 2: Overall advantages of greenways.

flows, and establishing tourism brands. Increased growth of greenways in tourist destinations through the establishment of pathway networks provides more pleasant routes for travel by bicycle or on foot across the regions, allowing tourists to better understand the region's culture, heritage, and landscape. This will contribute to the region's overall economic development and decrease CO_2 emissions. As a result, it is critical to make studies on whether investments in greenways deliver both economic and environmental benefits accessible to the political authorities and scientific community. One method of analyzing such economic data is to compare the costs of using greenways to the expenses of building infrastructure. Visitors come to Greenway from all over the world. It is not uncommon for towns to see the economic benefits of attracting visitors to these popular tourist attractions. Construction and maintenance jobs, as well as tourism-related possibilities such as river rafting trips, bicycle renting, restaurants, and hotels, are all encouraged by the development of trails and greenways.

Green infrastructure allows towns to be certified as protected tourism destinations. One billion visitors and one billion possibilities were coined by "the United Nations World Tourism Day Conference" in 2015 to emphasize the tourism industry's power to influence people's environmental behaviour. According to research, approximately 93% of Chinese tourists are choosing eco-friendly destinations when making vacation plans, and this trend is expected to continue in the future [4]. From a purely ecological perspective, Shanghai's ecological green belts have developed into both ecological and recreational spaces [5].

2.2. Psychological Well-Being. It is predicted that by 2050, more than 70 percent of the globe's population will live in cities, as per the "World Health Organization, 2013." There are 78% of Taiwanese people living in urban areas, according to the "Directorate General for Budget, Accounting, and Statistics, 2018." As urban green space disappears, so does the opportunity for urban people to interact with nature. With almost 20% of the population over the age of 65, Taiwan is on course to become a "super-aged" society by 2026 [6]. One-fifth to one-ninth of those 65 and older suffer from depression [7]. Suicide is more common among older people in Taiwan than in other age groups, and it climbs with age. Researchers, health care providers, and communities are now giving closer attention to older adults, which includes not only disease therapies but also effective psychological transitions, the advent of a healthy lifestyle, participation in physical activities, and societal engagement [8].

It is critical to maintain and prolong the lives of older persons, and urban green areas give senior inhabitants accessible access to nature. For example, UGWs are open spaces with varying widths that run parallel to roads and provide public access for recreational purposes. It is possible that greenways may improve the environment of urban regions, replace the lack of green space in advanced metropolitan regions, and offer stress-relieving and recreational possibilities for urban residents, strengthening their emotional links to the area [9]. There has been a lot of research on the advantages of urban green spaces, but most of it has concentrated on park visitors rather than those who live near them. People over the age of 65 in Taichung, Taiwan, prefer

to spend time in parks close to their homes because of the importance of design, location, and upkeep.

Greenways may help communities improve their health by providing a safe, designated place for fitness activities, leisure, and walking and cycling (nonmotorized transportation) [10]. The health advantages of UGCs have been shown in several cross-sectional research studies. The causal link between UGC and inhabitants' physical and mental health, on the other hand, is yet unknown [11]. Using a natural experimental study approach, [12] investigated the influence of a large-scale UGW, the "102-kilometer-long East Lake Greenway in Wuhan, China," on inhabitants' health outcomes. For elderly individuals to retain an active lifestyle, urban green areas are crucial. UGWs not only offers a physical environment that encourages older persons to get involved in outdoor activities in their communities, but they also provide social capital and social cohesiveness. As a result, the accessibility of these sites promotes older individuals' engagement in outdoor activities [13]. They facilitate social interaction and community cohesion by providing a natural gathering spot for locals.

2.3. Environmental Preservation.
To minimize habitat fragmentation and manage the principles of natural ecosystems, UGCS are "multifunctional linear green spaces." Greenways provide a wide range of benefits, including flood control, soil stabilization, water purification, and more. Green infrastructure refers to linked networks of green areas that may help with water quality enhancement, stormwater management, and wildlife conservation, among other things [14]. In China, the emphasis on greenway building is changing from production conservation and aesthetics to ecological protection [15]. In order to encourage new secondary sectors, such as production and innovative materials manufacture, the municipal government needs a clean environment. Maanshan's emergence as a tourist destination, real estate market, and other service sectors demands an appealing landscape and a good quality of life, which necessitates innovative approaches to mobilize and strengthen Maanshan's natural benefits in China. To date, Maanshan's environmental competitiveness has been bolstered by the municipal government's long-term focus on environmental protection. The presence of the ecological green belt is utilized to efficiently regulate urban land development and fulfill ecological operations such as "biodiversity maintenance, enhancement of overall protection and use of environmental resources in rural areas, regulation of the microclimate in cities and pollution absorption" [16]. Recently, academics have focused more on the compound operations of a green belt surrounding a city, and they've looked into new green belt planning strategies to assure the long-term growth of an ecologically protective green belt that can contribute to today's cities [17]. Green infrastructure may take the shape of multifunctional ecological networks that are natural, semi-natural, or manmade. New forms of transportation, entertainment, leisure, environmental preservation, cultural conservation, as well as protection against natural calamities like wind, soil erosion, and floods, have historically been connected with greenways.

2.4. Problems in China's Greenway Construction and Sustainable Economic Development.
The greenway is associated with several benefits regarding economic, environmental, and individual aspects. But the implementation of an effective greenway in China is linked with certain challenges, which are explained below.

2.4.1. Legal Issues.
The rate of urbanization in China is increasing at a fast speed, but the building of UGCs has not kept pace with this rapid growth. As a primary factor, China lacks a legislative framework for UGCs, which is one of the main causes. Even while the People's Republic of China Urban and Rural Planning Law explicitly states that the green channel is an essential aspect of China's national greening, most of the relevant implementation depends on this law. As far as we know, there are no specific legislative requirements or indications that the UGC building is important for city planning and growth. In contrast, the relevance of UGC buildings has not been stressed in the design of urban infrastructure. To speed up the building of UGCs, a thorough legislative and regulatory framework is required.

2.4.2. Lack of Environmental Awareness.
China's economy rose at an annual rate of 9.7 percent after reform and opening up, propelling it to the position of the globe's second biggest economy [18]. Anyhow, China's heavy utilization of energy and the environment for economic expansion has resulted in a number of economic and social issues. Since China has become the world's top energy user, the World Bank has highlighted this fact. The reduction of CO_2 emissions in China, which accounts for 26.6 percent of the globe's total emissions, is very difficult. In 2018, China was placed 120th out of 180 nations in terms of environmental performance [19]. As of 2010, the price of environmental deterioration in China has climbed from 3% of GDP to 3.5% of GDP during the period from 2004 to 2010. To sum up, environmental deterioration in China is endangering the country's long-term growth. A lack of societal engagement in the greenway design might have a negative effect on the greenway's aims and priorities, notably its ecological functions [20, 21]. One of the most difficult aspects of greenway development is how to get local government agencies, community organisations, and other interested parties to participate in the planning and decision-making procedure. For the sake of environmental sustainability, it is critical to understand and change human behaviour.

To reduce the possible detrimental impact of technological growth on human and environmental health, it is required to properly address difficulties like rising energy usage, waste and greenhouse gas emissions, and the utilization of natural and nonrenewable raw resources. A greener

future is one that replaces current technology, the Internet of Things (IoT), and the economy with environmentally friendly options, each of which has the potential to increase human well-being dramatically and so contribute to a more sustainable smart world [22].

2.4.3. Planning and Management Issues.

Landscape variety is an underappreciated contributor to a greenway's visual appeal and ecological diversity. When it comes to greenway alignment design, site appropriateness is the most important factor. The "one-size-fits-all" approach to greenway planning actions has resulted in the composition and variability of greenways in Shenzhen, making multifunctional greenway development in urban areas difficult [23]. UGCs' diverse landscapes will need greater attention in the future, as will the tactics for improving them in light of particular policy objectives.

China's cities are rapidly expanding and becoming denser. In turn, this puts more strain on the city's open spaces, hindering the growth of urban green infrastructure. Existing green spaces in highly populated areas have significant challenges, which raise questions about ensuring that all inhabitants have equitable access to urban green spaces. In order to enhance greenway planning in other cities, it is important to record and reflect on successful greenway design practices in emerging cities [24].

In an ideal world, roads would be as ecologically friendly as possible while also having aesthetically pleasing features to reduce the amount of environmental harm they cause. Researchers, engineers, and artists must collaborate to reshape human environments by designing roads depending on more than just engineering concerns; future roads must take ecological concerns into account but also have aesthetic properties to meet the needs for environmental protection during road construction [25]. Artistic landscapes can only be achieved via collaboration between science and artists [26]. Another important issue in greenway planning is the lack of innovative, energy-efficient technologies. A massive transportation network (railway, highway, and so on) has been created as a result of this rapid economic development and growth, including environmentally sensitive locations like Tibet that need careful engineering to protect the environment.

2.4.4. Limited Concern on Visitor's Experience.

Tourists' perceptions regarding greenways and eco-friendly locations are critical in developing an effective greenway development strategy. Previous greenway research has largely focused on physical-ecological design difficulties, with little concern for visitor experience [27]. Only a tiny number of greenway scholars have looked at visitors' personal values and established the degree to which such values are helpful in understanding tourists' motivation and making responsible judgments. Residents who reside near greenways have gotten minimal consideration in this design [28]. Though studies have looked at users' attitudes and preferences for greenways, little focus has been devoted to the changes in perceptions between nonresidential and residential users.

Previous greenway research has viewed greenway users as single aspects, intercepting all greenway users, whether they are local residents, tourists, or people who go to the greenway from other regions of the city. While considering all greenway users and considering them as a single entity has offered important insights into the overall preferences of users and utilization patterns, it fails to distinguish between residential and nonresidential perceptions, resulting in broad management suggestions that may be in conflict with residential preferences [29]. Those who live near greenways are possibly the most significant stakeholders, as they must struggle wwith both the beneficial and negative consequences of greenways in their communities. Because local people have the most access to the greenway, which acts as a transportation and recreational corridor for their surrounding communities, their opinions, concerns, and preferences are possibly more significant than those of commuting users [30].

2.5. Countermeasures for Effective Greenway Economic Development.

The technological innovations in the computer Internet provide better solutions for greenway development and sustainable economic development. Some of the countermeasures for effective greenway economic development are discussed below.

2.5.1. Effective Policymaking.

Many municipal development plans tend to recognize that environmental protection is quickly becoming a must for economic success rather than a choice. Despite the rising physical and political constraints imposed by the environment, China's governments have started to address environmental problems as part of their objectives to keep their cities' economies growing. Urban economic and environmental governance, according to this perspective, calls for a holistic approach rather than a fragmented one. In order to align their goals with those of the environmental movement, China's development regimes are becoming increasingly interested in implementing environmental laws in their cities.

After the Guangdong Provincial Government created the "Pearl River Delta (PDR) greenway network" in 2010, China's contemporary greenway movement was born. The PDR greenway network was designed to reduce gaps between urban and rural areas in terms of economic, environmental, and social factors [1]. Greenways have been more popular in China, with 31 provinces planning or implementing provincial greenways and 163 cities working on municipal greenways by the end of 2016. According to the "Ministry of Housing and Urban-Rural Development," greenways have been officially acknowledged as a new national policy by the federal government.

Road development in China's Hubei Province's Shennongjia area since 2006 has seen a shift away from traditional methods, which were costly and harmful to the environment. This time, the goal is to build "aesthetic greenway systems" that safeguard both nature's ecosystem and society's values, resulting in environmentally and socially harmonious road networks. In China, this means that the

networks account for Chinese aesthetic principles while also being in harmony with the environment and society's economic growth.

2.5.2. Environmental Education Programs.

To promote ecological consumerism and proenvironmental behaviour, environmental education is essential. We must never lose sight of our basic goal in environmental education: spreading environmental awareness to the general public. As a result of these arguments, environmental information and education have been included as two of the most important determinants of environmental knowledge [31].

The livelihoods of conservation participants across the globe have been radically transformed, and other socio-economic processes have been impacted. In order to be successful, such conservation projects need to be well understood by the public. The Grain-for-green programme (GGP), the biggest conservation initiative in the world, has so far shown little proof of its impact on the lives of its members. Research by [32] examined whether the GGP programme achieves its purpose of protecting participants' lives while also improving their environmental awareness. Educational awareness programs, according to [33], consider the following aspects:

   (i) Must enhance awareness regarding environmental conservation

   (ii) To facilitate personal reflection on environmental actions

   (iii) Must support the application of environmentally sustainable behaviours

2.5.3. Greenway Planning Using Computer Internet Technology.

Tourists and the growth of service industries may be encouraged by efficient planning of greenways in rural areas. Figure 3 shows the Internet technologies used in greenway planning and greenway approaches. It is possible to identify and demonstrate UGC planning techniques by analyzing the recent UGC structure and relating physical aspects to theories on "landscape architecture and ecology" as well as technologies like geographic information system (GIS), remote sensing, and graph theory [34].

The Shenzhen greenway network's activity distribution is mapped using a GIS. In their research, they have shown that greenways with dense housing, mixed land use, sophisticated street network, and huge parks enhance physical activity. An area's natural circumstances and level of environmental degradation must be taken into consideration before designing any greenway network. Afforestation requirements may be determined based on this information. A novel computer modeling technique, a tree belt modeling system (TBM), is proposed to assess the suitability of a particular location for greenway services [35]. TBM is a predesign step in the creation of an ideal greenway structure that determines whether or not tree belt functions will be available in the proposed network.

Using data fusion and computer virtual reality (VR), the work in [36] undertakes a study on urban green landscape design to illustrate the potential of a more planned design. SketchUp software is used to create models, and a digital camera is used to take continuous digital photos. To begin, a panoramic picture is created by stitching together many individual digital landscape photographs. For final tweaks to materials and lighting, they upload the reduced version into our VR platform, where they can compare it to the original model's volume and do a plan check. It is important to note that the study in [5] analyses greenway branding as a sort of green resource branding with a specific focus on "media promotion, users' outdoor activity trajectories, and the picture content of geotagged images." Participation of outdoor users in greenway development may enhance their enjoyment of the greenway's benefits.

GIS was used in [37] in order to build a greenway. There were three different greenway alignment scenarios offered in [38], each based on the criteria of compatibility, variety, and integration in Wuchang, China. Incorporating land appropriateness and variety in the environment, they came up with the concept of a greenway potential index. Combining the greenway potential index and connectivity research, they found an ideal route for future greenways between their origin and destination that met both site suitability and landscape variety criteria. The study in [39] intended at determining the best way to build urban ecological landscapes using edge computing from the Internet of Things.

(1) Green Comple. In China, fast urban expansion has been reduced by the employment of several green-space planning techniques. For instance, Shanghai, Chengdu, and other megacities have created an ecological network of green spaces, while Beijing has established a greenbelt surrounding the city. However, because of the scarcity of available land and the low demand for natural space, encroachment and fast urban growth have persisted. Therefore, in the context of modern megacity growth, green space with a single purpose and single shape may not be the ideal option.

(2) Green Ecological Corridor. The green ecological corridor is a key component of the landscape's overall security and stability because it serves as a conduit for energy flows between various ecological source patches, a link between people and the environment, and a link between urban and rural regions. According to a review study, the role, connotation, denotation, and construction methods of the green ecological corridor have been significantly expanded and altered from the Emerald Necklace-style Park corridor system in the Boston area in the 1960s to the European Ecological Network in the early 21st century. The expansive Three-North Shelter Belt in dry areas is only one of several instances of green ecological corridors in China at various dimensions.

(3) Greenway Network. The most frequent human activity and the most powerful interference are connected to roads. Urbanization and an increase in vehicles have led to the creation of several broad major roadways in recent years. The road network's density has continuously risen, improving travel convenience and enhancing the connectivity
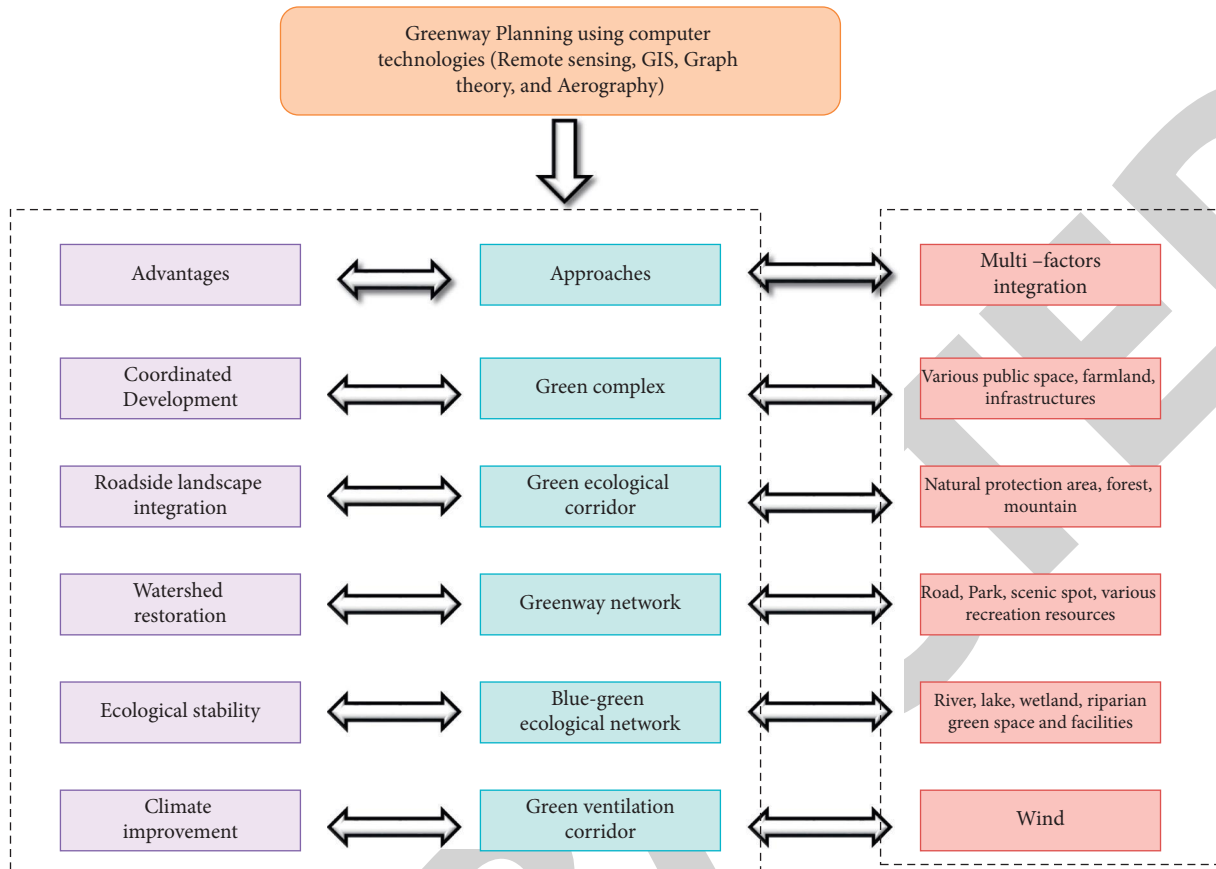
FIGURE 3: Computer Technology in greenway approaches.

between cities. However, forests and other natural environments have been fragmented as a result. The creation of a network of greenways, which connects roads with nearby greenery, can increase the amount of public space available for urban residents' leisure and recreational activities as well as the ecological areas available for wildlife to live and migrate. Greenways can also play a crucial role in a region's overall ecological security and the advancement of sustainability.

(4) Blue-Green Ecological Network. The river is a crucial component of the urban environment since it has flowing, linear, and natural elements. Rapid development has resulted in the artificialization of many natural rivers over the last several decades, as well as a significant loss of wetlands. Along with other problems, rivers have lost some of their natural characteristics and are no longer as effective at filtering water. This might be the root of Beijing's severe urban floods and Wuhan's severe waterlogging, among other major Chinese cities. The sustained growth of human settlements depends on river ecological restoration and watershed management. Restoring watersheds and urban-rural ecosystems may be accomplished by enhancing river connections and using watershed resources to create a blue-green ecosystem.

(5) Green Ventilation Corridor. Numerous case studies have been published on how green areas actively control the temperature via means like ventilation, cooling, and air purification. Green ventilation corridors may improve city ventilation, bring in fresh air, and reduce the heat island effect. There is a noticeable cooling and humidifying impact in the green ribbon area. Recent years have seen a boom in studies on the connection between urban climate (wind and temperature) and green space.

2.5.4. Significance of Internet Technology in Green Economic Development. Technology use that does not damage, overexploit, or deplete natural resources, reduces energy consumption, produces completely reused or reclaimed products, and reduces waste and emissions during production and use are the main goals of "green, environmental, or clean technology" [40]. In other words, green technology is characterized by environmentally sustainable design, manufacture, use, and disposal with little or considerable environmental effect. Figure 4 depicts the differences between conventional and green production industry.

Sustainability in economic development is a major issue, and this study examines how greening technology might help to improve the environment and economy while also enhancing human well-being and global modernity in the process [41]. According to research, Internet technologies have the potential to transform and bring numerous benefits (among them customer satisfaction, environmental
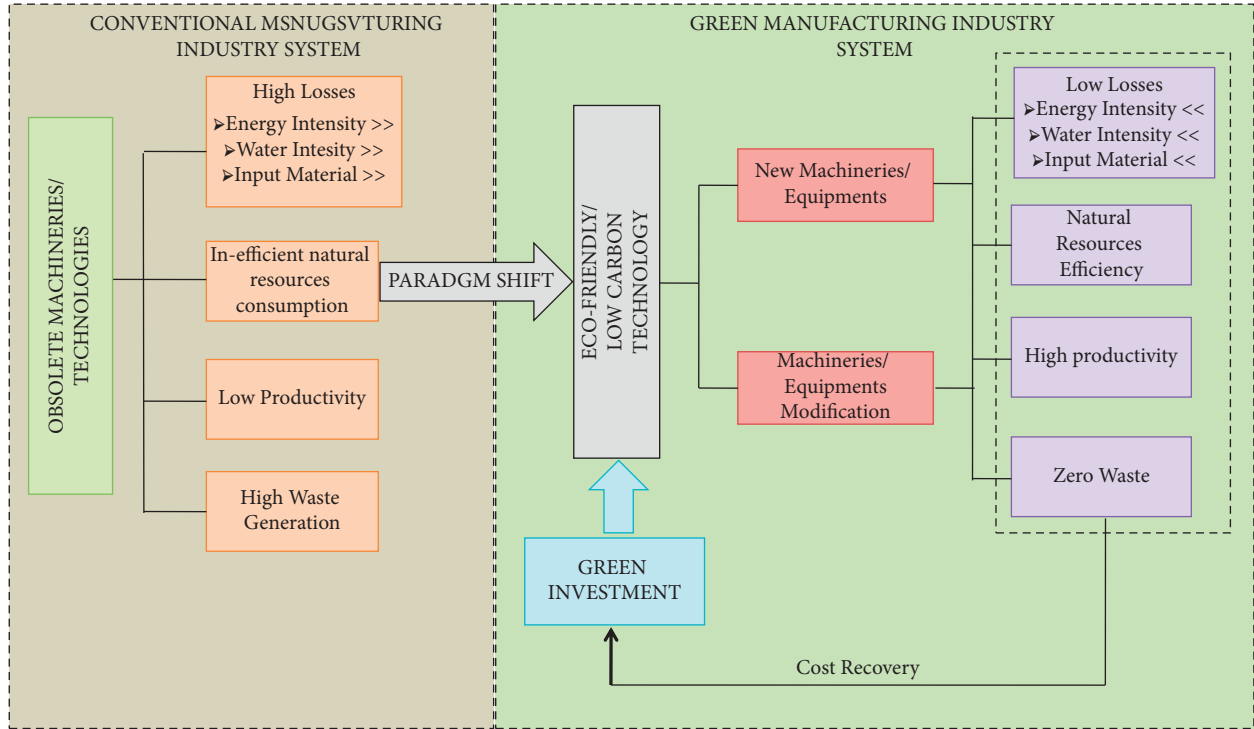
FIGURE 4: Green economic development.

protection, and increased profit) in a wide range of industries using the latest technology approaches and solutions, as well as eliminate or minimize the negative impact on human health and the environment. Information and computer technology (ICT) development has a favourable influence on boosting green total factor energy efficiency (GTFEE) and this role is controlled by environmental restrictions [42]. GTFEE is influenced by ICT development in various ways depending on the intensity of environmental control. In addition, the influence of ICT development on GTFEE varies greatly from area to area. It is possible that ICT infrastructure might play a big role in reducing carbon emissions from companies in the industrial sector. Industries may reduce their use of resources by using ICTs [43]. The study in [44] stated that the rise in ICT cuts carbon dioxide emissions in China.

3. Statistical Analysis

In statistical analysis, we use chi-square test and ANOVA test.

3.1. Chi-Square Test.
The chi-square analysis is one of the most efficient statistical tools for testing a hypothesis when the factors are minimal. Unlike the other statistics, the chi-square may provide specific information on not just the importance of any significant variability but also which groups are responsible for those distinctions.

$$\sum y_{j-k}^2 = \frac{(P - F)^2}{F}. \tag{1}$$

Here, P = present point, F = real point, Y^2 = Chi-square value, and $\sum Y^2$ = To total all of the cell chi-square values, whether they had been treated or not, the following formula is used to calculate expected chi-square values:

$$F = \frac{N_S \times N_D}{o}, \tag{2}$$

where F = reflects the work value of the unit, N_D = denotes that cell nucleus row edge, N_S = denotes that cell's row edge, and o = reflects the sample group as a whole.

The sample size is split by the product of the row marginal and the column marginal for each cell.

$$y^2 = \frac{(P - F)^2}{F}. \tag{3}$$

Correlation measures are statistical assessments of the strength of a relationship. The Cramer's V test is the most often utilized chi-square strength test. Using the formula, it is easy to calculate

$$\sqrt{\frac{y^2/o}{(l - 1)}} = \sqrt{\frac{y^2}{o(l - 1)}}. \tag{4}$$

The chi-square is a useful data analysis tool that reveals a lot about the nature of research data.

3.2. ANOVA.
Analysis of variance, or ANOVA, is a quantitative tool for dividing reported variability data into multiple parts for use in subsequent tests. A one-way

ANOVA has been used to discover the relationship between variables when there are 3 or more data sets. The classic ANOVA F-statistic is the proportion of average sums squared of the null model with an anthropic principle to the whole model. The parameters are calculated using the least-squares approach, with all variances being identical. This may be expressed as

$$S = \frac{MH_{\text{between}}}{MH_{\text{error}}}, \tag{5}$$

where

$$MH_{\text{between}} = \frac{\sum_{i=1}^{k} l(\overline{x} - \overline{x})^2}{n - 1}. \tag{6}$$

Also,

$$MH_{\text{error}} = \frac{\sum_{i=1}^{k} \sum_{j=1}^{l_i} (x_{ij} - x_i)^2}{l - k}. \tag{7}$$

The Welch-test-statistic is defined as

$$Y = \frac{\sum_{i=1}^{K} y_j \left[(x_i - \widetilde{x})^2 / (K - 1) \right]}{1 + 2(K - 2)/K^2 - 1 \sum_{i=1}^{K} \left[\left(1 - y_j/u \right)^2 / (l_i - 1) \right]}, \tag{8}$$

where $l_i = n/t_1^2$, $u = \sum_{i=1}^{K} l_i$, and $Y = 1/u \sum_{i=1}^{K} l_i y_j$; we have

$$S = \frac{k^2 - 1}{s \sum_{i=1}^{K} \left[\left(1 - y_j/u \right)^2 / (l_i - 1) \right]}. \tag{9}$$

The Brown–Forsythe-test-statistic is defined as

$$S^* = \frac{\sum_{i=1}^{K} o\left(\overline{x}_j - \overline{x} \right)^2}{\sum_{i=1}^{K} \left(1 - l_i/H_i^2 \right)}. \tag{10}$$

When L_o is factual, the allocation of S^* is appropriate by a central S distribution with degrees of freedom $K - 1$ and s, where s is defined as

$$\frac{1}{s} = \frac{\sum_{i=1}^{K} c_i^2}{(l_i - 1)}, \quad c_i = \frac{\left(1 - l_i/H_i^2 \right)}{\sum_{j=1}^{h} \left(1 - l_i/H_i^2 \right)}. \tag{11}$$

To calculate the generalized p value, the generalized p value is now computed as $p = 1 - r$, where s is the sample size.

$$s = A\left(J_{k-1 l - k} \left(\frac{l - k}{k - 1} \widetilde{t}_b \left(\frac{l_1 t_1^2}{B_1 B_2, \ldots, B_{K-1}}, \frac{l_2 t_2^2}{B_1 B_2, \ldots, B_{K-1}}, \frac{l_3 t_3^2}{(1 - B_2)B_3, \ldots, B_{K-1}}, \cdots \frac{l_1 t_K^2}{(1 - B_{K-1})} \right) \right) \right). \tag{12}$$

The prediction is calculated with regard to separate Beta stochastic process in an F-distribution having $h - 1$, $V - h$ dof.

$$B_K \sim \text{Beta}\left(\sum_{i=1}^{k} \frac{(l_i - 1)}{2}, \frac{(l_{k+1} - 1)}{2} \right), \quad l = 1, 2, \ldots, k - 1. \tag{13}$$

The p value is calculated by numerically integrating the anticipated value in the p value formula with regard to the Beta random variables.

4. Conclusion

China's recent growth has been the quickest among the world's major economies. Many nations have been able to change from a development mode focused only on economic expansion to one that prioritizes resource conservation and environmental preservation via green economic growth. An urban green area provides a secure and healthy atmosphere that is ideal for physical and recreational activities as well as for social contact. This paper presents a comprehensive review of problems and solutions for greenway economic development in China. The survey clearly emphasized the advantages of greenways for socioeconomic, individual, and environmental development. In addition, the challenges in greenway development like legal, planning, and environmental awareness issues are explained in detail. Finally, the importance of Internet technologies in greenway planning and green economic development is illustrated. Hence, this survey provides detailed insights into greenway development for greenway researchers and designers.

4.1. Future Work and Recommendation.
The findings point to ramifications for the future design and development of greenways, with a particular emphasis on safeguarding the rights of private landowners in addition to the public's right to a healthy environment. Directions for future study should centre on investigating the ways in which information-education initiatives have the potential to greatly raise community awareness and understanding of the advantages of greenway corridors. The findings of the study point to the need to do more research into a number of topics, each of which has the potential to provide vital information that will be relevant to the future design and development of rural greenways. It is recommended that a continuing and in-depth investigation and analysis of the differences in people's perceptions of streamland be carried out, with the comparison being made between rural homeowners of agricultural land, rural homeowners of nonagricultural land, suburban homeowners, and urban residents. Pursuing research that aims to lead to the creation of information-education programmes that prioritise environmental quality, aesthetics, and cultural-historical preservation should be

a priority. These programmes should enhance community and landowner understanding of the advantages that streamlands provide.

Data Availability

The data used to support the findings of this study are available from the author upon request.

Conflicts of Interest

The author declares no conflicts of interest.

Acknowledgments

References

[1] Z. Liu, Y. Lin, B. De Meulder, and S. Wang, "Can greenways perform as a new planning strategy in the Pearl River Delta, China?" *Landscape and Urban Planning*, vol. 187, pp. 81–95, 2019.

[2] M. Hu and B. Xia, "A significant increase in the normalized difference vegetation index during the rapid economic development in the Pearl River Delta of China," *Land Degradation & Development*, vol. 30, no. 4, pp. 359–370, 2019.

[3] M. Li, H. Sun, F. O. Agyeman, M. Heydari, A. Jameel, and H. Salah ud din Khan, "Analysis of potential factors influencing China's regional sustainable economic growth," *Applied Sciences*, vol. 11, no. 22, Article ID 10832, 2021.

[4] W. Ahmad, W. G. Kim, Z. Anwer, and W. Zhuang, "Schwartz personal values, theory of planned behavior and environmental consciousness: how tourists' visiting intentions towards eco-friendly destinations are shaped?" *Journal of Business Research*, vol. 110, pp. 228–236, 2020.

[5] W. Zhao, Y. Wang, D. Chen, L. Wang, and X. Tang, "Exploring the influencing factors of the recreational utilization and evaluation of urban ecological protection green belts for urban renewal: a case study in Shanghai," *International Journal of Environmental Research and Public Health*, vol. 18, no. 19, Article ID 10244, 2021.

[6] P. J. Chang, C. W. Tsou, and Y. S. Li, "Urban-greenway factors' influence on older adults' psychological well-being: A case study of Taichung, Taiwan," *Urban Forestry & Urban Greening*, vol. 49, Article ID 126606, 2020.

[7] R. N. Mody and A. R. Bhoosreddy, "Multiple odontogenic keratocysts: a case report," *Annals of Dentistry*, vol. 54, no. 1-2, pp. 41–43, 1995.

[8] B. Han, D. Li, and P. J. Chang, "The effect of place attachment and greenway attributes on well-being among older adults in Taiwan," *Urban Forestry and Urban Greening*, vol. 65, Article ID 127306, 2021.

[9] M. Jabbar, M. M. Yusoff, and A. Shafie, "Assessing the Role of Urban green Spaces for Human Well-Being: A Systematic Review," *GeoJournal*, vol. 12, pp. 1–19, 2021.

[10] X. Liu, Z. Zhu, L. Jin, L. Wang, and C. Huang, "Measuring patterns and mechanism of greenway use–A case from Guangzhou, China," *Urban Forestry & Urban Greening*, vol. 34, pp. 55–63, 2018.

[11] H.. Garg, "Digital twin technology: revolutionary to improve personalized healthcare," *Science Progress and Research (SPR)*, vol. 1, p. 1, 2020.

[12] B. Xie, Y. Lu, and Y. Zheng, "Casual evaluation of the effects of a large-scale greenway intervention on physical and mental health: A natural experimental study in China," *Urban Forestry & Urban Greening*, vol. 67, Article ID 127419, 2022.

[13] P. J. Chang, "Effects of the built and social features of urban greenways on the outdoor activity of older adults," *Landscape and Urban Planning*, vol. 204, Article ID 103929, 2020.

[14] S. Y. Pan, M. Gao, H. Kim, K. J. Shah, S. L. Pei, and P. C. Chiang, "Advances and challenges in sustainable tourism toward a green economy," *Science of the Total Environment*, vol. 635, pp. 452–469, 2018.

[15] Y. Wang, M. Duan, Q. Wang, J. Xie, and C. Yang, "Ecological Health Care Effect of Healthy Greenways in Urban Community belt parks," *Landscape and Ecological Engineering*, vol. 3, pp. 1–16, 2022.

[16] B. Ahmed and A. Ali, "Usage of traditional Chinese medicine, western medicine and integrated Chinese-western medicine for the treatment of allergic rhinitis," *Science Progress and Research*, vol. 1, pp. 1–9, 2020.

[17] M. Zhao, H. Lu, J. Liang, and C. S. Chan, "Evaluating green resource branding using user-generated content data: The case study of a greenway in eastern Guangzhou, China," *Urban Forestry & Urban Greening*, vol. 66, Article ID 127395, 2021.

[18] F. Han, R. Xie, and J. Fang, "Urban agglomeration economies and industrial energy efficiency," *Energy*, vol. 162, pp. 45–59, 2018.

[19] A. Shahabaz and M. Afzal, "Implementation of high dose rate brachytherapy in cancer treatment," *SPR*, vol. 1, no. 3, pp. 77–106, 2021.

[20] Z.. Li, *Treatment and Technology of Domestic Sewage for Improvement of Rural Environment in China-Jiangsu: A Research*, China, 2022.

[21] S. O. Salihu and I.. Zayyanu, "Assessment of Physicochemical parameters and Organochlorine pesticide residues in selected vegetable farmlands soil in Zamfara State, Nigeria," *Science Progress and Research (SPR)*, vol. 2, p. 2, 2022.

[22] M. Maksimovic, "Greening the future: green Internet of Things (G-IoT) as a key technological enabler of sustainable development," in *Internet of Things and Big Data Analytics toward Next-Generation Intelligence*, pp. 283–313, Springer, Cham, 2018.

[23] Z. Liu, Y. Lin, B. De Meulder, and S. Wang, "Heterogeneous landscapes of urban greenways in Shenzhen: t," *Urban Forestry and Urban Greening*, vol. 55, Article ID 126785, 2020.

[24] I. Kowarik, "The "Green Belt Berlin": establishing a greenway where the Berlin Wall once stood by integrating ecological, social and cultural approaches," *Landscape and Urban Planning*, vol. 184, pp. 12–22, 2019.

[25] B. Lin and Y. Zhou, "Measuring the green economic growth in China: influencing factors and policy perspectives," *Energy*, vol. 241, Article ID 122518, 2022.

[26] B. Cheng, Y. Lv, Y. Zhan, D. Su, and S. Cao, "Constructing China's roads as works of art: a case study of "esthetic greenway" construction in the Shennongjia region of China," *Land Degradation & Development*, vol. 26, no. 4, pp. 324–330, 2015.

[27] S. Zhang, W. P. Stewart, and E. S. Chan, "Place-Making upon Return Home: Influence of Greenway Experiences," *Leisure Sciences*, vol. 1, 25 pages, 2020.

Research Article

Location Mechanism and Perception Scheme of Remote Transmission Equipment Layout in Smart Water

Shikai Xing (iD) **and Xiang Xie** (iD)

School of Economics and Management, Beijing Jiaotong University, Beijing, China

Correspondence should be addressed to Xiang Xie; xxie@bjtu.edu.cn

Whether the traditional water equipment is equipped with remote transmission unit depends on whether the power consumption of the installed equipment can meet the expected life cycle. In the paper, by using intelligent strategies to save the power consumption of sending data, more sites can be selected to install remote transmitting units. In the process of transmitting data packet sequence, the remote transmission device needs to perceive the environment, interact with the environment and make decisions, and adjust the strategy according to the effect of the decision actions. Therefore, in this paper, the transmission process is modeled as Markov sequence decision model. And the real time signal interference noise ratio of the channel and the transmission delay of the data packet are defined as the state space. The decision action space consists of immediate transmission and delayed one, and the minimum total power consumption is taken as the objective function. The model is solved by Proximal Policy Optimization (PPO) algorithm, and the optimal decision sequence of site selection threshold is obtained.

1. Introduction

Different from solving multi-device game problems through reinforcement learning, solving single-device decision-making problems through reinforcement learning can be modeled as a Markov decision process [1]. When the state transition probability of a single device is known, the iterative equation of dynamic programming can be used to solve its decision-making problem. Due to the wireless environment of the installation point of water equipment, its state transition probability is often unknown, and reinforcement learning algorithm can explore the environment through trial and error, and obtain a good strategy to maximize the overall expected return. In this paper, a deep reinforcement learning algorithm will be used to solve the Markov sequence decision problem of a single intelligent device.

2. Literature Review

In recent years, wireless communication technology is developing rapidly, especially the Internet of Things communication technology represented by NB-IoT (Narrowband Internet of Things) [2], which has the characteristics of deep coverage, low power consumption, and low remote maintenance cost, which is very suitable for the water industry, the deployment location, data transmission, battery power supply, and other characteristics of the remote transmission equipment. With the full coverage of the NB-IoT network, a large number of IoT devices have been installed or replaced in the urban water supply and drainage network and copy collection system. The number of remote transmission equipment such as water meters and manhole cover monitors installed with NB-IoT modules in urban pipeline networks and copy collection systems is also rapidly increasing.

These remote transmission devices have intelligent features such as on-demand data reporting, intelligent sleep and access to the platform, and online detection with the platform. These intelligent devices are mainly composed of three modules, namely, data acquisition and calculation module, compression and storage module, and data transmission module. These modules are located in the application layer, storage layer, and communication layer of the smart device, respectively.

According to the agreement between the device and the platform [3], the device collects data at a fixed frequency every day, and the amount of data accumulated in each time period is basically fixed. The sensors of the device collect data according to certain rules and frequencies. The basic data is the measurement data of the user's water flow. In addition, there may be additional data such as water pressure and water temperature. The collected data is stored in the remote transmission device. Metering and sensing accumulated data to the platform in the form of incremental or absolute amount through the communication unit on the equipment at a predetermined time every day. The data accumulated during upload will be split into multiple packets and sent.

After the data is sent on the same day, in order to save power consumption and prolong the working life, the remote transmission device will power off itself to sleep or enter the PSM (power save mode) state after sending data, and wake up on the next day or when data is sent, the unsent data will continue to be sent the next day.

Most of the equipment of intelligent water affairs is installed in the weak coverage area of wireless channels such as corridors and wells. On the coverage benchmark, there will also be various frequency shifts, random disturbances generated by interference, and superposition of noises, such as white noise, occlusion, and co-frequency interference. The result of superposition makes the channel quality change randomly.

Therefore, the channel quality of wireless communication is jointly determined by signal coverage and noise interference. Generally, two quantities are used to evaluate the channel quality: RSRP (Reference Signal Receiving Power) and SINR (Signal to Interference plus Noise Ratio).

3. Problem definition

Most of the equipment in the urban pipe network system and the copying system are deployed in underground pipelines or corridors. The remote transmission equipment is battery powered and power consumption sensitive. Due to building block reflection or stray interference, the wireless signal will be greatly attenuated, and the battery life is very dependent on the installation environment and data transmission strategy. If the installation location and transmission strategy are not good, some batteries will be consumed quickly from the start of the remote transmission equipment to the use of the battery, which cannot support the data collection and transmission business needs of a period of 6–10 years. Since the installed equipment is unattended, when the equipment needs to replace the battery, it generally needs to be manually operated on-site. When the battery is manually replaced on-site or the traditional manual reading is used, it is often difficult to enter the home and the equipment maintenance cost is relatively high.

Due to the relatively high labor cost of replacing batteries for these devices, from the perspective of investment and cost, it is expected that these devices can maintain a longer life cycle and send more data, thus resulting in the problem of installation and location selection of remote transmission devices.

The installation site selection of the remote transmission equipment depends on two aspects [4]: the wireless signal coverage quality of the installation point, and the ability of the equipment to perceive the installation environment.

Before remote transmission equipment is installed, it is necessary to measure the wireless signal coverage quality at the installation site to ensure compliance with a certain threshold. The determination of the threshold is related to the ability of the device to perceive the installation environment. That is, the stronger the device's ability to perceive the installation environment, the lower the requirements for the quality of wireless signal coverage; on the contrary, the weaker the device's ability to perceive the installation environment, the higher the requirements for the quality of wireless signal coverage. By improving the device's ability to perceive the environment, the requirements of the device on the quality of on-site signal coverage at the installation point are reduced, so that more installation points have the conditions to install remote transmission equipment.

When the channel state changes, the SNR fluctuates. It will affect the power (power consumption) consumed by the remote device when sending data. When the signal gets worse, it takes longer to send data and requires more power consumption. The relationship between power consumption and signal quality presents a nonlinear relationship, that is, when the signal quality gradually deteriorates, the power consumption gradually increases. When the signal quality falls below a certain threshold, the power consumption starts to increase rapidly.

Based on the above-mentioned nonlinear relationship between channel quality and power consumption, there is no clear judgment value, and it is necessary to determine the best action value that should be taken according to the environment and a certain strategy. That is, the remote transmission device detects the channel quality of the wireless communication through the sensor to determine whether to send the data immediately or temporarily. Under the condition of a certain battery capacity of the remote transmission equipment, after all data of the day is sent correctly, the power consumption is the smallest, so that the service life of the remote transmission equipment is the longest. Therefore, when the life cycle of the equipment is 6–10 years, it is necessary to determine the location range of whether the remote transmission unit can be installed according to the amount of data to be sent.

Therefore, the installation site selection and perception improvement of remote transmission equipment is essentially an optimization problem for smart equipment to adopt a certain decision-making mechanism in an unknown environment to minimize the power consumption of the equipment.

4. System Model

In this paper, the environment-aware decision-making process of data packets sent by remote equipment is modeled as MDP [5] (Markov decision process) to describe the state transition of the system. And the derivation of the process is

equivalent to the layout and location selection mechanism of intelligent water remote transmission equipment.

Remote device is divided into n data packets each time, the data packet set $T_i = \{t_1, t_2, \ldots, t_n\}$, is a sample of task distribution, and each task sample corresponds to an MDP process. This MDP process is defined by a quintuple $\{S, A, P, R, \gamma\}$, where S is the state space, A is the action space, P is the state transition matrix, and R is the state transition benefit, which γ is the forward benefit discount.

4.1. Channel Status.

Remote transmission device sends each data packet in turn. According to the agreement, the amount of data sent is fixed. However, due to the random characteristics of wireless channels, the channel quality will vary randomly. Some packets need to be resent if poor channel quality results in higher BER or packet loss.

The signal-to-noise ratio (SINR) of the channel obeys a small-scale Rayleigh distribution, and its probability density function [6] is expressed as

$$p(snr) = \frac{1}{\overline{snr}} \exp\left(-\frac{snr}{\overline{snr}}\right). \tag{1}$$

Among them, snr the threshold value set of SINR is $\overline{snr}$ expressed, and the expectation of SINR is expressed. The NB-IoT network snr divides the channel quality into N grades according to the value, and sets a total of N -1 signal quality thresholds, $snr = \{snr_1, snr_2, \ldots, snr_{N-1}\}$. According to the snr threshold, the channel quality can be divided into N states, $C = \{c_0, c_1, \ldots c_{N-1}\}$. The channel state probability is

$$p_c(c_i) = \int_{snr_i}^{snr_{i+1}} p(snr) d(snr). \tag{2}$$

Assuming that the next moment of the channel can only be transferred to the adjacent interval, the state transition probability of the channel [6] is

$$p_c(c_i, c_{i+1}) = \frac{N(snr_{i+1}) * T_f}{p_c(c_i)}, \quad i \in \{0, 1, \ldots, N-2\},$$

$$p_c(c_i, c_{i-1}) = \frac{N(snr_i) * T_f}{p_c(c_i)}, \quad i \in \{1, 2, \ldots, N-1\}. \tag{3}$$

Among them, $N(snr_i) = \sqrt{2 * \pi * snr_i/\overline{snr}}$, *, $f_D *$ $\exp(-snr/\overline{snr})$, f_D is the maximum Doppler frequency shift.

4.2. Device Latency Status.

Device latency status of the device is described by the total delay after the device sends the i-th data packet, and the total delay corresponds to the power consumption. Suppose the sending time of the first data packet is, and the sending time of t_0 the i-th data packet is t_i. Obviously, the total delay after the device transmits the i-th data packet is not only *related* to the immediate decision, but

also depends on the delay accumulation of the previous $i-1$ data packets. The solution of the total transmission delay of the i-th packet can be transformed into a combination of immediate decision-making and solution of the total delay of the i-1-th packet.

$$\tau_i = \tau_{i-1} + \Delta\tau_i + T_f, \tag{4}$$

where the device latency status is $\tau = \{\tau_1, \ldots, \tau_n\}$, $\Delta\tau_i$ denote the waiting time determined by the immediate decision is related to the channel state and τ_{i-1} depends on the state of the previous i -1 data packets, T_f is a constant which is the transmission time of one data packet.

4.3. System Status.

System state S can be defined as the combined state of the channel state and the device delay state, i.e., $S \triangleq C \otimes \tau$. The above equation is a recursive process and cannot be solved by dynamic programming because the transition probability is unknown.

4.4. Action.

A denotes the remote transmission device taken when the new data packet is to be transmitted and it can be expressed as $A \triangleq \{0, 1\}$: When $a = 0$, it means to transmit immediately; when $a = 1$, it means to suspend transmission.

5. MDP Process Analysis of Remote Equipment

5.1. System State Transition.

When the remote transmission device sends data packets in sequence in a queuing manner, according to the channel state, the remote transmission device may choose to send the data packets immediately. This strategy works in the connected state, corresponding to the instantaneous SINR, and uses the transmit power to transmit with a certain probability. And transfer to the new channel state and device delay state.

The remote device may also adopt a policy of suspending sending a packet. This strategy waits in an idle state, expecting a better channel state to transmit, and the smart device $\Delta\tau_i$ continues to transmit after an interval according to the change of the channel state. After waiting for transmission, the system state transitions to the new channel state and in-device delay state. Since each data packet is small, the time to transmit a data packet is very short compared to the waiting time, which is defined as one frame T_f.

$\Delta\tau_i$ set as the single-step delay time of the i -th data packet, which is randomly determined $\Delta\tau_i$ by the intelligent device according to the environmental state. It is the pause waiting time. This waiting will cause the overall delay of the time for the device to complete the transmission of all data packets $\Delta\tau_i$. τ_i is the total delay of sending the first i data packets. t_i In order to send the time to be transmitted for the i - th data packet, the delay before the i-1 data packet $t_{i-1} - t_0$ is to be transmitted is

$$\tau_{i-1} = \begin{cases} t_{i-1} - t_0 + T_f, & \text{sentting the } (i-1)\text{th Packets immediately,} \\ t_{i-1} - t_0 + \Delta\tau_{i-1} + T_f, & \text{suspending the } (i-1)\text{th packet,} \end{cases} \tag{5}$$

the ith data packet is decided, the recursive formula of the delay state of the device is

$$\tau_i = \begin{cases} \tau_{i-1} + T_f, & \text{sentting the ith Packets immediately,} \\ \tau_{i-1} + \Delta\tau_i + T_f, & \text{suspending the ith packet.} \end{cases} \tag{6}$$

Set $\tau_0 = 0$. Then, the one-step delay of the i-th packet is

$$\Delta\tau_i = \left\{ \text{int}\left[\frac{N - snr_i}{(snr_N - snr_1/N)} \right] \right\} * \delta. \tag{7}$$

$SINR$ is related to the delay. The better the $SINR$, the smaller the delay. δ In order to suspend the parameter of the minimum time segment, this paper assumes that the channel quality remains unchanged in the minimum time segment to ensure that the retransmission is an incoherent channel, and N is the number of states. And int is a rounding operation function.

5.2. System Benefits and Costs.

After the device sends a data packet, if it does not receive the confirmation message from the platform within a certain period of time, it will resend the data packet. Due to the different acknowledgment mechanisms of different platforms, the additional power consumed by the device to resend the data packet is different. Therefore, an equivalent model is established: in different environments, the device increases the transmit power to ensure the same bit error rate during the transmission and transmission of data packets.

NB-IoT network adopts two rates to transmit data according to the $SINR$, corresponding to BPSK modulation (Binary Phase Shift Keying) and Q PSK modulation (Quadrature Phase Shift Keying). When the channel quality is relatively poor, the NB - IoT network uses the BPSK modulation method to obtain the minimum transmit power consumption when the bit error rate requirements are met under different channel states [7],

$$P_{ber}(s_i) \le 0.5 * erfc\left(\sqrt{\frac{snr_i * P_i')}{\sigma}} \right), \tag{8}$$

where $P_{ber}(s_i) = ber_i/\sigma$ is the bit error rate probability and $erfc(x) = 2/\sqrt{\pi} * \int_0^x e^{-\eta^2} d\eta$ is the Gaussian error function.

Ber is the allowable bit error rate of the wireless channel, σ is interference noise power. It is the P_i' equivalent transmit power that guarantees no packet loss or retransmission when the bit error rate is met. When the channel quality is relatively good, the NB-IoT network uses the QPSK modulation method to obtain the minimum transmission power consumption when the bit error rate requirements are met under different channel states, namely,

$$P_{ber}(s_i) \le 0.2 * exp\left(\frac{-1.6 * snr_i * P_i'}{3 * \sigma} \right). \tag{9}$$

When the device suspends sending the first i data packet, the power consumption is $I_0 * \Delta\tau_i + P_i'/v * T_f$, where I_0 is the idle state working current of the device, and when the device is in the waiting state, the idle state current of the I_0 device within $\Delta\tau_i$ the time is. v is the device voltage and T_f is the length of time in the connected state to transmit a data packet.

When the device sends immediately, $\Delta\tau_i = 0$, the power consumption of a single packet is

$$W_i = \Delta\tau_i * I_0 + \frac{P_i'}{v} * T_f. \tag{10}$$

After sending each data packet of all samples, the total power consumption of one day can be expressed as

$$W_{day} = \sum_{m}^{M} \sum_{i=1}^{n} \left(\Delta\tau_i * I_0 + \frac{P_i'}{v} * T_f \right). \tag{11}$$

Among them, M is the number of samples sent in one day, each sample is the data to be sent once agreed in the protocol, and n is the number of data packets corresponding to one sample. The goal of the system is to find the most efficient decision to send packets for maximum benefit, i.e., minimum total power consumption.

Defined t_i immediate reward function is the negative value of the power consumption of a single data packet, that is, $R_i = -W_i$, the immediate reward function of the amount of data sent per day is the negative value of the total power consumption, i.e., $R_{day} = -W_{day}$.

5.3. Location Range of Equipment.

The expected life days that the remote device can use is d, and the battery capacity is set to be W_{total}, then $d = W_{total}/W_{day}$, the expected life years $D = d/365$.

If the remote device only has data to be sent once a day, the location decision can be calculated by the following formula

$$W_{total} = \sum_{j=0}^{d} \sum_{i=1}^{n} \left(\Delta\tau_i * I_0 + \frac{P_i'}{v} * \tau_0 \right). \tag{12}$$

Since the battery capacity of the device W_{total} is a known value, when the expected number of days or years of use is given, an optimal environment-aware decision chain needs to be obtained to obtain the optimal $\Delta\tau_i$ sum P_i' scheduling.

Under the condition that the battery capacity, expected service life, and data volume of the device are all determined, whether a remote transmission device can be installed at an installation point depends entirely on $\Delta\tau_i$ and P_i' there is a

feasible solution. In order to delay the service life of the equipment, it is also necessary to seek the optimal strategy A to obtain the optimal solution.

According to the data collection and transmission rules of the device and the platform, each time data is sent every day during actual operation, it will be divided into n data packets and sent in a sequence. During the sample learning period, data can be sent continuously for M times, parameter training can be performed, battery power consumption can be measured, and it can be converted into whether the battery life meets the expected number of years.

6. Optimal Decision-Making Based on Deep Reinforcement Learning

For the above MDP problem, this paper will use deep reinforcement learning algorithm to solve. In the random and aligned MDP process, the state s and the decision action a have randomness and can $\pi(a|s)$ be represented by a conditional distribution, representing π the mapping from the state to the decision action a, that is, the policy. The following defines a class of neural networks to fit policies.

6.1. Sequence Representation of Neural Networks. The smart device divides the data sent each time into n data packet sequences to send, and needs to make a sending decision on the n data packets in turn to form a decision sequence. This paper uses seq2seq (sequence-to-sequence) [8] neural network to express the above process sequence, that is, input a sequence and output a sequence. The network structure is shown as Figure 1.

Seq2Seq neural network consists of an encoder and a decoder, both of which are R NN networks (Recurrent Neural Network, recurrent neural network). Assuming that the parameters of the neural network are θ, then when the state s is input, the conditional probability of outputting the optimal decision a can be rewritten $\pi_\theta(a|s)$. According to the input t_i, the neural network first encodes the learning and memory through the encoder, and then outputs through the decoder according to the memory of the network d_j, and then passes through two different activation functions, corresponding to the output state value function $v(s)$ and decision sequence probability $\pi_\theta(a|s)$.

In order to reflect the characteristics of the data to be sent, the sequence of data packets can be converted into a sequence of embedded vectors and input to the neural network. Input $t_i =$ [data packet sequence number i, data packet size], which is a 2-dimensional vector. For different types of devices, the amount of data to be sent each time may be different, so the size of the data packet is different. The size of the last data packet of each device is also different from the previous $n-1$ data packets, which is the margin of the data amount divided by $n-1$.

Denote the encoder and decoder [9] as f_{dec}, respectively, f_{enc}, then.

The output of the encoding part is

$$e_i = f_{enc}(t_i, e_{i-1}). \tag{13}$$

The output of the decoding part is

$$d_j = f_{dec}(z_j, s_{j-1}, a_{j-1}). \tag{14}$$

The input of the decoder consists of 3 parts [10], including the weighted sum of the output of the encoder z_j, and the decision execution result of the previous step s_{j-1}, a_{j-1}. z_j is the context of the decoder in step j, which contains the attributes of the data to be sent.

Seq2seq neural network is an n-dimensional vector d. After nonlinear activation of this vector, an n-dimensional probability vector π_θ and a value function v, (s_i), respectively, are obtained. The probability vector corresponds to π_θ the probability that the decision action takes a certain a, and the sum is 1. Then, the decision action of the $a_j = \text{argmax}_a(\pi_\theta)jth$ step can be obtained through the greedy algorithm.

6.2. Parameter Update and Optimal Decision. To update the neural network parameters, this paper adopts the PPO2 [11] algorithm (Proximal Policy Optimization, proximal policy optimization). The remote transmission equipment may send 1 or more samples per day, and the daily objective function can be defined as

$$J = \text{ER}_{\text{day}}. \tag{15}$$

Based on the sequence representation of neural network, the objective function J for each sample (θ_m) can be rewritten as

$$J(\theta_m) = E_\theta(R_{\text{day}}(\theta_m)). \tag{16}$$

According to the law of large numbers,

$$J(\theta_m) \approx \sum_{m=1}^{M} (R(\theta_m)) = E(-W_{\text{day}}), \tag{17}$$

where M is the number of samples, the data volume of each sample is divided into n data packet sequences and sent, m is the sample sequence number, that is, the data packet sequence. $R(\theta_m)$ is the power consumption of each sample, namely,

$$R(\theta_m) = \sum_{i=1}^{n} \min(\omega_i * A_i, \text{clip}_{1-\Delta}^{1+\Delta}(\omega_i) * A_i), \tag{18}$$

Δ is the clipping constant, and the trend function [12] is

$$A_i = \sum_{j=0}^{n-i+1} (\gamma * \omega_i)^j * (r_{i+j} + \gamma * v_\pi(s_{i+j+1}) - v_\pi(s_{i+j})). \tag{19}$$

Among them, γ is the advantage function discount coefficient, r is the immediate return, and v_π is the state value function. ω are the adjustment coefficients for variance and bias,

$$\omega_i = \frac{\pi_{\theta_m}(a_i|s_i)}{\pi_{\theta_m^0}(a_i|s_i)}. \tag{20}$$
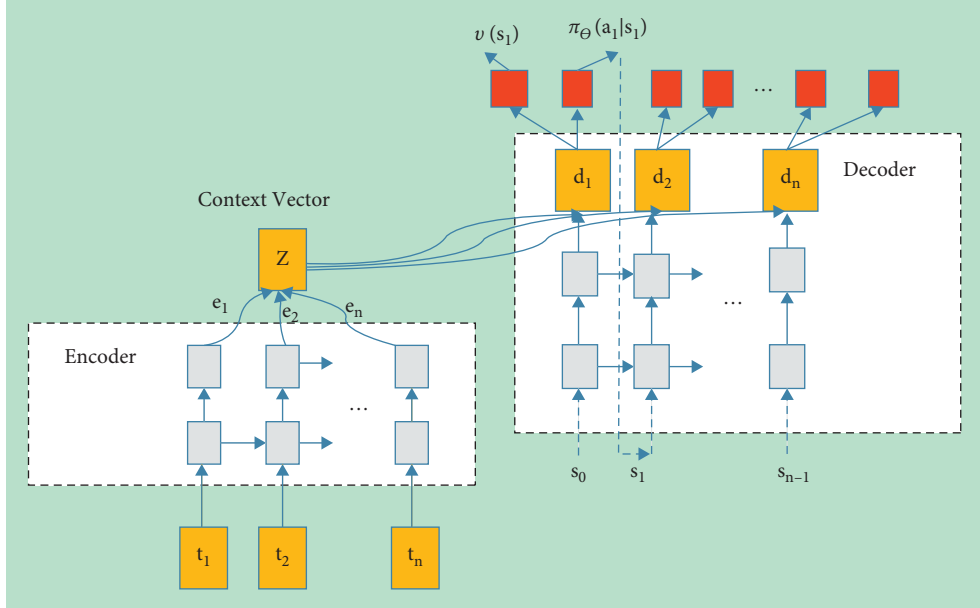
FIGURE 1: Architecture of the seq2seq neural network.

According to the formula (10), t_{i+j} the immediate reward function that can be obtained is

$$r_{i+j} = E\left(-W_{i+j}\right). \tag{21}$$

gradient update formula of the neural network parameters is

$$\theta'_i = \theta_i + \alpha * \nabla_{\theta_i}\left(J\left(\theta_m\right)\right). \tag{22}$$

Among them, α is a training coefficient, which is the learning rate, that is, the gradient descent factor. When the neural network parameters converge iteratively, the output of seq2seq is the optimal decision.

7. Experiment and Result Analysis

The list of simulation parameters used in this paper is as follows, including model parameters and algorithm parameters. In the simulation process, a sample is divided into 10, 12, 15, and 20 data packet sequences for training. In order to reduce the channel correlation, the minimum delay time segment of data packet transmission is taken as 5 s.

The water model and neural network, as well as the parameter settings of PPO2 are shown in Table 1.

After encoding the delay state and channel state, respectively, they are input into the neural network. Figure 2 shows the change of the device delay state. When the decision is to send immediately, the delay is much smaller than the delay caused by suspending transmission. Figure 3 shows the change of the channel state. The channel state, as the environment in which the smart device is located, is relatively stable when the signal-to-noise ratio is good, and is more likely to change randomly when the signal-to-noise ratio is poor.

Each time the far-transmission device sends a data packet, it may either suspend the sending or send it

TABLE 1: Simulation parameters.

Parameter	Value description
Number of packets	$n = \{10, 12, 15, 20\}$
Suspend the smallest time segment	$\delta = 5\,s$
Packet transmission time	$T_f = 1\,s$
Packet size	<200 bytes
SNR threshold	snr = [1.28 3.28 5.28 6.28]
Equipment voltage	$V = 3$
Device idle current	$I_0 = 250\mu A$
Number of samples	$M = 5$
Expected life cycle	$d = 10 * 365$ days
Learning rate	$\alpha = [0.002, 0.005]$
Return discount factor	$\gamma = 0.9$ _
Advantage function discount factor	$\phi = 0.95$
Clipping constant	$\Delta = 0.2$
Number of neurons	unit $s = 2\ 56;$
Overfitting factor	0.5 _
Hidden unit	Layers = 256
Coding layer	Layer 1 = 2
Decoding layer	Layer 2 = 2

immediately. The transmit power each time a data packet is sent is related to the state and environment it is in. Figure 4 shows the power when different data packets are transmitted, and the different power levels reflect the equivalent model of repeated transmission. The power consumption of suspending the transmission of data packets is related to the idle current and the waiting time. The power consumption of sending a packet immediately is related to the transmit power/device voltage, and the transmit duration. The immediate return in Figure 5 is the negative value of power consumption, and the minimum power consumption is equivalent to the maximum return.

After the water MDP model is established, it is solved by the PPO2 algorithm. For the sake of convenience, the reference objective function value is -6 in the experiment, and
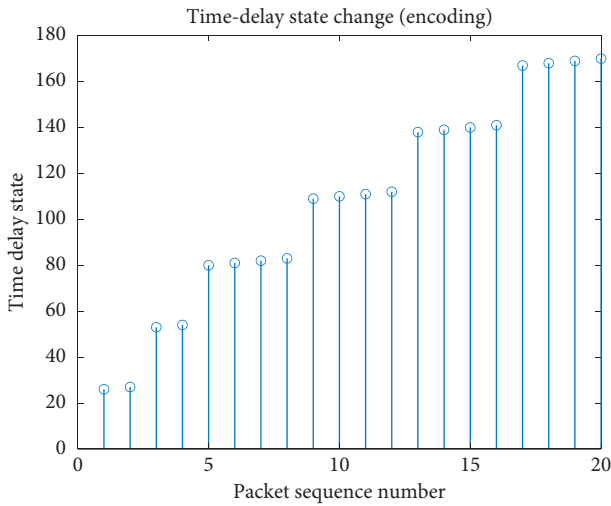
FIGURE 2: Delay change process of packet sequence.
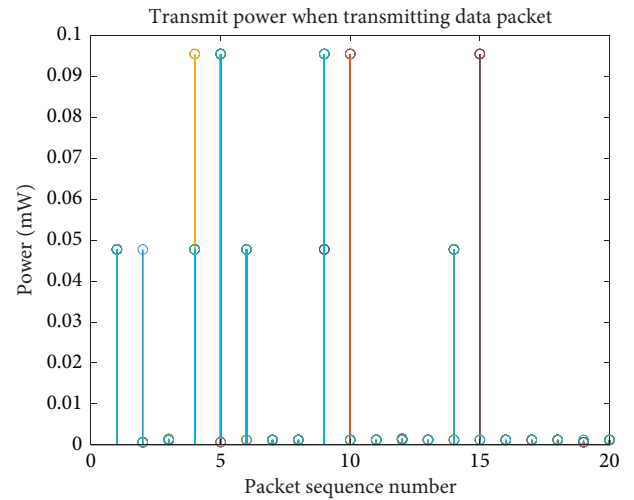
FIGURE 4: Transmit power of data packet sequence.
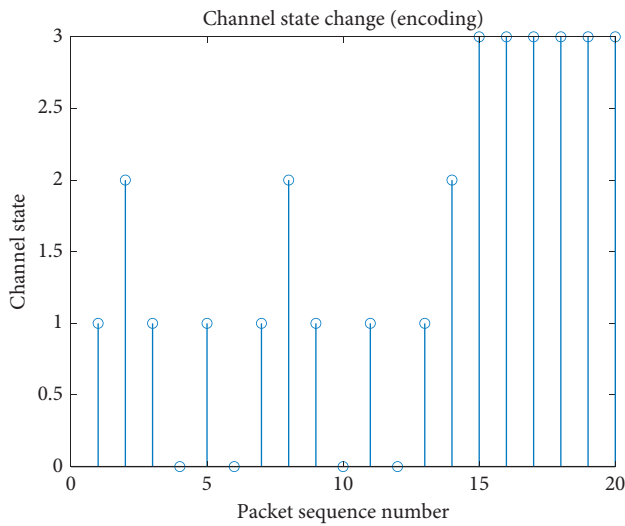
FIGURE 3: Channel change process of packet sequence.
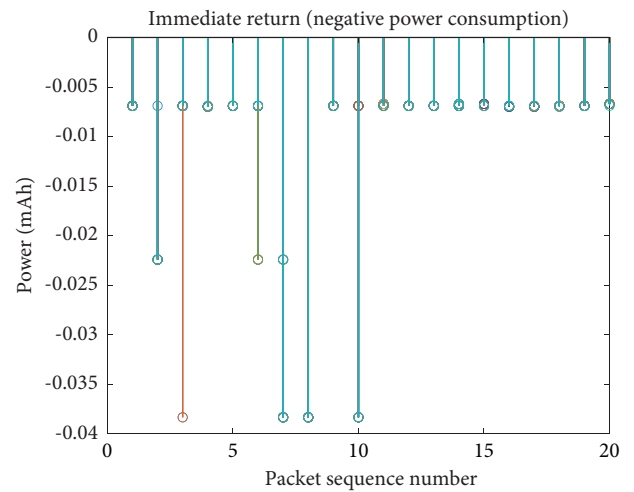
FIGURE 5: Immediate reward of packet sequence decision.

the difference between the training objective and the reference objective value is defined as the loss function. The parameters of the neural network converge as the number of iterations increases, and the strategy is also more optimized. When the loss function is stable, the parameters of the neural network converge to the optimal, and the corresponding strategy is the optimal expression of the neural network. Figure 6 is the parameter iteration process when the learning rate is 0.002, and Figure 7 is the parameter iteration process when the learning rate is 0.0005. When the learning rate is increased, the convergence is faster, but the oscillation is also larger.

The agreement between the water remote transmission equipment and the platform stipulates that a single data should be less than 200 bytes, so when the amount of sample data to be sent is large, it will be split into more data packets.

On the other hand, when the amount of data to be transmitted is constant, the more data packets are split, that is, the smaller the data packets, the lower the cost of retransmission, which is more suitable for the requirements of the narrowband Internet of Things. At this time, the loss function is smaller and easier to converge. Figures 6 and 7 are, respectively, the convergence of neural network parameters for 10, 12, 15, and 20 data packets.

Figure 8 randomly distributes the full connection bias range of the neural network from 0 to 1. After adjusting to a normal random distribution, the convergence of the loss function will be more oscillating due to the increase in the dynamic range.

As mentioned above (Figure 5), if the perception model and algorithm in this paper are not used, the power consumption of one data packet is 0.04 mAh. If 20 data packets are sent every day and the life cycle of 10 years is calculated,
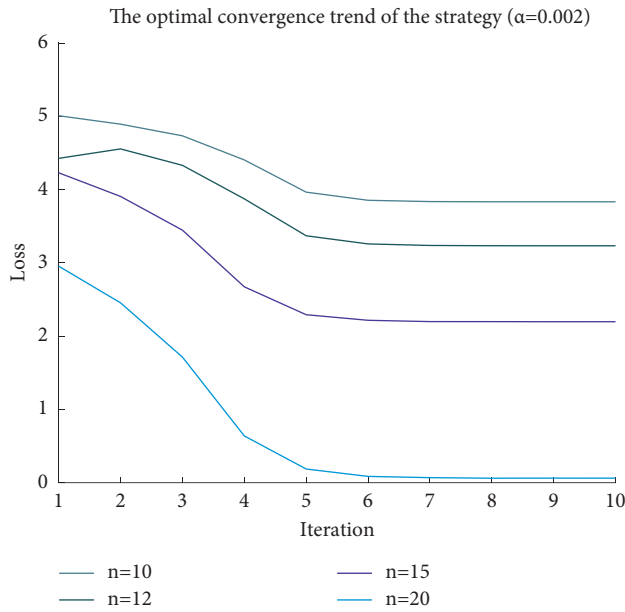
Figure 6: Convergence process of policy (learning rate 0.002).

Figure 7: Convergence process of policy (learning rate 0.005).
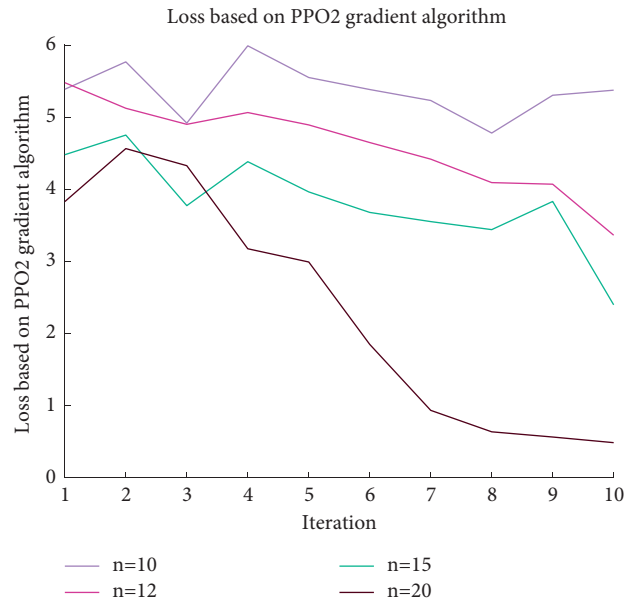
Figure 8: Convergence process with wider bias range (normal random distribution).

8. Conclusion

Starting from the actual problem of the layout and location mechanism of remote transmission equipment of intelligent water affairs, this paper establishes a system model based on MDP, namely, the Markov sequence decision model. The state, action, and reward function definitions of the model are given in this paper, and the simulation results of these definitions are given in the experimental part. The sequence decision is expressed based on the Seq2Seq neural network, and the PPO2 algorithm is used to solve the MDP sequence decision problem. The solution process and results are given in the experimental part. And when the model and method are applied to the actual water management problem, based on the life cycle of the remote equipment and the battery capacity value, the application scope of the equipment layout and location selection is solved.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] X. Lin, Yu Tan, J. Zhang, C. Yang, and J. Liu, "High energy efficiency policy based on Markov decision in wireless

the required battery capacity is $0.04*20*365*10 = 2920$ mAh. Generally, the medium battery capacity is 5000 mAh, which can support the device to send 1.71 samples per day, and each sample includes 20 data packets. After adopting the perception model and algorithm in this paper, the average power consumption of a data packet in Figure 5 is 0.018 mAh, which can support the device to send 3.7 samples per day. With the increase in the amount of data collected by equipment, the perception model and algorithm in this paper can support more installation point equipment with remote transmission units.

transmission," *Systems Engineering and Electronics*, vol. 36, no. 07, pp. 1433–1438, 2014.

[2] Bo Dai and G. Yuan, *NB-IoT Standards and Key Technologies*, People's Post and Telecommunications Publishing House, China Gongxin Publishing Group, Beijing.

[3] *TCMA SB 040-2019, Technical Guide for Layout and Location, Acceptance and Use of NB-IoT Automatic Meter Reading System*.

[4] T. IoT Industry Alliance, *5G Smart Water White*, p. 12, Paper 2020 Edition, China, 2020.

[5] X. Zhao, Q. Wang, and K. Zhao, "Research and application of reinforcement learning based on constraint Markov," *Computer Science*, vol. 4, no. 4, p. 36, 2009.

[6] Z. Jiang, T. Wang, Y. Song, and Y. Liu, "Transmission scheduling scheme based on deep Q learning in wireless network," *Journal of Communication*, vol. 39, no. 04, pp. 35–44, 2018.

[7] S. T. Chung and A. J. Goldsmith, "Degrees of freedom in adaptive modulation: a unified view," *IEEE Transactions on Communications*, vol. 49, no. 9, pp. 1561–1571, 2001.

[8] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal Policy Optimization Algorithms," 2017, https://arxiv.org/abs/1707.06347.

[9] W. Zhan, J. Wang, Q. Zhu, H. Duan, and Y. Ye, "Computational unloading scheduling method based on deep reinforcement learning in moving edge computing," *Computer Application Research*, vol. 38, no. 01, pp. 241–245+263, 2021.

[10] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *Proc. ICLR*, pp. 1–15, 2015.

[11] J. Schulman, P. Moritz, S. Levine, M. Jordan, and P. Abbel, "High-demensional continuous control using generalized advantage estimation," *Proc. ICLR*, pp. 1–14, 2016.

[12] J. Wang, J. Hu, G. Y. Min, A. Y. Zomaya, and N Georgalas, "Fast adaptive task offloading in edge computing based on meta reinforcement learning," *IEEE Transactions on Parallel and Distributed Systems*, vol. 32, no. 1, pp. 242–253, 1 Jan 2021.

Research Article

An Energy-Efficient Strategy and Secure VM Placement Algorithm in Cloud Computing

Devesh Kumar Srivastava [ID],[1] **Pradeep Kumar Tiwari** [ID],[1] **Mayank Srivastava** [ID],[2] and **Babu R. Dawadi** [ID][3]

[1]*Manipal University Jaipur, Jaipur, India*
[2]*Department of ECE, National Institute of Technology, Jamshedpur, Jharkhand, India*
[3]*Department of Electronics and Computer Engineering, Faculty of Computer Engineering,*
 Institute of Engineering-Pulchowk Campus, Tribhuvan University, Kirtipur, Nepal

Correspondence should be addressed to Babu R. Dawadi; baburd@ioe.edu.np

One of the important and challenging tasks in cloud computing is to obtain the usefulness of cloud by implementing several specifications for our needs, to meet the present growing demands, and to minimize energy consumption as much as possible and ensure proper utilization of computing resources. An excellent mapping scheme has been derived which maps virtual machines (VMs) to physical machines (PMs), which is also known as virtual machine (VM) placement, and this needs to be implemented. The tremendous diversity of computing resources, tasks, and virtualization processes in the cloud causes the consolidation method to be more complex, tedious, and problematic. An algorithm for reducing energy use and resource allocation is proposed for implementation in this article. This algorithm was developed with the help of a Cloud System Model, which enables mapping between VMs and PMs and among tasks of VMs. The methodology used in this algorithm also supports lowering the number of PMs that are in an active state and optimizes the total time taken to process a set of tasks (also known as makespan time). Using the CloudSim Simulator tool, we evaluated and assessed the energy consumption and makespan time. The results are compiled and then compared graphically with respect to other existing energy-efficient VM placement algorithms.

1. Introduction

In recent years, cloud computing has experienced a meteoric rise in popularity. The latest advances in virtualization technology have elevated it to a significant position currently. Cloud computing has given new life to a wide range of applications, including speech and signal processing [1, 2], thanks to the increased bandwidth and flexibility provided by 5G [3]. A variety of applications are submitted to data centers to obtain pay-per-use services. Because of resource restrictions, on-demand workloads are shifted to other data centers on a regular basis. As a result, workload scheduling in a heterogeneous multicloud environment is a hot topic and a challenging undertaking due to the wide variety of

cloud resources available [4–7]. Cloud workloads are planned using VMs, which are assigned based on the amount of traffic and requests. Users may benefit from on-demand, high-quality programs and services from a shared pool of configurable computing resources without having to worry about data storage and upkeep on their home computers [8]. Due to the fact that users with weak computer capabilities no longer have direct access to the outsourced data, data integrity security in cloud computing is difficult for them to maintain. Furthermore, users should be able to use cloud storage as if it were local storage, without having to worry about the integrity of the data stored there.

The application requires a secure infrastructure due to the presence of sensitive data or proprietary procedures [8].

As a result, we treat sensitive requests differently and execute them on a distinct VM. VM placement is an operating mechanism in cloud computing that is implemented to map the most appropriate server or PM to host VM. To improve and enhance the utilization of computing resources, power efficiency, and quality of service (QoS) in cloud computing, choosing the best and most suitable host is crucial and critical. Implementing VM placement in cloud computing is highly difficult, time-consuming, and complex. This is because we cannot predict or forecast how VM starting requests will arrive and study their patterns. Moreover, it is often difficult to accurately find the most suitable and best scenario due to the large size of the data center for the required load.

The data centre resources can be optimised, and energy consumption can be decreased with the use of a suitable VM placement algorithm. This algorithm can manage job scheduling , which helps in reducing the energy usage of data centres. Each VM needs a guaranteed number of computing resources such as memory, CPU, bandwidth, links, and much more to manage the application's security, isolation, and overall performance. Costs can be minimized only when the compute resources are exclusively made available for the application, not based on the load of the work benchmark. To run and maintain all applications' efficiencies, it needs to optimize, monitor, and measure continuously and in real time the data traffic. The data center requires proper precision planning of its network architecture to host hundreds of thousands of devices, including routers, switches, and servers. Until now, lots of advances have been made and have successfully created energy-efficient networking devices and compute servers. Up to 20% of energy conservation can be accomplished by data centers, which also saves up to 30% of the energy requirements that are used for cooling various hardware devices. Currently, a common standard cloud computing deployment model uses quite a large amount of energy, which indirectly produces a large amount of carbon dioxide (CO_2).

Hence, the goal for cloud providers is to conserve energy as much as possible by implementing highly advanced energy-efficient algorithms, protocols, and top-notch highly advanced techniques. Cloud computing is built to be energy-efficient inherently since it has the features of scalability, meaning it can scale multitenant usage and compute resources automatically. It is also to be noted that low-quality VM placement implemented in data centers results in maximum energy consumption. A Service Level Agreement (SLA) is an agreement that bonds the cloud provider and the customer. This agreement guarantees that at least a minimum level of service is being provided. In VM placement, the allotment of task requested to a set of VMs that are running/computing on several different hosts while following the agreements that contain terms and conditions is declared in SLAs. In this article, we focus on the problem of mapping and allocating VMs to PMs in the data center.

This pertains to decision-making like, when a VM is allocated, which VMs are to be reallocated, which PMs are to be switched off, and which VM is to be allocated to which PM.

The VMs are then grouped/clustered based on the type of resources they require. The output is later used as the reference for making decisions for the customized VM instances. These decision-making processes are crucial because they help in saving the energy usage of the entire system by modifying the state of nonactive PMs into disabled mode. The mentioned solution in this article helps in conserving resources and reducing resource wastages in cloud computing system models. This is made possible with the help of virtualization mechanisms and allocating suitable policies for efficiency. We have also formed varieties of subtypes of VMs based on the capability of their resources. The total load of inputs in the data center is calculated by the definite number of input tasks where each individual task is associated with many VMs for execution. During this work, my primary and main aim was to allocate an input task to an already existing VM or else make or create a new VM based on the task and allocate the newly created VM to a currently active host or PM. Because new VMs must be given to the right hosts on a regular basis, there are problems with assigning tasks.

As a result, to deal with the problem of job assignment, we employ the concepts of consolidation and optimal solution. The primary aim is to reduce energy consumption while also appropriately allocating resources in cloud-based applications. To handle this task assignment problem while decreasing energy consumption and ensuring optimal resource allocation in a cloud system, we employ the concept of consolidation and optimum solution approach developed by the authors. Data center time and energy consumption can also be cut down because of the consolidation method.

Consolidation techniques are divided into four subapproaches: server underload detection, server overload detection, VM placement, and VM selection. Each of these subapproaches is discussed in detail below. Consolidation is a tough and time-consuming task; thus, heuristic and meta-heuristic approaches are used to obtain solutions as quickly as possible. The system model proposed in our work is shown in Figure 1, where multiple users can make the request simultaneously.

The proposal made by users will reside in the task queue, which will be allocated to the VM based on the requirements. The critical point here is that, in the IaaS system architecture, the user request can either be a program, task, software, or the operating system itself. The task manager can divide the schedule and tasks and distribute them among multiple VMs to execute the parts of them. In addition, if the request is for system software or an operating system along with a program, then the division of those requests cannot be done in multiple VMs. Therefore, our system model should assign those VMs requests to anywhere in the physical system but not in total allocation. The VM request is being handled by the host as it is requested, depending upon the availability of the resource. The host manager uses a machine learning classifier to ensure the optimal allocation of resources at any moment. We can summarize our contribution as follows:

(1) A new cloud system model has been developed in which the input tasks are interpreted as user requests and the cloud computing resources are heterogeneous.
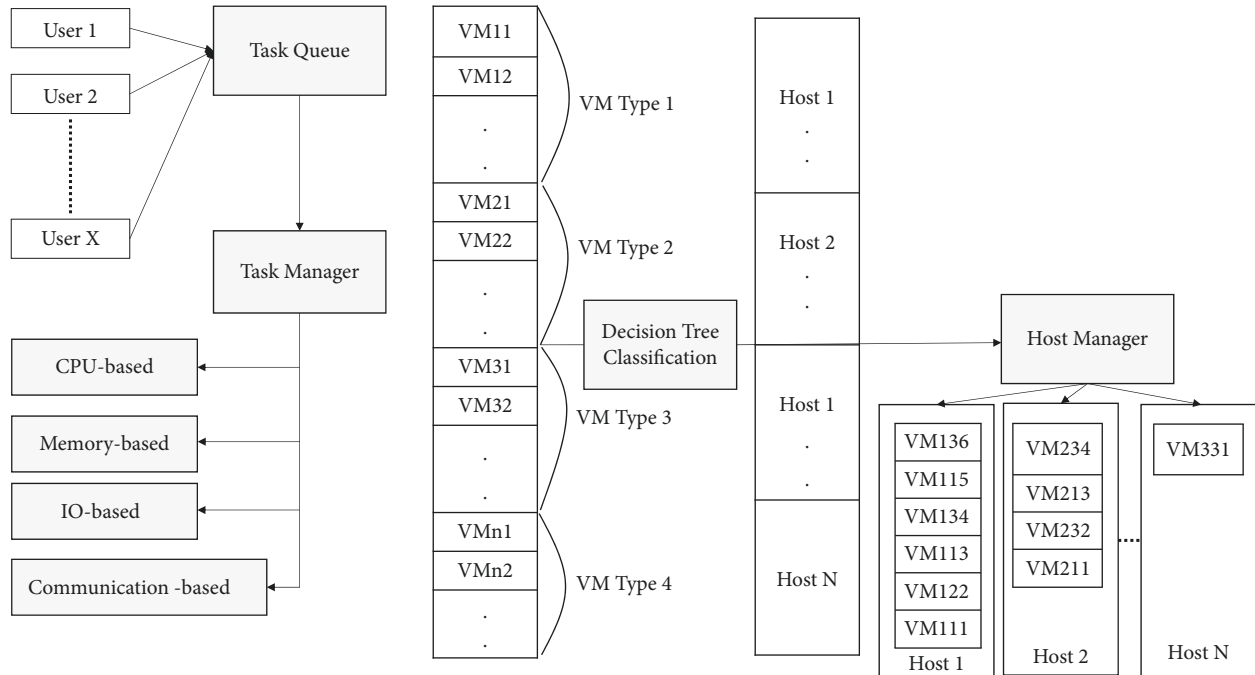
FIGURE 1: Cloud system model developed.

(2) To produce a VM Placement Algorithm named "Energy-Efficient VM Placement Algorithm" (EEVMPA) based on task requests, where this algorithm helps in reducing energy consumption, minimizing the rejection rate of tasks, and reducing the makespan time.

(3) In cloud computing, various requests are common and frequently made. The tasks can be classified based on their parameters, such as platform, space, priority, type, and sensitivity. We are preparing our data set to build a compatible situation for any given task and its required environments. Based on this, we prepare the decision tree to optimize the allotment. We are using multiple entries, out of which 20% will be used for training and 80% of the data will be used for validation.

(4) An overall evaluation of the algorithm is made with the help of the CloudSim tool, and graphical analysis is also done and compared with the other existing energy-efficient algorithms currently available in the market.

2. Literature Surveys

Mousa and Hussein proposed an unmanned aerial vehicles (UAVs)-based offloading system; this research goal is reducing the latency and the energy consumption of the UAV. There are two distinct subproblems for the suggested system that are looked into. The suggested offloading method divides the ground devices into clusters, and the UAV passes over each cluster head to perform the tasks that have been offloaded for the members of the cluster. This leads to the first subproblem, which is the clustering of IoT devices, the

shortest route for the UAV to travel through the cluster heads [9]. Cloud computing has enormous potential with the advancement of virtualization technologies. Large hardware resources are frequently virtualized. Clients are then assigned to these tiny units. However, these services must be resource-efficient. Many scheduling, allocation, and provisioning issues are framed as optimization issues. Many algorithms were introduced in the past that focused on resource allocation, consumption of energy, and issues faced during scheduling. A detailed literature survey was the need of the hour. Gabhane et al. [10] describe the meta-heuristic model that is extensively studied for VMP in cloud computing. Based on a lot of research and surveys, we found some of the most common related works, like this one. We also looked at their workings, benefits, and drawbacks. Karmakar et al. proposed [11] high-performance computing requires many VMs to meet user demand (VMs). This computation generates a lot of VM-to-VM traffic. Therefore, to satisfy user expectations, VMs must be deployed on actual computers in a way that reduces communication costs and delays. By consolidating VMs into less active physical computers, service providers can reduce operational costs. Finding the best way to reduce deployment or communication costs is NP-Hard. Moreover, focusing on one without the other may result in a cheaper but slower solution (or vice versa). Ant colony optimization is used to manage VM consolidation. The performance of this algorithm was compared with existing mono- and multiobjective algorithms.

Omer et al. [12] focused on large number of complex applications with varying priorities, and resource demands are arising in recent telecommunication paradigms like big data, IoT, UEC, and machine learning. Most researchers use a set of VMs with a set traffic load. Cloud data centers

(CDC), a key component of UEC, perform differently depending on how they are placed. VM placement is NP-hard; therefore, no optimal solution exists. This study presents a priority, power, and traffic-aware VM placement strategy in a CDC. A fat-tree topology and the proposed approach were compared. The proposed method decreases total network utilization by 29%, electricity consumption by 18%, and resource waste by 68% compared to the second best findings. Raman et al.'s [13] study is focused on knowing the makespan, and execution cost is required for workflow scheduling efficiency. Optimal workflow scheduling is tough in the cloud since estimating makespan and cost is tricky.

Cloud resources are scheduled to suit user demand. The virtualization approach provides for scalability. This study proposes the PBF-NN hybrid scheduling approach for calculating makespan and execution cost. The backfill algorithm schedules work. This method decreases migration compared to First Come, First Serve. It is then used for resource allocation equity. The system assigns jobs fairly. An energy-efficient VM placement approach is provided by the backpropagation neural network-genetic algorithm (BPGA). By distributing tasks dynamically, it saves resources. Experimental research illustrates how the recommended technique saves money and time. Salami et al. [14] highlight that the VMPP is a frequent provisioning issue. VMPP hosts VM requests on a small number of real computers. The VMPP is solved via Cuckoo Search (CS) and creates new cost and perturbation functions. Using two popular benchmark data sets, Reordered Grouping Genetic Algorithm, Best-Fit Decreasing, and MultiCSA also performed well. Gharehpasha et al. [15] work on store applications, data, and files; cloud computing employs a vast network of systems which reduces power computation, application hosting, content storage, resource waste, and delivery costs. By focusing on corporate goals rather than increasing consumer hardware resources, cloud computing helps organizations achieve their goals. VMs with real equipment in cloud data centers are challenging to place. Cloud data centers may better manage resources by putting VMs ahead of real ones. They also propose a hybrid discrete multiobject whale optimization method with a multiverse optimizer and chaotic function. The suggested method reduces the number of active physical computers in cloud data centers. VMs over real machines in cloud data centers decrease resource consumption. This approach slows the VM to PM migration. A final comparison was made using first fit, virtual machine placement ant colony system (VMPACS) and Modified Best Fit Decreasing (MBFD). Bouhank and Daoudi [16] use placing VMs on appropriate servers in a cloud context for NP-hard issue. They propose a Non-dominated Ranking Biogeography Based Optimization (NRBBO) algorithm to decrease the overall resource waste and power consumption across all servers in a VM installation. To test the approach's efficacy, many experiments were run using generated data from the literature. The findings demonstrate that NRBBO outperforms other multiobjective systems in terms of efficiency, convergence, and coverage.

Mousa and Hussein proposed a UAV-Assisted Computing Mobile-Edge Computing UCMEC system to increase the performance of offloading workloads from mobile devices and reduce task latency and system energy consumption. The suggested method divides the ground devices into regions so that the UAV can fly over each zone and complete the jobs. The partitioning technique employs a VD. Using a Graphics Processing Unit (GPU)-based Particle Swarm Optimization PSO, the UAV's trajectory over the areas is optimised. Authors' algorithm shows the best result among the compared algorithms [17].

Farzaneh and Fatemi's [18] work focused on the critical optimization challenge in cloud architecture that is VM placement (VMP). The solution impacts prices, energy, and performance. PM processing power and VM workloads have powered VMP. The semiconductor industry is also interested in chips having numerous homogeneous or heterogeneous PEs. The newest chip has general-purpose cores with reconfigurable fabric (RF). This study provides a random PE design technique for VMP algorithms. In the given solution, VMP employs RF components in the cloud infrastructure. It's a heuristic to tell the difference. Performance depends on parameter extraction. The obtained parameters are utilized to determine which PM should host the VM. Others outperform our suggested VMP algorithm in our proposed cloud architecture model.

Ghetas [19] proposed a multiobjective Monarch Butterfly Algorithm to manage the VM placement. This study helps to find the location of VMs to place the load.

Gohil et al.'s [20] works on the cloud computing load management is a new way to deliver web services. People, businesses, and academics have all noticed demand-based resource allocation in cloud computing. Rapid data computation and storage expansion in the cloud may produce workload imbalances that breach SLAs and degrade system performance. Workload balancing is crucial in cloud computing. When creating a new VM, the best PM in a cloud data center is chosen. This study examines VM placement in cloud data centers to maximize resource usage. This problem has many solutions. Most available methods balance cloud resources due to inefficient resource consumption. We describe an algorithm that minimizes resource imbalances, waste, and leakage while maximizing resource consumption. Our system selects a host for an incoming VM from a list of hosts based on cosine similarity. The simulation findings show significant gains over existing algorithms like Round-Robin and the default Worst Fit. Chen et al.'s [21] work focuses on reducing energy consumption and thermal costs. It also reduces the number of hot spots in the cloud computing system platform. In this research, Jayanetti et al. proposed a novel hybrid actor-critic technique for resource scheduling in an edge-cloud setting, together with a novel hierarchical state space simulation. The resulting deep reinforcement learning system dramatically decreases the size of the action space handled by each actor network while simultaneously encouraging a distinct separation between edge and cloud nodes, with numerous actor networks directed by a single critic network.

Authors also used proximal policy optimization to get around known issues with conventional actor-critic approaches. In order for the deep reinforcement learning

framework agent to develop a balanced trade-off between latency and energy consumption, we also made use of existing works to break down workflow deadlines into individual task deadlines. These were then used as soft upper bounds during the training process. Results from simulation experiments show that the deep reinforcement learning framework works better than all other comparison algorithms by consuming less energy while keeping an algorithm's total execution time that is comparable to other algorithms. [22]. Scheduling problems is proposed by work done by Casas et al. [23] which focuses on the problems of scheduling of scientific workflows in cloud computing systems. Allocation of computing resources is proposed by work done by Hameed et al. [24], which mostly explains how different resource allocations in cloud computing work and gives information on how they work. Kloh et al.'s [25] research work contributes to optimizing running time and lowering costs. This scheduling mechanism is much better than an approach called "join the shortest queue," but it does not guarantee the best utilization of resources. Deng et al.'s [26] work on a scheduling algorithm for Bag-of-Task (BoT) applications gives us an idea of how scheduling algorithms work in BoT applications. It is a priority based on the tasks and VMs. It then selects the best and most appropriate VM group/cluster for a distinct task cluster. Mosa and Paton's work focused on the management and optimization of energy usage through lower SLA rate violations.

Mosa and Paton developed a self-managing VM placement solution in cloud data centers that dynamically assigns VMs to hosts in accordance with resource use that was provided in this work. It is built on utility functions. The approach's major objective is to maximize an IaaS provider's profit by lowering energy consumption costs and the price of various SLAV sources. The effectiveness of the proposed utility-based approach and an existing heuristic-based solution have been compared through experiments. The proponents of the heuristic method were demonstrated to perform well after applying it to a comprehensive review in comparison to other heuristics.

In both lightly loaded and more heavily loaded cloud data centers, analysis revealed that the suggested utility-based solution beat the current heuristic-based method in terms of energy savings and SLAV minimization [27]. In this paper, we explored the energy consumption optimization of service provisioning within a volunteer cloud. We give a comprehensive solution that includes the service provisioning problem modelling, a hardness analysis, alternative resolution approaches, and experimental examination of those ways. Our investigation demonstrated that we can find effective algorithms for the real-time optimization of our service supply [28]. Belgacem et al. proposed an Intelligent Multi-Agent Reinforcement Model (IMARM) for optimizing cloud resource distribution in cloud. It provides a comprehensive solution for cloud service providers since it addresses the problem of resource allocation from several angles.

Using mechanisms for checkpointing and VM migration, the suggested approach achieves load balancing and provides fault tolerance. Sensor agents transmit system failure status, energy use, and workload on a periodic basis to track system reaction (recovery, downtime, migration, etc.). Q-Learning enables the indication of the action to be taken in each system state, recognizing that the weight of the virtual machines, overall energy consumption, and quantum time are the major factors employed to optimize the system. With the option to choose whether a task or virtual machine should be moved, restored, or shut down, it offers a fascinating new approach for determining the optimal response to shifting cloud settings. The outcomes and experimental analyses of proposed algorithm work better than other methods. By providing a minimal execution time while utilising resources effectively, it demonstrated a positive influence on the execution time [29]. Work of Mastroianni et al.'s [30] research is based on the idea of Bernoulli trials, and it works on two main principles: self-organization and adaptability. This makes it more efficient at managing large data centers.

3. Materials and Methods

Based on SLA policies agreed upon between the cloud provider and users, VM placement can dynamically map VMs onto PMs/hosts. Inside the PM/host, VMs are controlled and managed by a specific layer, which is a software known as the Hypervisor. There are two types of VM placement schemes. Those are as follows: (i) Static VM Placement: this is a VM mapping methodology in which the mapping technique is fixed and cannot be recomputed for an extended period of time. (ii) Dynamic VM Placement: it is a mapping methodology of VMs in which the initial placement can be changed because of some changes (dynamics) that occur during the system computing load or other types of loads. A Dynamic VM with a placement mechanism is used in this research due to its flexibility during execution and its ability to minimize energy consumption. To create the energy-efficient algorithm, we first developed a cloud system model, which is a scheduling model where the tasks are advanced through various entities, such as task queue, task manager, task segregator, various VM-type allocation, host manager, and various hosts. These are all included in a cloud data center. The cloud computing system has m hosts ($H = \{H1, H2, \ldots, Hm\}$) which are heterogeneous in nature with respect to the capacity of resources. Initially, tasks are received from the users over the Internet. These tasks are known as "service requests." A cloud service provider (CSP) provides these services to users when they request any service. Users present their tasks to the task queue, and the task manager separates and clusters the tasks into four main, uniquely different categories of tasks. These four categories are CPU-based, memory-based, I/O-based, and communication-based. These tasks are separated into four types because it could map similar types of tasks to VMs quite easily. Hence, this methodology is implemented. The task manager later gets to know about the types of VMs and learns about the subtypes of VMs that are offered. As a result, four types of VMs are created to map smoothly with the VM. They are named as VM1 for CPU-based tasks, VM2 for memory-

based tasks, VM3 for I/O-based tasks, and VM4 for communication-based tasks. Each VM-type further contains a subtype (VMab), which represents ath VM of bth type). After mapping the tasks into the VM-types, the four VM-types are submitted to the host manager, where the host manager gets to study the whole information about the VM-types and record it. The host manager later chooses the appropriate VM type based on the task type to be processed and then creates a suitable VM (VM) on the PM (PM) or host. By doing this, energy consumption can be minimized and makespan time can also be reduced significantly. At the start, all hosts are in an inactive or sleep state. The host manager checks if any active host has adequate resources to create a VM. If that condition is satisfied, then it creates a VM on that active host. This methodology helps in saving energy. If there are no active hosts, then the host manager goes towards a nonactive/sleep host, changes its mode status to an active state, and then creates a brand-new VM on that host. When there is a situation where no tasks are present in the task queue, then the connection between the VM and the PM host is terminated, and the resources are set free from usage.

A specific energy model is adopted throughout this project/research, which helps in reducing energy consumption. In this model, energy consumed by a task can be calculated. Ultimately calculating the total energy consumed by individual tasks results in finding out the overall energy consumption of the whole data center. Figure 1 shows the host manager managing the available hosts, and these hosts contain the serial number of VMs. Each host is defined by a unique ID to identify the host, and each host VM also has a unique ID to identify the VMs. It helps to manage the load on different hosts. The task manager manages the tasks of different userbases, and these userbases are also identified by a unique ID. A unique ID helps to manage the requested resources and capacity of infrastructure that are already assigned to a particular userbase.

Energy consumed by a task (t_x) is represented as E_{xyz}, where energy is consumed by xth task which is present on yth VM (VM) and zth PM (PM) or host. The formula for calculating E_{xyz} is as follows:

$$E_{XYz} = E_{xyz}^{tr} + E_{xyz}^{Pr}, \tag{1}$$

where E_{xyz}^{tr} is the energy consumed during file data transfer and E_{xyz}^{pr} is the energy consumed during task execution.

Energy consumed during file data transfer, i.e., E_{xyz}^{tr}, is calculated by the following formula:

$$E_{xyz}^{tr} = \left\{ \sum_{l=1}^{h} T_i metr\left(f_{xYz}^l\right) \right\} \times P_{avg}, \tag{2}$$

where E_{xyz}^{tr} is calculated by the product sum of all needed files transfer time with the transfer of data's average power consumption (P_{avg}).

The host's average power consumption (P_{avg}) in the data center is calculated as follows:

$$P_{avg} = \left(\sum_{z=1}^{m} Pactive_z \right), \tag{3}$$

where P (active) is the power consumed by the zth host which is in active state.

Energy consumed during task execution, i.e., E_{xyz}^{pr}, is calculated by the following formula:

$$E_{xyz}^{Pr} = ET_{xyz}^{I} = P_{yz}^{I}, \tag{4}$$

where E_{xyz}^{pr} is calculated by the product of task execution time (ET_{xyz}^{I}) and the average power consumed by yth VM (I-type) of zth host. I is the VM type where I can have four values, i.e., $(I = \{1, 2, 3, 4\})$. If $I = 1$, then it is CPU-based tasks. If $I = 2$, then it is memory-based tasks. If $I = 3$, then it is I/O-based tasks. If $I = 4$, then it is communication-based tasks.

$$NETC_{ijk} = \frac{\left(ETC_{ijk} - Minimum_{ij}\right)}{Maximum_{ij} - Minimum_{ij}}. \tag{5}$$

$$f(x) = \begin{cases} MCXXS(i; m2 + 1); & if\ m = odd, \\ MCXXS(i; m2) + MCXXS(i; m2 + 1)2; & otherwise. \end{cases} \tag{6}$$

Some of the important tools such as CloudsSIM 3.0.3 and NetBeans 7.4(IDE) are used for the analysis and simulation: the median to perform the task over cloud can be calculated using the below equation. The normalized ETC matrix element NETCij; k is formed by taking the ratio between the difference in ETCij; k and the minimum execution time of task Ti; j and the difference in maximum and minimum execution time of task Ti; j.

4. Problem Statement and Formulation

VM Placement plays a crucial and important role in the cloud computing data center. It helps in a large way to reduce energy consumption and enables proper resource utilization in the data center. Here in this project/research, we have considered m hosts, which are in a heterogeneous state. Each host is either in an active or inactive (sleep) state at the beginning of the stage. There are also k types of VMs in this cloud system that are classified based on their resources. By using the same idea for VM types, tasks are also categorized into the same task clusters so that they can fit into one suitable type of VM. Hence, the task manager plays a crucial role in this algorithm as it gets complete information about the service request tasks from the task queue, categorizes these tasks into four main task groups, and uses this information to decide whether to create new VMs or not in the host/PM. Therefore, due to this circumstance, a job assignment problem arises as it constantly deals with the allocation of newly created VMs to appropriate hosts. Hence, we implement the idea of the consolidation method and optimal solution methodology to solve this job assignment problem. However, our prime focus is to reduce energy consumption and proper resource allocation in cloud systems.

Now implement the idea/concept of the consolidation method and optimal solution methodology to solve this job assignment problem while keeping our primary focus on

reducing energy consumption and proper resource allocation in the cloud system. Algorithm also helps us to make span time and energy consumption-aware in the data center. The consolidation method is a complicated approach which can be further divided into four subapproaches, those are server underload detection, server overload detection, VM placement, and VM selection. Finding solutions using the consolidation method is difficult and time-consuming. Hence, heuristic-based and meta-heuristic-based approaches are used to find solutions quickly and efficiently.

5. Methodology

The following are the methods implemented during the research work: (i) a new cloud system model is developed where the input tasks are taken as the requests from the users and the cloud computing resources are in a heterogeneous state. (ii) A VM Placement Algorithm named Energy Efficient VM Placement Algorithm (EEVMPA) is produced based on task requests, where this algorithm helps in reducing energy consumption, minimizes the rejection rate of tasks, and reduces the makespan time. (iii) In cloud computing, heterogeneous requests are common and requested frequently. The tasks can be classified based on their parameters, such as the platform, space, priority, type, and sensitivity. We are preparing our data set to build a compatible situation for any given task and required environments. Based on this, we prepare the decision tree to optimize the allotment. We are using multiple entries, out of which 20% will be used for training and 80% of the data will be used for validation. (iv) An overall evaluation is made on the algorithm with the help of the CloudSim tool, and graphical analysis is also done and compared with the other existing energy-efficient algorithms currently available in the market.

6. Proposed Work

In this research work, we developed seven algorithms, where one of the algorithms is then in the form of MB (megabyte). IO_x is the xth task's input/output needs. b_x is the bandwidth of the xth task and is in the form of MB (megabytes). The EEVMPA consolidation algorithm is Algorithm 1 where tasks set (TS) with n number of tasks, deadline set (DLS) with a n number of deadlines, hosts set (HS) with a m number of hosts, and VM-types (VMt) with four types, i.e., $\{y = 1, 2, 3, 4\}$, are given as input resources to the algorithm. This algorithm gives us output as Algorithm: in Algorithm 2, in Step 2, another subalgorithm is called "CategorizedTask"—Algorithm 3, which helps to categorize all the tasks into the mentioned four categories. Here the sorted, updated, and refreshed task queue RQ is given as input. Here, the task resources, which are made into five tuples, are used and represented as $RPTK_x$, which should also contain a lower bound and upper bound for each tuple. $RPTKx = \{Len_x, dls_x, MM_x, IO_x, b_x\}$. CPUB is the upper bound and CPLB is the lower bound for task lengths. DLSU is the upper bound and DLSL is the lower bound for tasks' deadline. MMU is the upper bound and MML is the lower bound for

tasks in main memory. IOU is the upper bound and IOL is the lower bound for task IO requirements. bU is the upper bound and bL is the lower bound for task bandwidth.

Algorithm 3 returns the four categorized tasks sets, i.e., CPU-based, memory-based, I/O-based, and communication-based. In Step 1 of Algorithm 3, processing speed's upper bound (PSU) is calculated from Step 2 to Step 14 in this algorithm, a loop is created and at the xth iteration, xth task is deployed at one of the four categories for the tasks. In Algorithm 1, we could notice that in Step 1, a subalgorithm is called (SortedQueueTask—Algorithm 2) with the help of set of tasks and deadlines. Algorithm 2 helps in sorting all the service request tasks in ascending order based on their deadlines. Algorithm 2 consists of Delete Min function which helps in deleting the deadlines, where r is of minimum value from the set of tasks, and then stores it in the RQ (Refreshed Queue) mentioned to calculate the values of vc_x, vm_x, vio_x, and vb_x, which have the value between 0 and 1. But when you add all these, then the sum should be equal to 1.

In Step 6, we could see that all the values of vc_x, vm_x, vio_x, and vb_x are multiplied with z_x. After multiplying, we compare all the four values and find out the Max value which determines the category of the task. In Algorithm 1, from Step 3 to Step 9, the loop runs x number of tasks. In Algorithm 4—VMsFree, the VMs which are in free/no work state change to sleep or inactive mode/state and those VMs compute resources are transferred to the next host. The active host connected with its allotted VM is provided as input to this algorithm.

The ultimate goal of this algorithm is to reduce the number of active VMs and regularly keep them updated with the host manager. In Algorithm 1, Algorithm 4 is a subalgorithm for it, and it is called. In Algorithm 5—Hosts Free, it is a subalgorithm for Algorithm 1 where it is called at Step 5 in it. This algorithm gets input as an active host set and an active VM set. Steps 1 through 5 of the loop assist in converting the host's state, i.e., it converts an active host into a sleep or inactive state if the host is not working or is in an idle state. The state of the host will be changed from Step 7 to Step 18. If VMs on a minimum loaded host/PM can be fully migrated to other hosts that are in an active state, then all the VMs are migrated to other active hosts and the previous host changes its mode to sleep or inactive state.

In Algorithm 1, in Step 6, a subalgorithm is called (Algorithm 6—VMTypeSelection). Algorithm 6 chooses a VM-type based on the task type and the requirements of the tasks. Here, tasks, task types, VM types, and VM subtypes are sent as input. VM types are of four types, that is VM type1 for CPU-based, VM type2 for memory-based, VM type3 for I/O- based, and VM type4 for communication-based. There are n number of VM subtypes in a VM type. For example, for VM type1, the VM subtypes are as follows: in Algorithm 1, in Step 7, a subalgorithm is called (Algorithm 7—host selection). Algorithm 7 helps in finding the best host for the specific VM type. The VM type that was found from Algorithm 6 is used as input here along with the hosts which are set in active state and the hosts which are set in a sleep/inactive state. In this algorithm, steps 1 to 6 help us find an active host so that a VM can be placed on it. In steps 7 to 13, this loop helps in searching for sleep state hosts if any other

Input: TS = {ts, ts$_2$, ..., ts$_n$} ⟶ Tasks set,
     DLS = {dls$_1$, dls$_2$, ..., dls$_n$} ⟶ Deadlines set,
     HS = {hs$_1$, hs$_2$, ..., hs$_m$} ⟶ Hosts set,
     VM$_{ts}$ = {vm$_1$, vm$_2$, vm$_3$, vm$_4$} ⟶ VM-types set.
   VMsFree () ⟶ function to check the VM is free
   HostsFree() ⟶ function to check the VM is free
Output: Energy consumed and makespan time
(1) Update RQ ← SortedQueueTask (TS, DLS) from Algorithm 2
(2) {Q_s, Q_t, Q_u.Q_v} ← CategorizedTask (RQ) from Algorithm 3
(3) for each and every task tk_x ∈ TS then do
(4)   VMsFree () from Algorithm 4
(5)   HostsFree () from Algorithm 5
(6)   vm ← VMTypeSelection (ts_x, VMType(ts$_x$, TS)) from Algorithm 6
(7)   h ← HostSelection (vm) from Algorithm 7
(8)   Allot ts$_x$ to *vm* which is placed on host *h*
(9) end for loop
(10) VMsFree ()
(11) HostsFree()

ALGORITHM 1: Energy efficient VM algorithm (EEVMPA).

Input: TS = {ts$_1$, ts$_2$, ..., ts$_n$} ⟶ Tasks set,
   DLS = {dls$_1$, dls$_2$, ..., dls$_n$} ⟶ Deadlines set.
Output: RQ (Refreshed Queue).
   Shorting all the service request task in ascending order based on their deadlines
(1) for x = 1 to n then we have to do
(2)   RQ[x] ← DeleteMin (TS$_x$)
(3)   DeleteMin (TS$_x$) will delete the task which contains less dls$_x$ value
(4) end for loop
     return RQ

ALGORITHM 2: Sorted queue task.

Input: TS = {ts$_1$, ts$_2$, ..., ts$_x$} ⟶ Tasks set,
   RPTK$_x$ = {Len$_x$, dls$_x$, MM$_x$, IO$_x$, b$_x$} ⟶ Tasks resource requirements,
   CPU: (CPUU, CPUL), Main memory: (MMU, MML), Deadline: (dlsU, dlsL), I/O: (IOU, IOL), Bandwidth: (bU, bL).
Output: Four types of tasks categorize (CPU-based, Memory-based, I/O-based and Communication-based).
(1) PSU ← CPUU/ddlU
(2) For each task tk$_x$ ∈ TS then do
(3) C_x ← Len$_x$/dls$_x$
(4) w_{cx} ← c_x/PSU, w_{mx} ← MM$_x$/MMU
(5) w_{iox} ← IO$_x$/IOU, w_{bx} ← b_x/bU
(6) z_x ← 1/w_{cx} + w_{mx} + w_{iox} + w_{bx}
(7) w_{cx} = $z_x X w_{cx}$, w_{mx} = $z_x X w_{mx}$
(8) w_{iox} = $z_x X w_{iox}$, w_{bx} = $z_x X w_{bx}$
(9)   Maximum = {w_{cx}, w_{mx}, w_{iox}, w_{bx}}
(10)   ts$_x$ ∈ CPU-based if w_{cx} = Max
(11)   ts$_x$ ∈ Memory-based if w_{mx} = Max
(12)   ts$_x$ ∈ I/O-based if w_{iox} = Max
(13)   ts$_x$ ∈ Communication-based if w_{bx} = Max
(14) end for loop
(15) return CPU-based, Memory-based, I/O-based and Communication-based

ALGORITHM 3: Categorized task.

Input: ACTH = {acth$_1$, acth$_2$, ..., acth$_k$} $\longrightarrow$ Hosts which are in active state,
ACTV = {actv1$_1$, actv1$_2$, ..., actv$_{21}$, actv$_{22}$, ..., actv$_{2p}$} $\longrightarrow$ VMs which are in active state.
Output: ACTV Updated.
(1) for each host in active state acth$_i \in$ ACTH then do
(2)   for each VM actv$_{ij} \in$ acth$_i$ then do
(3)   if actv$_{ij}$ in idle state then
(4)   Deallocate actv$_{ij}$ resources to acth$_i$
(5)   end if loop
(6)   end for loop
(7) end for loop

ALGORITHM 4: VMs free.

Input: ACTH = {acth$_1$, acth$_2$, ..., acth$_k$} $\longrightarrow$ Hosts which are in active state,
ACTV = {actv$_{11}$, actv$_{12}$, ..., actv$_{21}$, actv$_{22}$, ..., actv2$_p$} $\longrightarrow$ VMs which are in active state.
Output: ACTH Updated.
(1) for each host in active state acth$_i \in$ ACTH from acth$_k$ to acth$_1$ then do
(2)   if acth$_1$ is in no working/idle state then
(3)   convert host acth$_i$ from no working/idle state to sleep state
(4)   end if loop
(5) end for loop
(6) StatusofVM = 0
(7) for each host in active state acth$_i \in$ ACTH from acth$_k$ to acth$_1$ then do
(8)   for each VM actv$_{ij} \in$ acth$_i$ then do
(9)   if migration of actv$_{ij}$ to ACTH—acth$_i$ then
(10)   StatusofVM = StatusofVM + 1
(11)   migration$_{ij} \leftarrow$ Host ID which is being targeted
(12)   end if loop
(13)   end for loop
(14)   if StatusofVM = $j - 1$ then
(15)   migrate all VMs using migration$_{ij}$
(16)   convert host acth$_i$ from no working/idle state to sleep state
(17)   end if loop
(18) end for loop

ALGORITHM 5: Hosts free.

Input: Task t_k, type of task PP, VM-Types, VM sub types.
Output: VM (VM) type.
(1) All VM-types contains VM sub types in sorted order
(2) for each VM sub type VMtypepp$_i \in$ VMtypePP do
(3)   if tk is able to fit in VMtypePP$_i$ then
(4)   VM-type $\leftarrow$ VMtypepp$_i$
(5)   return VM-type and then stop
(6)   end if loop
(7) end for loop

ALGORITHM 6: VM type selection.

active hosts are not available. Using the heuristic and meta-heuristic consolidation methodologies here, we can achieve our main aim of finding the best active host for the VM.

7. Results and Discussion

After developing all the algorithm modules, these modules are transferred and performed in the CloudSim 3.03 simulator. Using the CloudSim Simulator tool, we evaluated and assessed the energy consumption and makespan time. The results are compiled and then compared graphically with respect to other existing energy-efficient VM placement algorithms. We also used a software called Xen, which is used as an operating system (Hypervisor) which constantly monitors the VM. This developed algorithm (EEVMPA) is implemented in the Java language. NetBeans 1.7 (IDE) and JDK 1.7 software are also used to provide a platform for the cloud simulator. This experiment was tested on a HP

Input: VM-type,
ACTH = {acth$_1$, acth$_2$, ..., acth$_k$} ⟶ Hosts which are in active state,
SLPH = {slph$_1$, slph$_2$, ..., slph$_k$} ⟶ Hosts which are in sleep state.
Output: Desired Host
Algorithm: Host Selection
(1) for each acth$_i$ ∈ ACTH then do
(2)   if VM of VM-type is able to fit in acth$_i$ then
(3)   Host ← acth$_i$
(4)   Return Host then stop
(5)   end if loop
(6) end for loop
(7) for each slph$_i$ ∈ SLPH then do
(8)   if VM of VM-type is able to fit in slph$_i$ then,
(9)   Host ← slph$_i$
(10)   Convert the state of slph$_i$ from sleep/non active state to active state
(11)   Return Host then stop
(12)   end if loop
(13) end for loop

ALGORITHM 7: Host selection.
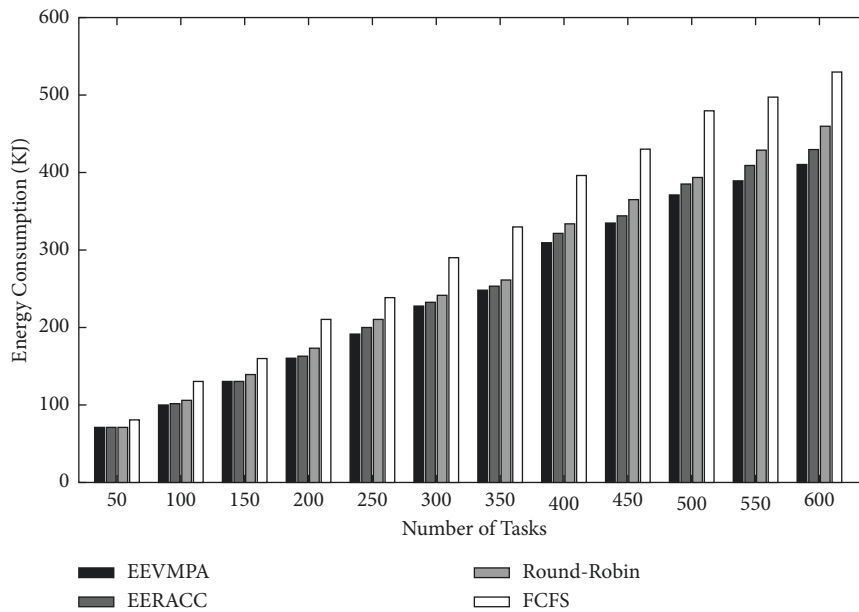
FIGURE 2: Energy consumption for EEVMPA, FCFS, Round-Robin, and EERACC.

workstation which has an Intel i3 1.19 GHz CPU and 12 GB of memory. During simulations, several cloud computing resources are used, such as user request tasks, VMs, and hosts that are heterogeneous in nature. The length of service requests by users and the requirement of resources are randomly created. During simulation, for every test case, the number of hosts is kept fixed, and the number of VMs (VMs) is kept in the range of 10 to 200. The VM resources are randomly generated and should satisfy the condition that these resources should be less than the capacity of the host on which they are deployed. In terms of calculating performance, we compare all VMs' energy consumption with respect to the entire server's energy consumption one by one. To avoid anomalies, 10 different runs of the entire

algorithms for the specific user service request tasks are implemented. In this experiment, we have used four different ranges for the length of tasks. The ranges are as follows: (3500–5500), (6500–8500), (8500–10,000), and (10,000–11,500). Each and every input task in the set contains 25% of the tasks from any one of the ranges of the task length mentioned before. In this experiment, we compare the developed EEVMPA with the other existing VM placement algorithms. Those algorithms are (1) FCFS (First Come, First Serve) Algorithm, (2) Round-Robin Algorithm, and (3) Energy-Efficient Resource Allocation (EERACC Algorithm). Then, many graphical comparative analyses (bar-graph representation) of these algorithms, including EEVMPA, is made. In terms of energy consumption in these
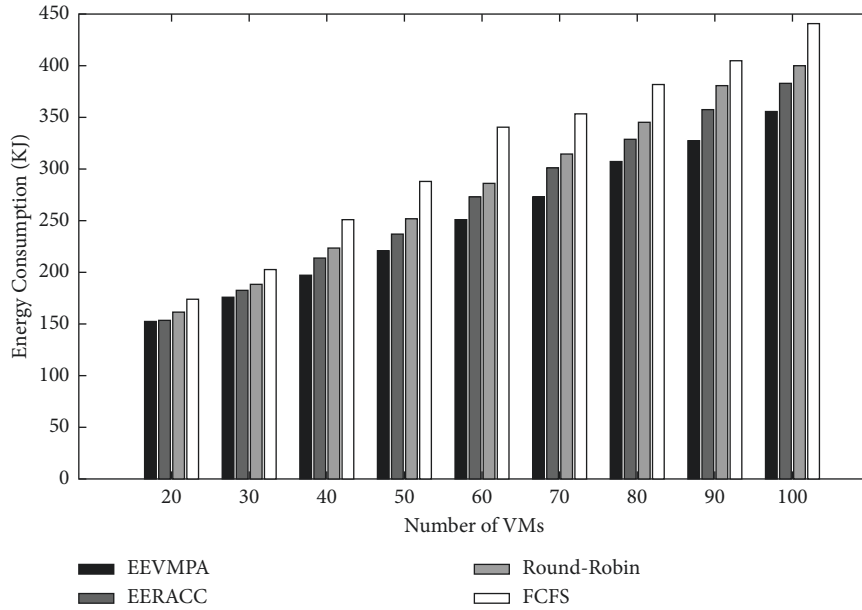
Figure 3: Energy consumption for EEVMPA, FCFS, Round-Robin, and EERACC.
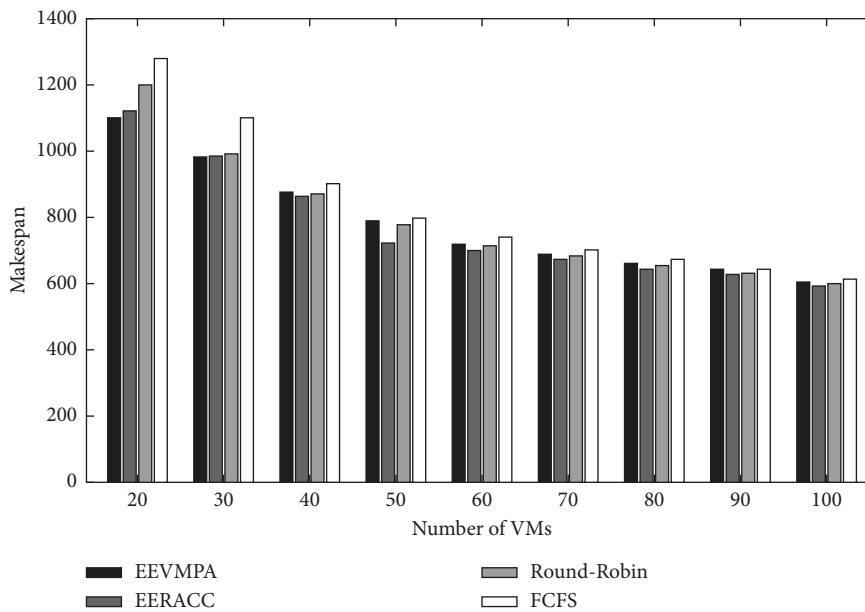
Figure 4: Makespan time for EEVMPA, FCFS, Round-Robin, and EERACC.

algorithms, we implemented graphical analysis based on two different conditions. In condition 1, we have compared all four algorithms when the number of VMs is fixed, and the number of tasks is varied. Figure 2 shows the graphical comparison of energy consumption for EEVMPA, FCFS, Round-Robin, and EERACC when the number of tasks is varied, and the number of VMs is fixed. From the figure, we can see that EEVMPA consumes less energy as compared to the rest of the VM placement algorithms when the number of tasks increases, and the number of VMs is kept fixed. In condition 2 in terms of energy consumption, we have compared all four algorithms when the number of VMs is varied, and the number of tasks is fixed.

The graphical analysis of this condition is shown below. From Figure 3, we can see that when the number of tasks is fixed and the number of VMs is varied, EEVMPA consumes less energy than the other VM placement algorithms. In terms of makespan time in these algorithms, we implemented graphical analysis based on two different conditions. In condition 1, we have compared all four algorithms when the number of VMs is fixed, and the number of tasks is varied.

Later, this situation is solved as the number of tasks increases, the EEVMPA is able to adapt and then provides less makespan time compared to other algorithms. This is especially true when the number of tasks is varied, and the
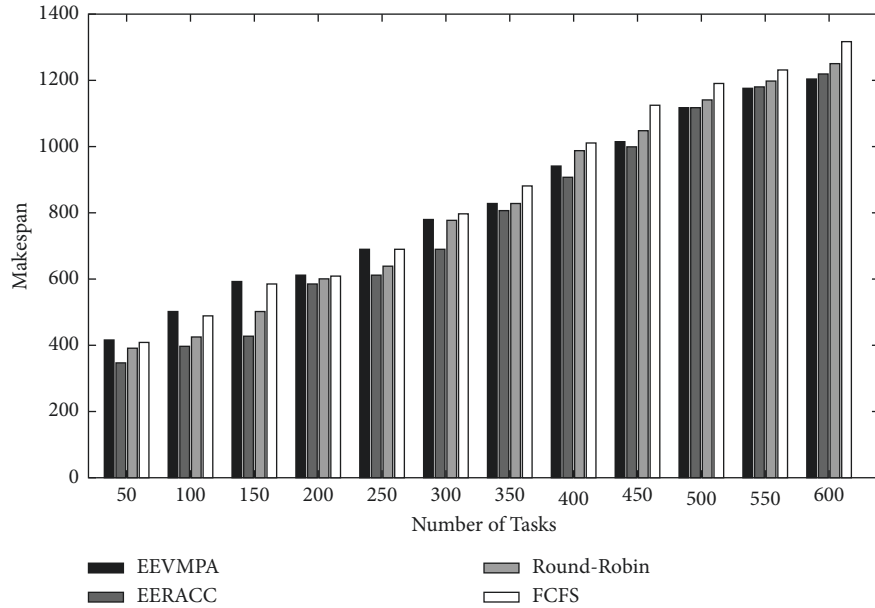
FIGURE 5: Makespan time for EEVMPA, FCFS, Round-Robin, and EERACC.

number of VMs is fixed. In condition 2, in terms of makespan time, we have compared all four algorithms when the number of VMs is varied, and the number of tasks is fixed. Figure 3 depicts the comparison of energy consumption for EEVMPA, FCFS, Round-Robin, and EERACC when the number of tasks is fixed and the number of VMs is varied.

EEVMPA initially performs worse compared to other algorithms as it takes maximum makespan time. This is because EEVMPA attempts to use the minimum number of hosts possible, but in general, cloud systems take a greater number of tasks, so this situation occurs.

Figures 4 and 5 depict the graphical comparison of makespan time for EEVMPA, FCFS, Round-Robin, and EERACC when the number of tasks is varied and the number of VMs is fixed. From Figure 5, we can see that at first, EEVMPA takes less time to make and performs better than other algorithms we compared. As the number of VMs grows, all the algorithms come to the same level of makespan. This is because the longer the VMs work, the performance slows down significantly, and hence, this result occurs when the number of tasks is fixed but the number of VMs is varied.

8. Conclusions

The main aim of this algorithm is to minimize energy consumption and reduce makespan time, which is crucial for getting the best performance from the data center. Both targets were successfully achieved in this work. In this research, we compared EEVMPA with other existing VM placement algorithms such as the First Come, First Serve (FCFS) algorithm, the Round-Robin algorithm, and the Energy Efficient Resource Allocation (EERACC) algorithm and did graphical analysis among them to determine which algorithm performed the best in each condition. Our

method is found to be the best in economizing energy consumption, related power costs, and reducing makespan time.

Data Availability

The manuscript contains all the data used in the study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] S. Debnath, B. Soni, U. Baruah, and D. K. Sah, "Text-dependent speaker verification system: a review," in *Proceedings of the 2015 IEEE 9th International Conference on Intelligent Systems and Control (ISCO)*, pp. 1–7, IEEE, Coimbatore, India, January 2015.

[2] D. K. Sah, C. Shivalingagowda, and D. P. Kumar, "Optimization problems in wireless sensors networks," *Soft computing in wireless sensor networks*, vol. 2018, pp. 41–62, 2018.

[3] D. N. K. Jayakody, K. Srinivasan, and V. Sharma, *5G Enabled Secure Wireless Networks*, Springer International Publishing, Cham, 2019.

[4] H. Hu, Z. Li, H. Hu et al., "Multi-objective scheduling for scientific workflow in multicloud environment," *Journal of Network and Computer Applications*, vol. 114, pp. 108–122, 2018.

[5] C. Georgios, F. Evangelia, M. Christos, and N. Maria, "Exploring cost-efficient bundling in a multi-cloud environment," *Simulation Modelling Practice and Theory*, vol. 111, Article ID 102338, 2021.

[6] Y. Kang, L. Pan, and S. Liu, "Job scheduling for big data analytical applications in clouds: a taxonomy study," *Future Generation Computer Systems*, vol. 135, pp. 129–145, 2022.

[7] S. A. Zakaryia, S. A. Ahmed, and M. K. Hussein, "Evolutionary offloading in an edge environment," *Egyptian Informatics Journal*, vol. 22, no. 3, pp. 257–267, 2021.

[8] S. K. Panda and P. K. Jana, "Efficient task scheduling algorithms for heterogeneous multi-cloud environment," *The Journal of Supercomputing*, vol. 71, no. 4, pp. 1505–1533, 2015.

[9] M. H. Mousa and M. K. Hussein, "Efficient UAV-based mobile edge computing using differential evolution and ant colony optimization," *PeerJ Computer Science*, vol. 8, p. e870, 2022.

[10] J. P. J Gabhane, S. Pathak, and N. M. Thakare, "Metaheuristics algorithms for VM placement in cloud computing environments—a review," *Computer Networks, Big Data and IoT*, vol. 2021, pp. 329–349, 2021.

[11] K. Karmakar, R. K. Das, and S. Khatua, "An ACO-based multi-objective optimization for cooperating VM placement in cloud data center," *The Journal of Supercomputing*, vol. 78, no. 3, pp. 3093–3121, 2022.

[12] S. Omer, S. Azizi, M. Shojafar, and R. Tafazolli, "A priority, power and traffic-aware virtual machine placement of IoT applications in cloud data centers," *Journal of Systems Architecture*, vol. 115, Article ID 101996, 2021.

[13] N. Raman, A. B. Wahab, and S. Chandrasekaran, "Computation of workflow scheduling using backpropagation neural network in cloud computing: a virtual machine placement approach," *The Journal of Supercomputing*, vol. 77, no. 9, pp. 9454–9473, 2021.

[14] H. O. Salami, A. Bala, S. M. Sait, and I. Ismail, "An energy-efficient cuckoo search algorithm for virtual machine placement in cloud computing data centers," *The Journal of Supercomputing*, vol. 77, no. 11, pp. 13330–13357, 2021.

[15] S. Gharehpasha, M. Masdari, and A. Jafarian, "Virtual machine placement in cloud data centers using a hybrid multiverse optimization algorithm," *Artificial Intelligence Review*, vol. 54, no. 3, pp. 2221–2257, 2021.

[16] A. Bouhank and M. Daoudi, "Non-dominated ranking Biogeography based optimization algorithm for VM placement in cloud computing," in *Intelligent Computing*, pp. 423–438, Springer, Cham, 2022.

[17] M. H. Mousa and M. K. Hussein, "Efficient UAV-based MEC using GPU-based PSO and voronoi diagrams," *CMES-Computer Modeling in Engineering & Sciences*, vol. 2022, 2022.

[18] S. M. Farzaneh and O. Fatemi, "A novel virtual machine placement algorithm using RF element in cloud infrastructure," *The Journal of Supercomputing*, vol. 78, no. 1, pp. 1288–1329, 2022.

[19] M. Ghetas, "A multi-objective Monarch Butterfly Algorithm for VM placement in cloud computing," *Neural Computing & Applications*, vol. 33, no. 17, pp. 11011–11025, 2021.

[20] B. N. Gohil, S. Gamit, and D. R. Patel, "Fair fit—a load balance aware VM placement algorithm in cloud data centers," in *Advances in Communication and Computational Technology*, pp. 437–451, Springer, Singapore, 2021.

[21] X. Chen, Y. Chen, A. Y. Zomaya, R. Ranjan, and S. Hu, "CEVP: cross entropy based virtual machine placement for energy optimization in clouds," *The Journal of Supercomputing*, vol. 72, no. 8, pp. 3194–3209, 2016.

[22] A. Jayanetti, S. Halgamuge, and R. Buyya, "Deep reinforcement learning for energy and time optimized scheduling of precedence-constrained tasks in edge-cloud computing environments," *Future Generation Computer Systems*, vol. 137, pp. 14–30, 2022.

[23] I. Casas, J. Taheri, R. Ranjan, L. Wang, and A. Y. Zomaya, "GA-ETI: an enhanced genetic algorithm for the scheduling of scientific workflows in cloud environments," *Journal of computational science*, vol. 26, pp. 318–331, 2018.

[24] A. Hameed, A. Khoshkbarforoushha, R. Ranjan et al., "A survey and taxonomy on energy efficient resource allocation techniques for cloud computing systems," *Computing*, vol. 98, no. 7, pp. 751–774, 2016.

[25] H. Kloh, B. Schulze, R. Pinto, and A. Mury, "A bi-criteria scheduling process with CoS support on grids and clouds," *Concurrency and Computation: Practice and Experience*, vol. 24, no. 13, pp. 1443–1460, 2012.

[26] L. Deng, Q. Yu, and J. Peng, "Adaptive scheduling strategies for cloud-based resource infrastructures," *Security and Communication Networks*, vol. 5, no. 10, pp. 1102–1111, 2012.

[27] A. Mosa and N. W. Paton, "Optimizing virtual machine placement for energy and SLA in clouds using utility functions," *Journal of Cloud Computing*, vol. 5, no. 1, p. 17, 2016.

[28] D. More, S. Mehta, P. Pathak, L. Walase, and J. Abraham, "Achieving energy efficiency by optimal resource utilisation in cloud environment," in *Proceedings of the2014 IEEE International Conference on Cloud Computing in Emerging Markets (CCEM)*, pp. 1–8, IEEE, Bangalore, India, October 2014.

[29] A. Belgacem, S. Mahmoudi, and M. Kihl, "Intelligent multi-agent reinforcement learning model for resources allocation in cloud computing," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 6, pp. 2391–2404, 2022.

[30] C. Mastroianni, M. Meo, and G. Papuzzo, "Probabilistic consolidation of virtual machines in self-organizing cloud data centers," *IEEE Transactions on Cloud Computing*, vol. 1, no. 2, pp. 215–228, 2013.

Research Article

An Improved Load Forecasting Method Based on the Transfer Learning Structure under Cyber-Threat Condition

Luo Zhao [ID],[1] **Xinan Zhang,**[2] **Yifu Chen,**[3] **Xiuyan Peng,**[1] **and Yankai Cao** [ID][4]

[1]*College of Intelligent Systems Science and Engineering, Harbin Engineering University, Harbin 150001, China*
[2]*Department of Electrical, Electronic and Computer Engineering, The University of Western Australia, Perth 6009, Australia*
[3]*Department of Chemical and Biological Engineering, University of Wisconsin-Madison, Madison 53706, USA*
[4]*Department of Chemical and Biological Engineering, University of British Columbia, Vancouver V6T 1Z3, Canada*

Correspondence should be addressed to Luo Zhao; zhaoluo425@163.com and Yankai Cao; yankai.cao@ubc.ca

Smart grid is regarded as an evolutionary regime of existing power grids. It integrates artificial intelligence and communication technologies to fundamentally improve the efficiency and reliability of power systems. One serious challenge for the smart grid is its vulnerability to cyber threats. In the event of a cyber attack, grid data may be missing; subsequently, load forecast and power planning that rely on these data cannot be processed by generation centers. To address this issue, this paper proposes a transfer learning-based framework for smart grid scheduling that is less reliant on local data while capable of delivering schedules with low operating cost. Specifically, the proposed framework contains (1) a power forecasting model based on transfer learning which can provide high quality load prediction with limited training data, (2) a novel adaptive time series prediction method with modeling time series from a covariate shift perspective that aims to train the forecasting model with a strong generalization capability, and (3) a day-ahead optimal economic power scheduling model considering a shared energy storage station.

1. Introduction

In recent years, the emergence of renewables and big data has prompted a reform of the electrical network. As a consequence, the concept of a smart grid becomes increasingly popular [1]. A smart grid is defined as the next generation electrical grid with power-flow control, self-healing, and energy reliability using digital communications. Compared to the conventional power system, the smart grid is designed to integrate millions of smart sensors and advanced computing technologies into the whole grid [2]; it can efficiently realize real-time automatic control, intelligent regulation, online analysis and decision making, cooperative interaction, and other advanced functions of the power grid. One feature of the smart grid is the high share of renewable generation, which poses a threat to its reliability due to the renewable intermittency. One widely adopted solution to this problem is to employ the energy storage system (ESS) in microgrids. Since wind and photovoltaics power are

nondispatchable parts, the dispatching strategy of ESS becomes an important component in the smart grid with renewable generation.

Most existing energy storage planning approaches rely on historical data or highly precise generation/load forecast [3, 4]. Current power forecasting methods can be broadly categorized into two types: methods based on statistical analysis [5] and methods based on artificial intelligence (AI) algorithms [6]. For methods in the first category, the Bayesian theory, multivariable linear regression, and autoregressive moving average (ARIMA) are often employed [7, 8]. These methods are better suited for fitting data that have periodic features; the high percentage of renewable energy sources considerably increases the randomness of power variations, making methods in this category less suitable for applications in smart grids. Methods based on AI algorithms, on the other hand, are theoretically favorable in predicting an output for systems with high nonlinearity and complex dynamic properties. In particular,

a recurrent neural network (RNN) [9] helps handle non-linear problems and capture more dynamic relationships between the input and the forecasted output compared to the artificial neural network (ANN) [10], and the long short-term memory (LSTM) unit proposed in [11] further improves its performance in the prediction of time series data.

A noteworthy concern is that all the abovementioned methods require sufficient training data. The heavy reliance on information networks leads to higher cyber risks for smart grids [12]. In the event of a cyber-attack, the local database could be tampered and lost, which leads to serious consequences [13]. For instance, Ukraine's electricity supply system was hacked in 2017 through an attack on the data aggregator, which is the central node containing data from data collection base stations [14]. This event caused massive power outages, paralyzed connected nodes, and prevented the control center from accessing customer load data in time for power control and scheduling operations. It is of great interest to reduce the grid control centers' heavy reliance on local users' data, so that the day-ahead or long-term power dispatch would not be blocked by insufficient data. The primary issue that needs to be addressed is load forecasting based on inadequate local data.

Transfer learning (TL) [15, 16] is a suitable tool for addressing the challenges discussed above. It is of great value to introduce transfer learning in power systems to efficiently utilize resources from different regions, discover the commonality of different datasets, and establish transfer learning-based forecasting methods. The key idea of transfer learning is to use the existing experience to solve similar tasks, exploit similarities between data and models, and apply the trained content to new tasks. Specifically, transfer learning allows the use of knowledge in the dataset with complete labels (i.e., the source domain) to solve problems in the dataset with missing labels (i.e., the target domain), using a trained model with good generalization capability [17]. The research in [18] presents the outstanding contribution of transfer learning in the field of image processing. Innovative breakthroughs have also been made in the field of classification and target detection [19] in recent years. Lu et al. [20] proposed a general transfer learning-based framework for load forecasting with limited data. The influence of adopting different kernel functions in transfer learning for fault diagnosis is studied by Li et al. in [21]. Scholars in [22] investigate the superiority of transfer learning in extracting features and aim to predict the wind speed in different environments. Yin et al. [23] proposed a hybrid transfer learning-based wind power forecasting model. Unfortunately, the potential relationship between statistical properties in time series and transfer learning is ignored in these works.

One critical issue for developing transfer learning-based forecasting methods is how to train a forecasting model with strong generalization capabilities. Many published forecasting methods for smart grids are based on the assumption that historical data follow the same distribution. The scholars in [24] made great improvements in load forecast based on machine learning in certain areas, but the generalization ability of their proposed method is not very promising since the differences between distributions of data are not considered. In a typical grid, especially those with high penetration of the renewable system that introduces high stochasticity, the distribution of the data in the temporary structure changes over time. Consider the illustrative graph in Figure 1, the probability distribution of P_x varies for different intervals, and the temporal covariate shift phenomenon happens after adding a new segment of data, where $P_a \neq P_b \neq P_c \neq P_{test}$. Here, the aforementioned issue is embodied in two aspects. First, how to build an adaptive prediction model to weaken the effect of covariate shift and accommodate the diversity of sample data. Second, how to develop a probability distribution algorithm to minimize the divergence between the distribution for different intervals.

Load prediction provides basic data for generation planning, day-ahead market offers, and intraday market trading, and it is important for the economic dispatch of power system. In the wind/photovoltaic/energy storage complementary microgrid, the generation plan of energy storage is the only dispatchable part. Researchers in [25] proposed a general method for the capacity and power of energy storage batteries and constructed a capacity allocation scheme for energy storage batteries, researchers in [26] introduced the control and communication technology, and operation principle of cloud energy storage based on the Irish power system. At present, the research on shared energy storage is in its initial stage, and the existing work takes shared energy storage systems as the main research object to analyze the business model and profitability of shared energy storage systems, while in-depth research on the charging and discharging behavior and economic benefits of users' participation in the shared energy storage system is limited. This paper introduces shared energy storage plants among different user groups and establishes an optimal scheduling model with the objective to minimize daily operation cost of user groups.

The main contributions of this paper are summarized as follows:

(1) To address the challenge that the generation center fails to develop power planning for grid operation due inadequate local user data, we propose a power forecasting method based on transfer learning, where data from the source domain can provide valuable reference information. Additionally, case studies provide a detailed analysis of how to choose the appropriate source domain, and the effect of negative transfer on model performance is analyzed.

(2) This work proposes to model the time series of load prediction from the covariate shift perspective. To train a forecasting model with strong generalization capability with transfer learning, we generate a combination mode where the probability distributions of time series vary for different intervals, and an optimal split method is proposed to ensure that the segments being divided are the most dissimilar ones. The temporal distribution matching algorithms are proposed to minimize the divergence between the distribution for different intervals. Dynamic
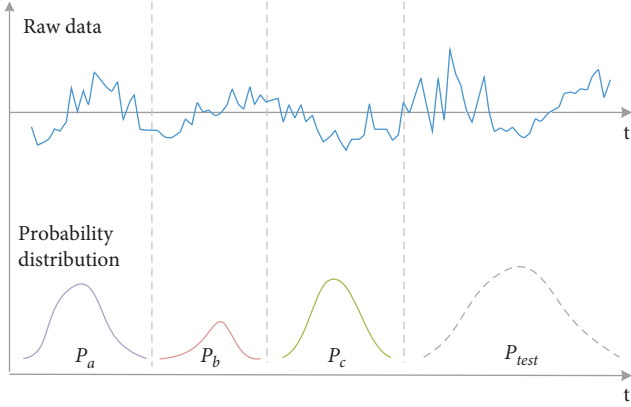
FIGURE 1: The temporal covariate shift phenomenon.

programming (DP) is applied to optimize the optimal division points. The case study shows that the maximum improvement of the proposed forecasting method is up to 52.8% in mean absolute percentage error (MAPE) compared to other transfer learning-based methods, and up to 64.4% compared to the traditional method.

(3) With the accurate load prediction obtained from (1) and (2), a novel shared energy storage station (ESS) concept is proposed to form a framework for optimal scheduling based on the transfer learning method. The case study shows that the proposed framework can reduce the overall operating cost of the microgrid and maximize the benefit of grid operation, thus addressing the issue of energy curtailment [27] as well as the high cost of energy storage.

This paper is organized as follows: Section 2 presents the structure of the transfer learning-based forecasting algorithm with the limited data set, and modeling time series from a covariate shift perspective, which is critical to determining the generalization ability of forecasting models. Section 3 proposes a framework of optimal dispatch planning for the distributed microgrid, a shared energy storage station (ESS) is formed as a solution for multisource power grid scheduling, and the dispatchability of the cyber-attacked

area is analyzed when using the proposed optimal economic dispatching method. The performance of transfer learning in addressing fragmented test data of the target domain is then developed. Moreover, using data obtained from the proposed forecasting method, the economic analysis of the proposed energy storage station is shown in Section 4. Finally, the main findings are included in Section 5.

2. Methodology of the TCS-Transfer Learning Model

2.1. Transfer Learning-Based Structure. In most machine learning tasks, the training and test sets come from the same feature space and are subject to the same probability distribution. When the actual conditions are not satisfied, it takes a lot of resources to collect data from the target domain to retrain a model. In particular, in a feature-rich smart grid, retraining the model becomes inefficient and time-consuming, which can seriously disrupt the schedule of the generation center, transfer learning [28] is a new solution to this problem.

Maximum mean discrepancy (MMD) is applied in this paper to define a more specific formula for the TL problem:

$$f^* = \arg\min_{f \in H} \frac{1}{N_s} \sum_i^{Ns} l(f(x_i), y_i) + \xi MMD(D_s, D_t), \quad (1)$$

where H represents the vector space that satisfies the objective function, N_s means the number of samples of the source domain D_s, and D_t means the source domain and target domain. N_t represents the number of samples of the target domain, and $l(x, y)$ denotes the loss function suitable for different structures, and x and y represent the sample and label, respectively. Argmin denotes the value of the variable that minimizes the objective function. ξ represents the metric coefficient. This paper selects root mean squared error (RMSE) as the measurement of error. It is considered that MMD is one of the most widely used distance metrics in TL among many statistical measures, which is an effective way to measure the correlation between any two different domains D_i and D_j, and it can be formulated using the following equations:

$$MMD[\mathbb{F}, D_i, D_j] = \left[\frac{1}{m(m-1)} \sum_{i \neq j}^m k(x_i, x_j) + \frac{1}{n(n-1)} k(y_i, y_j) - \frac{2}{mn} \sum_{i,j=1}^{m,n} k(x_i, y_j) \right]^{1/2}, \quad (2)$$

$$k(x, x') = e^{-\|x - x'\|^2/2\sigma^2}, \quad (3)$$

where D_i and D_j represent any two different domains, and their sample sizes are m and n. $k(\cdot, \cdot)$ represents the Gaussian kernel function. Radial basis kernel function $k(x, x')$ is used for mapping to the higher dimensional spaces. x' is the center point of the kernel function, σ denoted the

expectation, which controls the range of action of the Gaussian kernel function; the larger its value, the larger the local range of the influence of the Gaussian kernel function. $\mathbb{F}$ denotes the unit ball in the regenerative nuclear Hilbert space (RNHS).

It is worth noting that not all source domains are suitable to be selected, Krizhevsky et al. [29] provides an in-depth analysis of the role of pretraining models for transfer tasks, a pretraining model can be used as benchmark models for the task of the target domain, researchers in [30, 31] also demonstrate that for datasets with different distributions, pretraining in an appropriate source domain can greatly improve the accuracy of the results, and source domain that has a weak correlation with the target domain may cause a negative transfer phenomenon. Therefore, it is essential to analyze the discrepancy between the $\mathscr{D}_s$ and $\mathscr{D}_t$; this paper measures the difference by calculating the MMD value, and the source domain can be selected for pretraining if the MMD is below the preset threshold value.

2.2. Problem Formulation of Temporal Covariate Shift (TCS).

The variate $\{x_i\}$ of the time series is assumed to follow the same probability distribution in most existing prediction methods, it may have achieved satisfactory results on specific datasets, such as the load of a stand-alone device in a traditional predict scene, which is relatively uncomplicated in its diversity. However, this assumption is not realizable in the actual application due to the huge amount of data and features, the variation of data distributions with the time changing cannot be ignored. Thus, our problem can be formulated as follows: split a given time series $\mathbb{S}$ with m labeled dataset into k segments with the most dissimilar distribution $\{\mathbb{S}_1 \dots \mathbb{S}_k\}$.

$$\begin{cases} \mathbb{S} = \{\mathbb{S}_1, \dots \mathbb{S}_p \dots \mathbb{S}_q \dots, \mathbb{S}_k\}, \\ \sum \mathbb{S} = m. \end{cases} \tag{4}$$

With reference to the definition of covariate drift [32] in the classification field, the temporal covariate shift can be presented such that the whole intervals in the same period i follow the same probability distribution $P_{\mathbb{S}_{p,q}}(x, y)$. The distribution will be different when the time period changes $P_{\mathbb{S}_p}(x) \neq P_{\mathbb{S}_q}(x)$ $(p \neq q \in (1, k))$. To train a prediction model with excellent generalization performance under temporal covariate shift, the main issue is to capture the common knowledge shared among different periods of $\mathbb{S}_{p,q}$.

According to the principle of maximum entropy [33, 34], finding intervals that are the most distinct from each other can help maximize the capacity of shared information within a time series under temporal covariate change issues. It is fair to make distributions of each interval as diverse and feasible to maximize the entropy of the overall distributions of an array. This enables for more generic and adaptable future data modeling. Worst-case training using the original sequence enables the model to cope with the stochastic nature of the unknown data. Figure 2 shows the structure of the proposed predicting method based on transfer learning, and the primary task is to split the time series into k segments. The splitting problem of the time series in (5) is solved by the greedy algorithm.

$$\max_{m_1, \dots m_k} \frac{1}{k} \sum_{1 \leq \lambda_i \neq \lambda_j \leq k} MMD(\mathbb{S}_p, \mathbb{S}_q), \tag{5}$$

$$s.t. \forall i, \quad \Delta_1 \leq |\lambda[j] - \lambda[i]| \leq \Delta_2,$$

where array $\lambda[1 \dots m]$ represents a set of time series, m represents the length of it, $MMD(\cdot, \cdot)$ means the distribution-based distance function to measure the distance between distributions of any two segments $\mathbb{S}_p, \mathbb{S}_q$, $\lambda[c], c \in [i, j]$ denotes the coordinate position of the cut at each time. $\lambda[i]$ and $\lambda[j]$ are the left and right endpoints of the segment to be segmented, respectively. $\lambda[j]_{-1}$ is the right endpoint of the previous interval and $\lambda[i]$ is the start point of the next interval. Δ_1 and Δ_2 represent predefined constants to prevent trivial solutions.

2.3. Matching Process and Fine-Tune.

The proposed method is designed to acquire the common information shared by distinct intervals by comparing their probability distributions, this section presents the process of how to pretrain a model after obtaining the optimal split intervals in raw data. In comparison to approaches that simply depend on local or statistical knowledge, the pretraining model $\mathscr{M}$ can produce a nice generalization on unknown datasets of the target domain. The pretraining issue formulated in Section 2.2 can be solved by the domain generalization (DG) [35] method, and the distribution matching loss of the network can be established as follows:

$$\theta^* = \arg\min_\theta \sum_{i=1}^{k} \mathscr{L}_{\text{pre}}(y_i, \hat{y}_i) + \xi \sum_{1 < i < j < k} MMD(\mathbb{S}_i, \mathbb{S}_j), \tag{6}$$

where $\mathscr{L}(\cdot, \cdot)$ denotes the MSE loss in the source domain, and $\mathbb{S}_i$ and $\mathbb{S}_j$ are any two different intervals of $\lambda[1 \dots m]$. In the proposed method, LSTM is employed as the main body of the network structure. Due to the special memory unit of LSTM, potential relationships between data in a time series can be preserved, which can provide high accuracy for the prediction results. Compared to the conventional recurrent neural network algorithm, which contains only one state h_t, the LSTM structure introduces cell states c_t to develop potential relationships in a long time series. The LSTM structure can be formulated as follows:

$$\begin{aligned} i_t &= \sigma(W_i * [h_{t-1}, x_t] + b), \\ \tilde{c}_t &= \tanh(W_c * [h_{t-1}, x_t] + b), \\ c_t &= f_t \odot c_{t-1} + i_t * \tilde{c}_t, \\ o_t &= \sigma(W_o[h_{t-1}, x_t] + b), \\ h_t &= o_t * \tanh(c_t), \end{aligned} \tag{7}$$
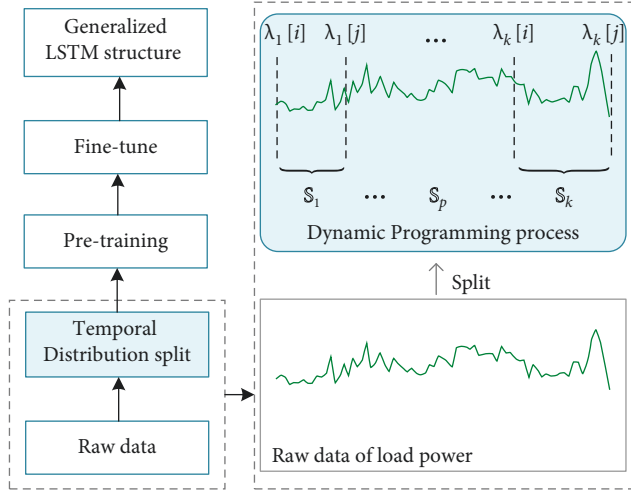
FIGURE 2: Schematics of the proposed predicting structure based on transfer learning.

FIGURE 3: LSTM unit structure.

FIGURE 4: Power forecasting structure based on the TCS-transfer learning model.

where x_t is the input data, $W_{i,c,o}$ represent the weight matrices, i_t, f_t, and o_t are the input, forget, and output gates of the LSTM structure, respectively. b represents the bias value. c_{t-1} denotes the state of the memory cell. The candidate value $\tilde{c}_t$ is generated by tanh layer. h_t presents the output value, h_{t-1} represents the output value of the previous unit, the sigmoid function is denoted as σ, and $*$ represents the dot product. $\odot$ is the element-wise product. Figure 3 shows the structure of LSTM; the type of data being inputted is described in Section 4.

Figure 4 shows the flowchart of the proposed transfer learning-based forecasting structure. The implementation steps of the whole network are as follows:

(1) Start by collecting the raw data of power fluctuations from neighboring cities and calculating the probability distribution of the candidate datasets

(2) Calculate the distance between the source and target domains according to eq. (2) and select the appropriate source domain as the input data of the pretrain model

(3) For the selected source domain, it is first divided into k most dissimilar segments using the proposed dynamic programming-based method according to eqs. (4) and (5).

(4) The k segments are considered as different domains. Eq. (6) is used as the new loss function in the model with LSTM as the main network. After obtaining a prediction network with strong generalization capability, the target domain with very less data is used to fine-tune by using Eq. (1)–(3).

The proposed temporal covariate shift issue focuses on an easily neglected problem in time series Dynamic programming solves the optimal spilt problem of the time series, considering the fragment as k independently distributed individuals. The generalization ability of the pretraining model using source domain data is greatly improved by taking the differences between these segments as an additional consideration in the loss function. The deep LSTM network can be trained with less learning time while overcoming the problem of lacking sufficient local data as training datasets.
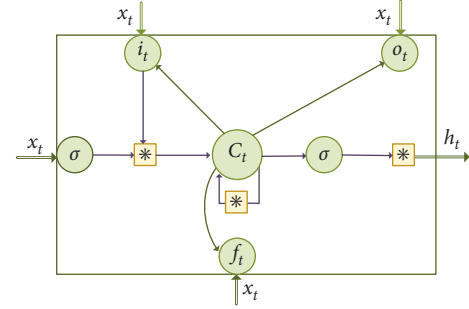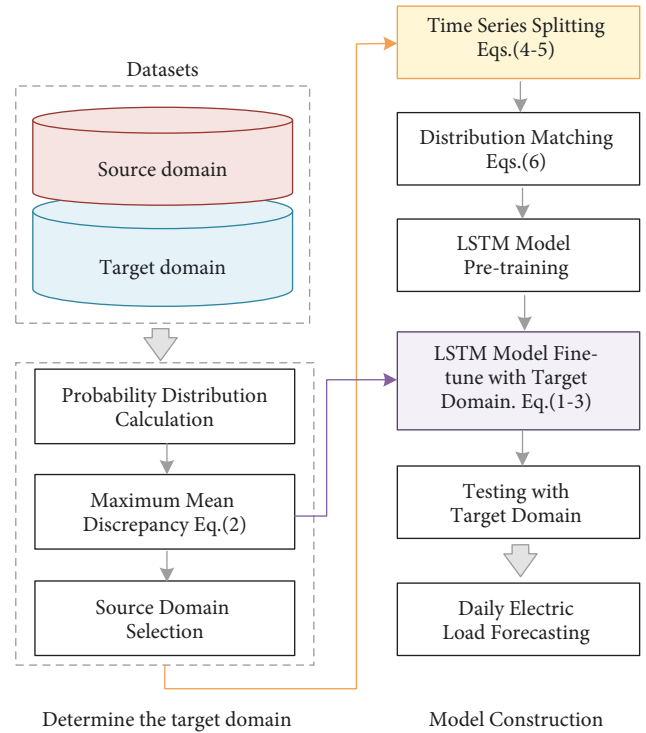
3. Optimal Dispatch Planning for Distributed Microgrid

3.1. Smart Grid with Multiple Sources of Energy Supplies. Smart grids are increasingly used in recent years for both residential and industrial purposes. In many microgrids, the hybrid wind-solar generation plant is employed owing to the complementary nature of wind and solar generation patterns. Therefore, the smart grid with renewable generations is studied in this paper as shown in Figure 5. From the perspective of the Internet of Things (IoT) network structure of the smart grid, the data is collected by a large number of
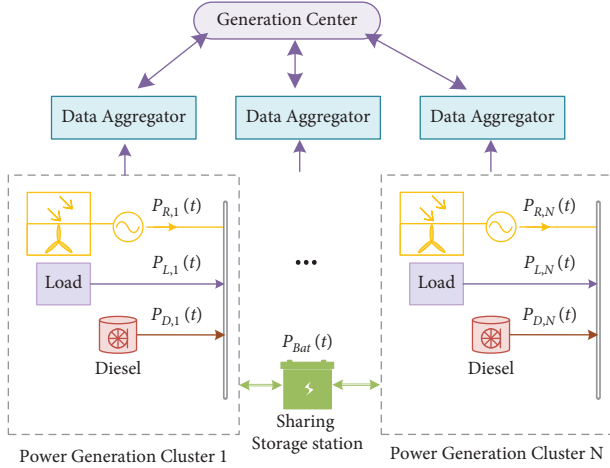
FIGURE 5: Conceptual framework of the multienergy complementary power generation system.

smart meters in the users' network, encrypted by the gateway, gathered by the data aggregator, and then transmitted to the control center, and this network data distribution characteristic is called the "funnel effect" [36]. In a smart grid with deeply integrated information systems, the form of fault propagation has more possibilities. When a node in the network fails, it will trigger the failure of the related nodes, and the existing scheme requires all users to work collaboratively. As long as one user fails, the whole system cannot operate normally. The transfer learning-based forecasting method proposed in this paper can substantially reduce the dependence of generation centers on the adequacy of local data.

The generation plant of the microgrid is divided into N clusters, each having wind turbine and solar photovoltaic (PV) systems. The output power of the wind/PV system in the j-th cluster can be expressed as follows:

$$P_{G,j}(t) = P_{W.j}(t) + P_{PV,j}(t) + P_{D,j}(t), \quad j \in \{1, 2, \ldots, N\}, \tag{8}$$

where $P_{w,j}$, $P_{PV,j}$, and $P_{D,j}$ represent the output power of wind turbines, photovoltaic systems, and diesel systems, respectively. In modern microgrids, generators are usually used as backup power sources, the term $P_{D,j}(t)$ in (8) can be ignored if no power failure occurs. It can be further expressed in per unit (p.u.) value denoted by $P_{*G,j}(t)$ as eq. (9), where $P_{G,\text{base}}(t)$ denotes the base power in the DC bus, $P_{G,j}(t)$ is the actual power value of the DC bus, and $P_{\text{Bat}}(t)$ represents the power of the battery.

$$\begin{cases} P_{*G,j}(t) = \dfrac{P_{G,j}(t)}{P_{G,\text{base}}(t)}, \\ \\ P_{PV,j}(t) + P_{W.j}(t) + P_{D,j}(t) + P_{\text{Bat}}(t) = P_{L,j}(t). \end{cases} \tag{9}$$

3.2. Operation Mode and Market Rules of a Shared Station.
The concept of a shared energy storage plant is shown in Figure 5, in which the operator of an energy storage station

uses the financial advantage to establish a large shared energy storage plant among a group of customers, and unifies the operation and management of the energy storage station to provide shared energy storage services to multiple customers in the same distribution network area. The generation center forecasts the load based on historical electricity consumption data and plans to use shared energy storage plants for charging and discharging, which minimizes the economic cost of operation of the storage devices and saves the customer the investment cost of installing and maintaining the storage devices. Power market rules require the dispatch plan submitted to the grid operator 24 hours ahead.

Figure 6 depicts the framework of the proposed power dispatch schedule. The main purpose of this work is to solve the issue that the predicting process cannot operate due to insufficient local data, thus, the power data of the neighbor area is employed to train highly generalizable models, and the precise prediction model can be obtained after fine-tuning with few local data.

The generation center of the energy shared storage power station transmits the remaining power of users who need charging directly to users who need discharging according to the charging demand and discharging demand of each user at each time. The generation station of energy storage station can utilize the complementary nature of customers' power consumption behavior, i.e., the difference in power consumption behavior of the same customers at different times and different customers at the same times, and can maximize the investment in the least amount of energy storage to meet customers' demand for energy storage use. The specific optimization strategy of shared storage stations is detailed in Section 3.3.

3.3. Optimal Economic Scheduling of Shared Energy Storage Station.
Having explained the forecasting method and the possible operational modes of the microgrid, the next task is to develop the dispatch strategy for the shared ESS to economically benefit users. The high investment cost of energy storage is the main reason that limits the application of energy storage technology on the demand side of the grid. This paper proposed the concept of a shared energy storage station, as shown in Figure 7, which is applied to the economic optimization scheduling of regional users, and the minimum daily operating cost of the user group is achieved by coordinating the charging and discharging power of the users. The energy storage system allows users to store electricity during the grid valley hours and release it during peak hours, thereby decreasing electricity costs and relieving the pressure on regulating the peak load. According to the charging demand and discharging demand of each user in each period, the generation center will deliver the remaining electric energy of the user who needs to discharge directly to the user who needs to charge. If the total charging and discharging demand of users in the same time period is discharged, the regulation center will decide whether the users' electricity needs to be purchased by the main grid or stored in the shared power station according to the electrovalence at that time.
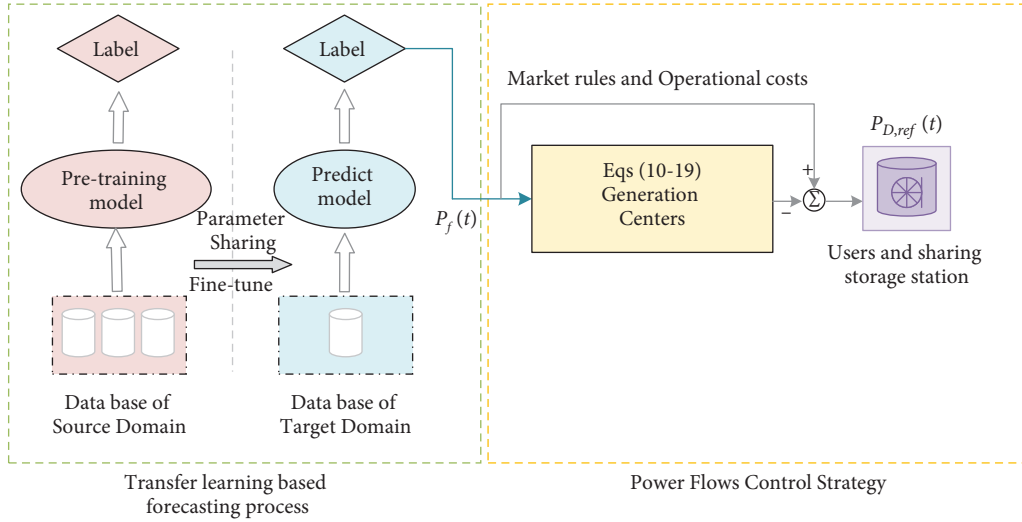
FIGURE 6: Schematics of the proposed optimal power planning procedure.
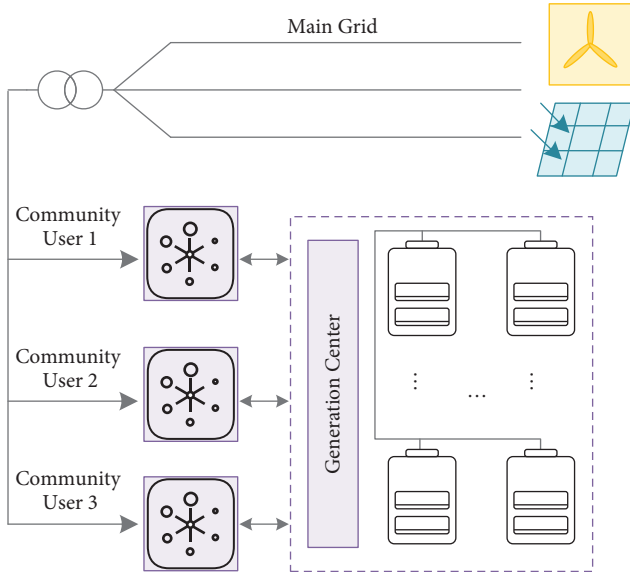
FIGURE 7: Schematic diagram of the shared energy storage station.

Based on the predicting method proposed in the previous sections, this part develops a scheme to obtain the minimum grid operating cost through proper charging and discharging of the shared energy storage station. By using the shared energy storage station, the user saves the investment costs for the installation and maintenance of energy storage devices. Users pay the service fee to the generation center in exchange for shared energy storage services. The service fee means the users pay to the generation center when they use the shared energy storage stations for charging and discharging, it is set as 0.16$/kWh.

3.3.1. Objective Function of the Optimization Model.
According to the market rules presented in Section 3.2, the user group connected to the shared energy storage station uses the typical daily operating cost optimization as the objective function to determine the capacity, the maximum charging and discharging power of the energy storage station, and the charging and discharging power of the storage station for each time period of the user. The daily operating cost of the customer group includes the cost of electricity purchased from the grid and the service fee paid to the energy storage station.

$$\min C = C_g + C_s, \tag{10}$$

where C represents the daily cost of electricity for the user community; C_g and C_s denote the cost of electricity purchased by the customer from the grid and the service fee paid to the energy storage station, respectively.

$$\begin{cases} C_g = \sum_{i=1}^{N} \sum_{t=1}^{T} \rho(t) \cdot P_{G,i}(t) \cdot \Delta t, \\ C_s = \sum_{i=1}^{N} \sum_{t=1}^{T} \delta(t) \cdot \left[P_{E,D,i}(t) + P_{E,C,i}(t) \right] \cdot \Delta t, \end{cases} \tag{11}$$

where N represents the serial number of the user group connected to the same shared energy storage station, user groups of three areas are selected as case studies in this paper ($i \in [1,3]$); $\rho(t)$ ($/kW \cdot h$) represents the price that the users purchase electricity from the grid. T denotes the scheduled time periods; $P_{G,i}(t)$ indicates the power value purchased by the user i from the grid at a given time interval t; Δt represents the unit time length of the power scheduled; $\delta(t)$ is the service fee of the shared energy storage station; $P_{E,D,i}(t)$ and $P_{E,C,i}(t)$ are discharge power and charging power of energy storage station at time t, respectively.

3.3.2. Constraint Condition.
The constrains should be met by the proposed power planning model. They include the following: electrical power balance constraints [26] and operational constraints of the energy storage stations:

(1) Power balance constrain of the whole grid:

$$P_G(t) + P_{PV,i}(t) + P_{W,i}(t) + P_{ESS}(t) - P_{L,i}(t) = 0,$$

$$P_{ESS} = P_{E,D,i}(t) - P_{E,C,i}(t) = \text{MAX}\left(P_{W,i}(t) + P_{PV,i}(t) - P_{L,i}(t)\right), \tag{12}$$

where $P_G(t)$ is the power purchased from the main network; $P_{PV,i}(t)$ represents the output power of the PV system of the i-th user in time interval t; $P_{W,i}(t)$ is the wind power of the i-th user in time interval t. $P_{L,i}(t)$ denotes the load power of the i-th user, which is a predicted value obtained by the method proposed in the previous section; P_{ESS} is the power of the energy storage station, which satisfies the maximum power difference between generation and load in the grid for any period of time; $P_{E,D,i}(t)$ and $P_{E,C,i}(t)$ represent the power under discharging and the charging status of the energy storage system.

(2) Charging and discharging power constraints for a shared energy storage station:

$$\begin{cases} 0 \le P_{E,D,i} \le P_{\max} \delta_{ESS\_D}^{\max}, \\[2mm] 0 \le P_{E,C,i} \le P_{\max} \delta_{ESS\_C}^{\max}, \\[2mm] \delta_{ESS\_D}^{\max} = \dfrac{SOC - SOC_{\min}}{SOC_l - SOC_{\min}}, \\[3mm] \delta_{ESS\_C}^{\max} = \dfrac{SOC_{\max} - SOC}{SOC_{\max} - SOC_h}, \end{cases} \tag{13}$$

$$\begin{cases} SOC = \left(SOC_0 - \dfrac{1}{Q \int_0^t \eta I dt}\right) \times 100\%, \\[3mm] SOC_{\min} \le SOC \le SOC_{\max}, \end{cases} \tag{14}$$

where $P_{\max}$ is the rated maximum power of the charging and discharging power of the energy storage station; $\delta_{ESS\_D}^{\max}$ and $\delta_{ESS\_C}^{\max}$ are defined as the discharging and charging state factor, respectively, which ensure that the energy storage is not in an overcharged and discharged state; $SOC_{\min}$ and $SOC_{\max}$ represent the operating range of the energy storage; SOC_l and SOC_h represent the optimal working interval for energy storage; SOC denotes the state of the charge value of the energy storage; SOC_0 is the initial state of charge; η is the charge and discharge efficiency; Q represents the electric charge quantity; I is the battery current.

(3) Power balance constrain of the energy storage station:

$$\sum_{i=1}^{N} \left|P_{E,D,i}(t) - P_{E,C,i}(t)\right| = \left|P_D(t) - P_C(t)\right|, \tag{15}$$

$$\begin{cases} \delta_{ESS\_D}^{\max} + \delta_{ESS\_C}^{\max} \le 1, \\[2mm] \delta_{ESS\_D}^{\max} \in \{0, 1\}, \delta_{ESS\_C}^{\max} \in \{0, 1\}. \end{cases} \tag{16}$$

The charging and discharging power of each user in a time period t needs to be balanced with the charging $P_C(t)$ and discharging $P_D(t)$ power of the energy storage station.

3.3.3. Resolve Method. The few nonlinear terms in the abovementioned constraints can significantly increase the difficulty and time during the solving process of numerically solving the aforementioned optimization problem. To overcome this challenge, Big-M [27] is adopted to linearize the nonlinear constraints in this work. The user scheduling model based on shared energy storage stations can be converted to a mixed integer linear programming problem. Thus, (15) can be reformulated as follows:

$$\begin{cases} 0 \le P_{E,D,i} \le P_{\max}, \\[2mm] 0 \le P_{E,D,i} \le \delta_{ESS\_D}^{\max} M, \\[2mm] 0 \le P_{E,C,i} \le P_{\max}, \\[2mm] 0 \le P_{E,C,i} \le \delta_{ESS\_C}^{\max} M, \\[2mm] \delta_{ESS\_D}^{\max} + \delta_{ESS\_C}^{\max} \le 1, \\[2mm] \delta_{ESS\_D}^{\max} \in \{0, 1\} \delta_{ESS\_C}^{\max} \in \{0, 1\}, \end{cases} \tag{17}$$

where the value of M is determined as 10^8 in this paper.

4. Case Study on the Smart Grid System

4.1. Setup and Experimental Result of the Proposed Forecasting Method. Field load data of three regions from Western Australia are utilized for the case studies, covering the time range of May 1, 2015, to July 1, 2021. A map of the regions of the presented case studies is shown in Figure 8, where Case 1 and Case 2 are two typical industrial type electricity consumption areas, and Case 3 is the residential user group. Each node represents a region of independent microgrids that do not interfere with each other and are powered primarily by renewable energy. It is noted that 80% of the historical data are used as the training set and the remaining 20% are used as a validation set. The train/valid sets are structured with a ratio of 8 : 2, the test set is from the real load data from users. When load forecasting is implemented, 168 steps are batched together to train the model and predict the net power in the next 24 steps, and the sampling interval is 1 hour and the forecasting horizon covers 24 hours. The input feature mainly consists of the population, temperature, and calendar data which contain the season, number of holidays, and weekdays. The features are summarized in Table 1.

Figure 9 shows the raw data collected from three cases presented in Figure 8. As shown by the yellow line, part of the data in the western mining smelter (Case 2) is absent because data loss occurred due to aggregator overload. The common deep learning-based predicting models are not universal for different data sources, and each region needs to
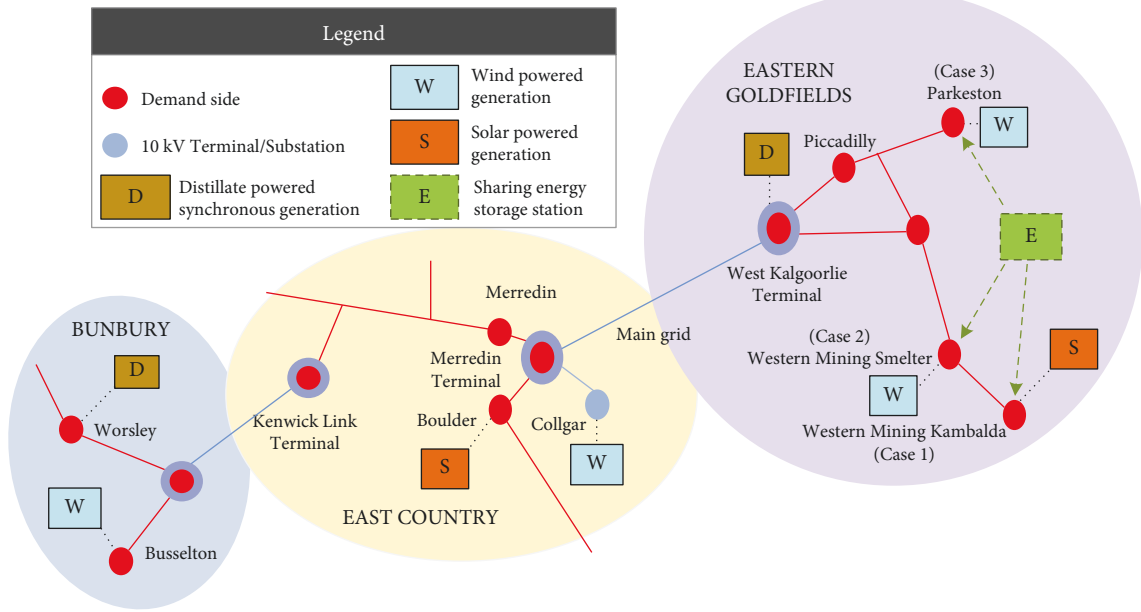
FIGURE 8: Geographical locations of the case study.

TABLE 1: Input feature of the forecasting model.

No.	Input feature	Description
1	No. of days	Integer
2	No. of holidays	Integer
3	No. of weekdays	Integer
4	Holiday length	Integer
5	Season	Binary
6	Population	Integer
7	Temperature	Integer
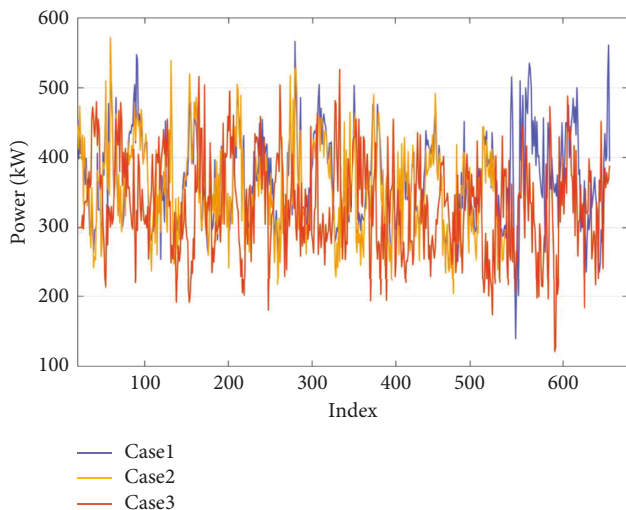
FIGURE 9: Raw data of three cases.

train a locally applicable predicting model based on its own database. According to the proposed transfer learning-based model, after comparing the MMD value of each database associated with the neighboring grid, the database of a microgrid in western mining Kambalda is chosen as the source domain.

To demonstrate the performance of the proposed structure, this work compares the proposed TCS-transfer learning model with four categories of methods listed as follows:

(i) Traditional time series forecasting model LSTM with classical gradient descent: there are also BP [37], ANN, and CNN [38] belonging to this category, and the most applicable time series predicting model–LSTM is selected in this category

(ii) The latest time series structure ELM [39] without backward propagation: the convergence time is substantially reduced and the training efficiency has been greatly improved. The first two categories are classical methods of time series forecasting.

(iii) Variants of popular domain adaptation methods include MEDA-LSTM [40] and MMD-RNN [41], which are also based on the concept of transfer learning

(iv) A branch of transfer learning is that a transformer with an attention mechanism [42, 43] has a stronger generalization capability than the classical method. This paper uses 6 encoder blocks of a transformer and 8 heads for self-attention.

The relevant parameters are listed in Table 2. The parameters that produce the best performance for each model are tuned by K-fold cross-validation. The following comparison discusses the effectiveness of the proposed TCS-transfer learning model mainly from two aspects: first, the power grid of the western mining smelter (Case 2) is with missing data, thus the traditional methods such as LSTM and ELM included in categories i and ii failed to forecast the load power due to the fragmented dataset in Case 2. For cases of missing data in the test set, transfer learning is the only method that can solve the problem. The first part of the

TABLE 2: Parameter set of different forecasting methods.

Model	Meaning	Value
TCS-transfer learning	Hidden layers	2
	Number of input layer nodes	4
	Number of hidden layer nodes	32
	Activation function	Sigmoid, tanh
	Learning rate	0.01
	Optimization function	ReLU
	Epochs of training	400
	Distance function	MMD
	Data preprocessing methods	Optimal splitting
MMD-RNN	Hidden layers	2
	Number of input layer nodes	4
	Number of hidden layer nodes	32
	Optimization function	ReLU
	Learning rate	0.01
	Epochs of training	500
	Distance function	MMD
	Data preprocessing methods	None
LSTM	Hidden layers	2
	Number of input layer nodes	4
	Number of hidden layer nodes	32
	Activation function	Sigmoid, tanh
	Optimization function	ReLU
	Learning rate	0.01
	Epochs of training	500

comparison in this section uses different transfer learning based approaches to predict the load power of Case 2. Second, this paper presents that the proposed method is also suitable for the time series forecast with a complete dataset. The comparison group of Case 1 and Case 3 contains both traditional and transfer learning based methods.

4.2. Performance of Transfer Learning in Addressing Fragmented Test Data in the Target Domain.

This section mainly discusses the differences between the proposed TCS-transfer learning model and other transfer learning-based methods in dealing with load forecasting problems. In the first set of experiments in this paper, transformer, MMD-RNN, and MEDA-LSTM are used to compare the superiority of the proposed approach among transfer learning-based learning methods, the forecast load result is shown in Figure 10. Table 3 presents the prediction error of each method with RMSE and MAPE. These prediction errors are evaluated by

$$RMSE = \sqrt{\frac{\sum_{t=0}^{n} (\widehat{x}_t - x_t)^2}{n}}, \tag{18}$$

$$MAPE = \frac{1}{n} \sum_{t=0}^{n} \frac{|\widehat{x}_t - x_t|}{x_t} \times 100\%, \tag{19}$$
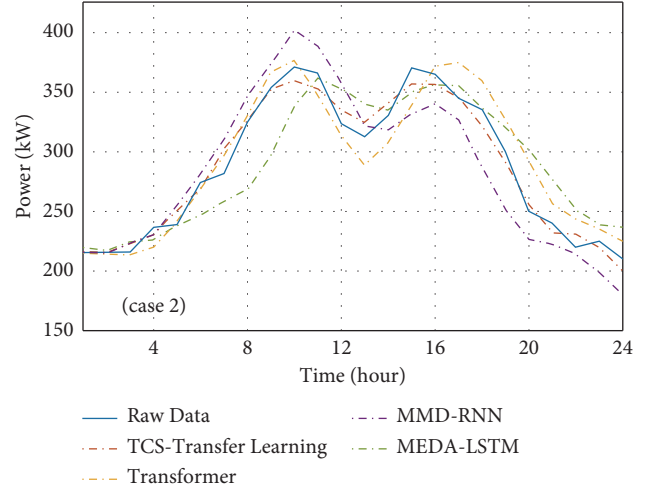
FIGURE 10: Accuracy comparison between TCS-transfer learning and other methods based on the database of Case 2.

TABLE 3: Comparison between MAPE and RMSE values obtained using different forecasting methods.

Methods	MAPE (%)	RMSE
Ref. [41] MMD-RNN	6.09	0.457
Ref. [40] MEDA-LSTM	4.94	0.395
Ref. [42] transformer	4.13	0.196
The proposed TCS-transfer learning	2.87	0.042

where $\widehat{x}_t$ denotes the predicted value of samples, x_t is the actual value of samples, and n represents the number of samples.

The convergence times of different forecasting methods are listed in Table 4; it takes 0.34 s for convergence with regard to the proposed TCS-transfer learning structure, which is a 60.4% increase compared with the MMD-RNN. The prediction accuracy (MAPE) is improved by 52.8% over the RNN-based method, considering GPU cycles with Intel Core i9-12900K. Although the proposed method is slightly slower than the transformer in terms of convergence speed, it has a great improvement in prediction accuracy.

To evaluate the effect of different splitting methods on the training results, two additional methods of dividing the time series in Figure 11, i.e., splits A and B, were designed in this work. Split A denotes that the sequence is randomly divided into k segments; split B is the method where all intervals are with similar distributions, which is the opposite of our proposed split method. The proposed split method has the objective of minimizing the cost function in (5). The "distance" on the Y-axis means the distribution distance MMD with the green line and the RMSE denote the error with the blue line. As a result, it is critical that we divide the periods according to the worst case, where the distributions are the most varied.

4.3. Results from the Case Study on Power Dispatch Strategy.

Figure 12 shows the results of power dispatch solved according to eqs. (10)–(19), whose location are illustrated in

TABLE 4: Convergence time (s) of different forecasting methods.

Methods	Time (s)
Ref. [41] MMD-RNN	0.86
Ref. [40] MEDA-LSTM	0.68
The proposed TCS-transfer learning	0.36
Ref. [42] transformer	0.21
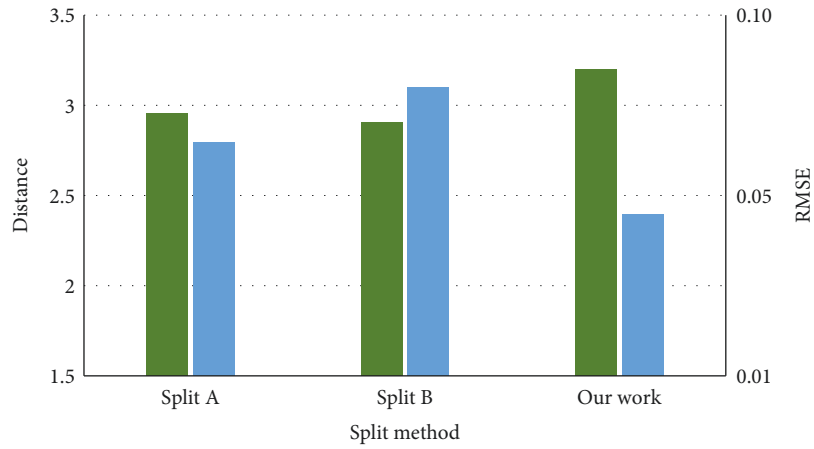
FIGURE 11: Different split methods.
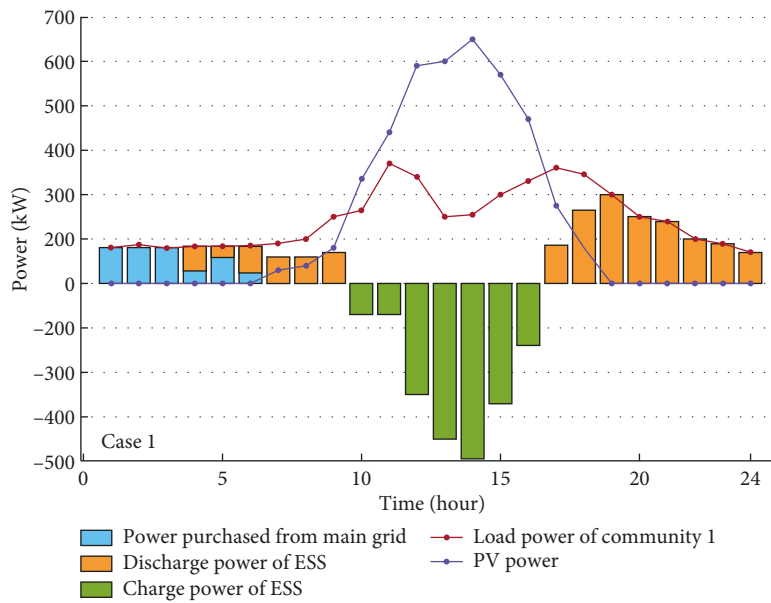
(a)

FIGURE 12: Continued.

(b)

(c)

FIGURE 12: Power load balance curves of communities 1–3.

Figure 8. It can be observed from Case 1 in Figure 12 that the PV output power of the user group in the morning is less than the demand side power; the demanded electricity for this period is purchased from the grid as well as using part of the shared power station considering the optimal economics. When the PV power is higher than the demanded power, the surplus power within the community is stored by the storage power plant to avoid energy curtailment. At the period of 15: 00–20:00, the demand of the community cannot be met by the PV system. While the electricity pricing is high in this time period, the undersupplied energy is provided by the

shared energy storage station. Table 5 provides the electricity rate obtained from the energy provider.

The configuration of the shared energy storage plant results in a capacity of 2,508 kWh and a maximum charge/ discharge power of 637 kW. It can be observed from Figure 13 that at the period of 01:00–06:00, the electricity of users is purchased from the main grid and the power station does not provide electricity to the customer. During the 10: 00–17:00, the energy storage station is in a charging state and the power rises from 382 kWh to the maximum value of 2,000 kWh, where the charging and discharging power value

TABLE 5: Electricity rates.

Level	Time frame	Electricity rates of the main grid [44] $AUD/kWh
Peak	Commercial users 09:00–12:00; 14:00–19:00	0.253
	Residential users 07:00–09:00; 17:00–20:00	0.386
Average	Commercial users 12:00–14:00; 19:00–24:00	0.172
	Residential users 09:00–17:00; 20:00–24:00	0.322
Trough	Commercial users 24:00–09:00	0.073
	Residential users 24:00–07:00	0.086
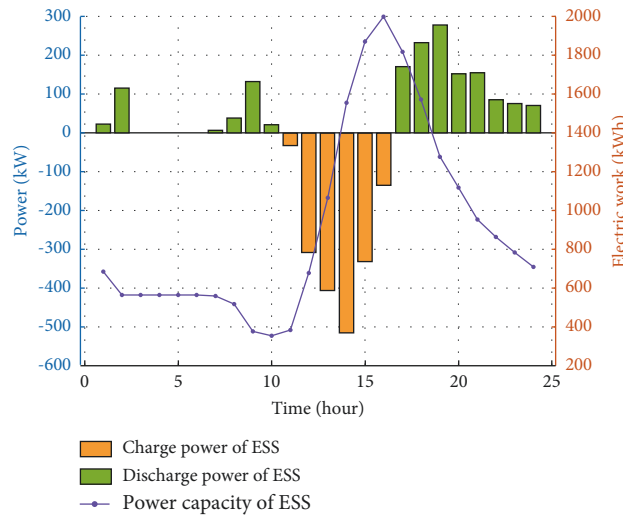
FIGURE 13: Output power and SOC curves of the shared energy storage station.

TABLE 6: User costs with using different energy storage configurations per day.

	Optimization planning results of schemes 1 [45]			Optimization planning results of our work		
No. of user communities	Capacity of ESS (kW·h)	Maximum charge/ discharge power (kW)	Operating costs ($AUD)	Power purchase from the grid (kW·h)	Service fee of shared ESS ($AUD)	Operating costs ($AUD)
Community 1	2820	621	1416.5	583	702.5	759.2
Community 2	1005	163	566.3	16	593.4	595.4
Community 3	376	78	635.8	78	311.4	318.1
Total	4001	862	2618.6	677	1607	1672.7

Optimization planning results of schemes 1: independent configuration of the energy storage system within each user. Optimization planning results of our work: each user has access to a shared energy storage station.

of the energy storage plant is negative, which means that the energy storage plant is charging; if the charge/discharge power value is positive, it denotes that the energy storage plant is discharging.

As can be seen from Figures 12-13, the electric loads of communities 1–3 have reached a balanced state, and there is no phenomenon of energy curtailment, meanwhile, the energy storage station returns to the initial operation state after one cycle of operation, ensuring the normal operation of the next cycle of the energy storage station.

Table 6 compares the daily cost of each case considering different configuration methods of energy storage, the first method presents the independent energy storage system within each user. The total operating cost is AUD 2618.6, which is 36% more expensive than our proposed planning method, and the capacity of the energy storage required by

the customer is reduced by 37.3% due to the complementary nature of the customer's power consumption behavior. Moreover, the energy curtailment is well addressed and the output power of renewables is fully utilized. One limitation with shared energy storage stations is that they are prone to cause harmonics after they are connected to the grid, which can compromise power quality; therefore, more management needs to be put into safe operation when using energy storage stations.

5. Conclusion

This paper proposed a framework for smart grid scheduling that is less reliant on local data while capable of delivering schedules with low operating costs. Specifically, the proposed framework contains the following: (1) a power

forecasting model based on deep transfer learning which can provide high-quality load prediction with limited training data; (2) a novel adaptive time series prediction method based on a neighboring area dataset that aims to train the forecasting model with strong generalization capability; (3) a day-ahead optimal economic power scheduling model considering the shared energy storage station. Results based on a case study with field load data in Western Australia showed that the maximum improvement of the proposed forecasting method is up to 52.8% in MAPE compared to other transfer learning-based methods, and up to 64.4% compared to the traditional method. The total operating cost after optimization according to the proposed method was reduced by 36.1%. These numbers indicate the proposed framework is a promising approach to solving power planning problems with incomplete datasets, in particular in addressing the cyber threats.

In the future, we plan to explore a deeper extension of TCS-transfer learning to a transformer for better performance. Moreover, this work only designed a centralized energy storage system. If multiple energy storage is needed, optimal coordination between these dispatch-oriented energy storage systems would be considered a promising area for future investigation.

Data Availability

The features data input into the predicting model used to support the findings of this study have been deposited in the following repository: (1) Exemplary Energy Partners Company. (http://www.exemplary.com.au/), (2) Office Holidays (https://www.officeholidays.com/countries/australia/2021), and (3) Australia Net Migration Rate 1950-2022 (https://www.macrotrends.net/countries/AUS/australia/net-migration). The net power data used to support the findings of this study are currently under embargo, while the research findings are commercialized. Requests for data, 6/12 months after publication of this article, will be considered by the corresponding author

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] Irena, "Planning for the renewable future," 2011, https://www.irena.org/energytransition.

[2] M. Dabbaghjamanesh, A. Kavousi-Fard, and Z. Y. Dong, "A novel distributed cloud-fog based framework for energy management of networked microgrids," *IEEE Transactions on Power Systems*, vol. 35, no. 4, pp. 2847–2862, 2020.

[3] Z. Min, B. Wang, S. Guo, and J. Watada, "Multi-objective prediction intervals for wind power forecast based on deep neural networks," *Information Sciences*, vol. 550, pp. 207–220, 2021.

[4] N. Victor and D. Lopez, "Sl-LSTM: a Bi-directional LSTM with stochastic gradient descent optimization for sequence labeling tasks in big data," *International Journal of Grid and High Performance Computing*, vol. 12, no. 3, pp. 1–16, 2020.

[5] M. S. Kivi, B. Blankely, M. Masters, and C. J. Bernacchi, "Development of a data-assimilation system to forecast agricultural systems: a case study of constraining soil water and soil nitrogen dynamics in the APSIM model," *Science of the Total Environment*, vol. 820, 2022.

[6] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[7] M. A. Alqahtani, "Cybersecurity awareness based on software and E-mail security with statistical analysis," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 6775980, 12 pages, 2022.

[8] M. Cai, M. Pipattanasomporn, and S. Rahman, "Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques," *Applied Energy*, vol. 236, pp. 1078–1088, 2019.

[9] M. S. Cui, "District heating load prediction algorithm based on bidirectional long short-term memory network model," *Energy*, vol. 254, no. 1, Article ID 124283, 2022.

[10] X. Luo and X. Zhu, "Deep learning based forecasting of photovoltaic power generation by incorporating domain knowledge," *Energy*, vol. 225, Article ID 120240, 2021.

[11] H. H. Dong, Y. Gao, Y. Fang, and M. Liu, "The short-term load forecasting for special days based on bagged regression trees in qingdao, China," *Computational Intelligence and Neuroscience*, vol. 2021, Article ID 3693294, 16 pages, 2021.

[12] A. Sultana, A. Bardalai, and K. K. Sarma, "Salp swarm-artificial neural network based cyber-attack detection in smart grid," *Neural Processing Letters*, 2022.

[13] Wikipedia, "Ukraine power grid hack," 2015, https://en.wikipedia.org/wiki/Ukraine_power_grid_hack.

[14] H. S. Lallie, K. Debattista, and J. Bal, "A review of attack graph and attack tree visual syntax in cyber security," *Computer Science Review*, vol. 35, Article ID 100219, 2020.

[15] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2009.

[16] W. Cui, G. Zheng, Z. Shen, and S. Jiang, "Transfer learning for sequences via learning to collocate," in *Proceedings of the International Conference On Learning Representations*, New Orleans, LA, USA, May 2019.

[17] M. N. Bisheh, X. Wang, S. I. Chang, and S. Lei, "Image-based characterization of laser scribing quality using transfer learning," *Journal of Intelligent Manufacturing*, 2022.

[18] Y. Liu, S. Liu, J. Xu, X. Kong, L. Xie, and K. Chen, "Forest pest identification based on a new dataset and convolutional neural network model with enhancement strategy," *Computers and Electronics in Agriculture*, vol. 192, Article ID 106625, 2021.

[19] J. Lines and A. J. Bagnall, "Time series classification with ensembles of elastic distance measures," *Data Mining and Knowledge Discovery*, vol. 29, no. 3, pp. 565–592, 2014.

[20] Y. K. Lu, Z. Tian, R. Zhou, and W. Liu, "A general transfer learning-based framework for thermal load prediction in regional energy system," *Energy*, vol. 217, no. 15, Article ID 119322, 2021.

[21] J. K. Li, M. Lin, Y. Li, and X. Wang, "Transfer learning network for nuclear power plant fault diagnosis with

unlabeled data under varying operating conditions," *Energy*, vol. 254, Article ID 124358, 2022.

[22] T. Liang, Q. Zhao, Q. Lv, and H. Sun, "A novel wind speed prediction strategy based on Bi-LSTM, MOOFADA and transfer learning for centralized control centers," *Energy*, vol. 230, no. 1, Article ID 120904, 2021.

[23] H. Yin, Z. Ou, J. Fu, Y. Cai, S. Chen, and A. Meng, "A novel transfer learning approach for wind power prediction based on a serio-parallel deep learning architecture," *Energy*, vol. 234, no. 1, Article ID 121271, 2021.

[24] C. Li, X. Liu, Y. Cao et al., "A time-scale Adaptive dispatch method for renewable energy power supply systems on islands," *IEEE Transactions on Smart Grid*, vol. 7, no. 2, pp. 1069–1078, 2015.

[25] C. Jiang and Z. Xia, "Application of a hybrid model of big data and BP network on fault diagnosis strategy for microgrid," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 1554422, 12 pages, 2022.

[26] Y. Li, M. Vilathgamuwa, S. S. Choi et al., "Design of minimum cost degradation-conscious lithium-ionbattery energy storage system to achieve renewable power dispatchability," *Applied Energy*, vol. 260, Article ID 114282, 2020.

[27] B. S. Pali and S. Vadhera, "An innovative continuous power generation system comprising of wind energy along with pumped-hydro storage and open well," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 1, pp. 145–153, 2020.

[28] H. Ismail Fawaz, "Transfer Learning for Time Series Classification," in *Proceedings of the International Conference on Big Data*, pp. 1367–1376, Seattle, WA, USA, December 2018.

[29] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Proceedings of the Advances in Neural Information Processing Systems*, vol. 60, pp. 1097–1105, 2012.

[30] M. Long, J. Wang, and M. I. Jordan, "Deep transfer learning with joint adaptation networks," 2018, https://arxiv.org/abs/1605.06636.

[31] X. Huang and S. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 1510–1519, Venice, Italy, October 2018.

[32] Y. Du, J. Wang, W. Feng, S. Pan, and T. Qin, "AdaRNN: adaptive learning and forecasting for time series," in *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, Virtual Event, Australia, November 2021.

[33] E. Tzeng, J. Hoffman, N. Zhang, and K. Saenko, "Deep domain confusion: maximizing for domain invariance," 2018, https://arxiv.org/abs/1412.3474.

[34] F. Y. Wang, J. Zhang, Q. Wei, X. Zheng, and L. Li, "PDP: parallel dynamic programming," *IEEE-caa Journal of Automatica Sinica*, vol. 4, pp. 1–5, 2017.

[35] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures," *Neural Networks*, vol. 18, no. 5-6, pp. 602–610, 2005.

[36] W. P. Liu, T. Peng, R. Tang, and Y. Umeda, "An Internet of Things-enabled model-based approach to improving the energy efficiency of aluminum die casting processes," vol. 202, Article ID 117716, 2020.

[37] X. M. Gao, S. F. Yang, and S. B. Pan, "Optimal parameter selection for support vector machine based on artificial bee colony algorithm: a case study of grid-connected PV system power prediction," *Computational Intelligence and Neuroscience*, vol. 2017, Article ID 7273017, 14 pages, 2017.

[38] S. Q. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," in *Proceedings of the 29th Annual Conference on Neural Information Processing Systems (NIPS)*, December 2015.

[39] G. B. Huang, Q. Y. Zhu, and C. K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, pp. 489–501, 2006.

[40] F. Z. Zhuang, Z. Qi, K. Duan et al., "A comprehensive survey on transfer learning," *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, 2021.

[41] Q. Zhou, W. Zhou, and S. Wang, "Semantic adaptation network for unsupervised domain adaptation," *Neurocomputing*, vol. 454, pp. 313–323, 2021.

[42] Z. Jin, J. Kim, H. Yeo, and S. Choi, "Transformer-based map-matching model with limited labeled data using transfer-learning approach," *Transportation Research Part C: Emerging Technologies*, vol. 140, Article ID 103668, 2022.

[43] A. Vecchietti, S. Lee, and I. E. Grossmann, "Modeling of discrete/continuous optimization problems: characterization and formulation of disjunctions and their relaxations," *Computers & Chemical Engineering*, vol. 27, no. 3, pp. 433–448, 2003.

[44] globalpetrolprices, "Retail energy price data," 2022, https://www.globalpetrolprices.com/.

[45] N. Lee, C. H. Nee, S. S. Yap et al., "Capacity sizing of embedded control battery–supercapacitor hybrid energy storage system," *Energies*, vol. 15, no. 10, 2022.

Research Article

Sustainable Energy Consumption Model for Textile Industry Using Fully Intuitionistic Fuzzy Optimization Approach

Sajida Kousar [ID],[1] **Urooj Shafqat** [ID],[1] **Nasreen Kausar** [ID],[2] **Dragan Pamucar** [ID],[3] **Yeliz Karaca** [ID],[4] **and Mohammed Abdullah Salman** [ID][5]

[1]*Department of Mathematics and Statistics, International Islamic University Islamabad, Islamabad, Pakistan*
[2]*Department of Mathematics, Faculty of Arts and Sciences Yildiz Technical University, Esenler 34210, Istanbul, Turkey*
[3]*Faculty of Organizational Sciences, University of in Belgrade, Belgrade, Serbia*
[4]*University of Massachusetts Medical School (UMASS), Worcester MA 01655, MA, USA*
[5]*College of Education, Applied Sciences and Arts Amran University, Amran, Yemen*

Correspondence should be addressed to Mohammed Abdullah Salman; masalman79@gmail.com

Consumption of renewable energy is on the rise because new technologies have made it cheaper and easier to meet the needs of a long-term energy source. In the present study, the idea of optimal usage of sustainable energy is discussed, taking into consideration the environmental and economic conditions that exist in Pakistan's textile manufacturing industry. By taking into account the regional potential for the application of renewable energy resources, solar energy generators are taken into consideration, and a fully intuitionistic fuzzy (FIF) textile energy model is constructed. Using the FIF model to determine the optimal distribution of solar energy units resulted in a tolerable number of unused energy units. These units may be returned to the central power supply station, which would save both money and energy.

1. Introduction

In the textile industry, electricity and thermal energy usage is the most common, as generated at industrial level through nonrenewable energy resources (petroleum, hydrocarbon gas liquids, natural gas, coal, etc.) or obtained directly through the government according to industrial policies. For example, in Pakistan, coal, nuclear, natural gas, hydroelectric, wind, and solar generators are the major electricity developing sources. Specifically, in Pakistan, the two crucial electricity producers are WAPDA (water and power development authority) that generates hydroelectricity and PEPCO (Pakistan electric power company) where distribution companies (DISCOs) under PEPCO work to maintain the path between electricity producers to consumer end by purchasing and selling it to representative area distribution companies. Except for one (K-electric), all these companies are owned by the government, and hence, the government

provides different policies to different consuming sectors. In Pakistan, the energy crisis has been a never-ending situation from last two decades due to imbalance production and consumption ratio. At this point of energy crisis, the government of Pakistan and APTMA (All Pakistan Textile Mills Association) are on continuous conflict regarding the availability and favourable rates of energy. Working of textile industry is based on multiple stages like spinning, weaving, rewinding, dying, and so on, which are further subdivided into several other stages (see Figure 1).

All those stages require a huge amount of energy for their processing. Proper energy management planning can help industries to survive in natural crises or any other unpredictable conditions. Only if a single factor like energy optimization is focused, firstly, it will help industries to minimize their investing cost; secondly, the buyer will be able to get the product at a low cost. Thirdly, the saved energy will be used by other sectors like domestic sectors to
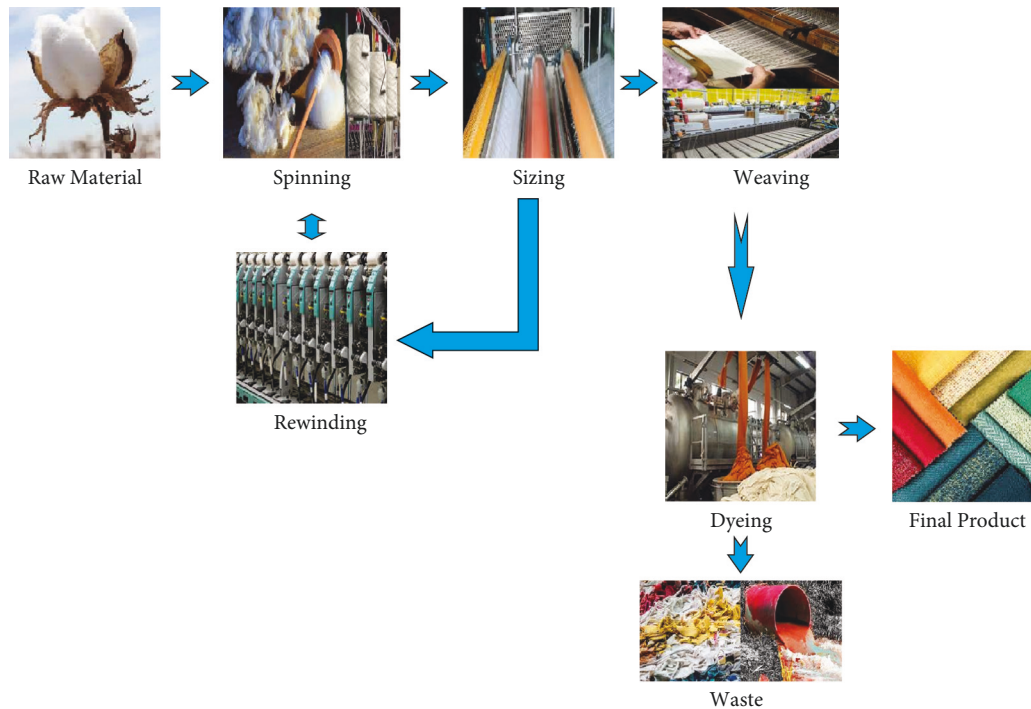
Figure 1: Textile industry processes.

fulfill their energy need and most importantly for the sake of healthy environmental and economic conditions, energy saving is a significant step. In the present time, when COVID-19 or other natural and economic disasters are happening, it is almost impossible to make an accurate and successful management plan. To ensure better energy management for textile industries under all these circumstances, developing an energy optimization model that can work in uncertain situations and generate the best optimal output is needed. To reduce the energy cost, one way is to reconsider the source of energy and replace it with the best convenient sustainable resource. Another way is to allocate energy units optimally. For the first way, considering the situation of Pakistan, it is best for textile industries to generate their own electricity and thermal energy through the Solar System. For the second way, it is best to modify the linear programming (LP) model for the textile industry according to the fuzzy environment. For this purpose, the conversion of the objective function, constraint equations, demand, and supply into fuzzy number is needed. Precisely that fuzzy number can overcome the uncertainty regarding material availability, working hours, financial and social barriers, labour, money, and space.

2. Literature Review

Although in Pakistan, textile industries cover 46% of the whole manufacturing sector and are a macro contributor with 8.5% of Pakistan's gross domestic product (GDP), it is also a huge environmental pollution contributor. In this setting, textile industries have a significant share in water and air pollution. The amount of clean water utilized by the textile industrial sector is substantially more as compared to

the agricultural one [1], and for the production of one kg fabric output at the wet processing stage, the amount of water needed is almost 80–150 liters in addition to other chemicals [2]. Not only that, the report of 2015 stated that textile sector solely accounted for 1.2 billion tons of carbon dioxide [3], while according to the present behaviour of this industrial sector, 26% of carbon emission and 0.3 billion tons of crude oil consumption by the year 2025 is estimated by the study as well [3]. This is because the textile industry's fuel consumption is not only related to the production level but also during the transfer of textile products through different transportation means to different areas [4] taking part in carbon emission. All these aspects clear the reason for being called textile industry as the most polluting industry. In Faisalabad, Pakistan, a chain of textile industries is located around the road connecting twin districts Khuranwala and Jaranwala as shown in Figure 2.

Due to industrial water waste and carbon emission, the Khuranwala/Jaranwala road of 25–28 km long is full of unbearable smell and smog. The residents and daily travellers nearby face exposure to this unhealthy environment and suffer from breathing and eye diseases. By controlling abrupt energy consumption and ways connected to pollution either directly (industrial waste) or indirectly, like nonrecyclable fabric formation, excess of textile products, and so on, can help to lessen down pollution. Several studies have been conducted to reuse industrial water waste. Nadeem et al. [5] discussed the recycling and treatment potential of textile waste water by using membrane technology. According to Dehghani and Yoo [6], the organic components and approximate temperature of the textile industry's waste water are sufficient enough for the production of biofuel that can be used as a cheap source for creating thermal energy.

Figure 2: Industrial chain of Khuranwala.

Efficient consumption of energy except proper planning is not possible in the region due to the complicated nature of the textile manufacturing sector. Spinning and weaving stages fully depend on electricity, while sizing, dying, and rewinding requires thermal energy to heat water or to dry the fabric [7, 8]. Studies have shown that, globally one trillion KWh of electricity is consumed per year to produce 60 billion kg fabric [9]. Usage of renewable energy resources like solar energy, geothermal energy, and landfill gas is helpful in the reduction of textile production costs. It lowers the total energy per unit usage that automatically degrade the environmental pollution factor as well. In Atlanta, seven textile manufacturing plants only use renewable electricity, and 89% of this electricity is renewable source [10].

Specifically, in Pakistan, the existence of energy self-sufficient textile industries is rare. The All Pakistan Textile Mills Association (APTMA) Punjab Chairman also insisted on adopting renewable solar energy resources to fulfill the need for electricity with affordable cost for the maintenance of industrial sustainability. According to their reporting, Pakistan textile industry is ready to shift to solar or hybrid energy generating systems [11]. This shift will not only reduce the cost but also reduce the carbon emission. In developing countries like Pakistan, where energy production is not proportional to the need, the cheap production and optimal allocation of energy is required. The usage of renewable energy resources can also be of great help for the industrial sector. Considering solar energy as a means for electricity is the best suitable option. The temperature and weather conditions of Pakistan as shown in Figure 3 can accelerate the outcome of the solar energy system.

The advancement of solar energy system makes electricity production much more efficient. Using this kind of system is considered as sustainable and ecological investment. Textile industry extensively uses electricity, so the optimal utilization of energy at each stage is regarded as an important initial target.

The conventional and most widely used method with reliable results for optimization is linear programming (LP) created by Kantorovich [13]. The only drawback of LP is its nonflexibility regarding nature. Natural scenarios are full of ambiguity, while classical optimization based on LP does not cover these uncertainties. After the revelation of fuzzy sets by Zadeh [14], LP also started to be modified. Zimmerman [15] used the concept of fuzziness in LP. He invented the technique to find a solution for fuzzy multiobjective LP. The fuzziness in this technique is due to the presence of fuzzy optimization conditions in it. Atanassov [16] improved the concept of fuzziness more by introducing a new generalization of a fuzzy set named intuitionistic fuzzy set, in which both the membership grade and nonmembership grade of element from decision set are necessary. This generalization created more optimization techniques [17]. Angelov [18] made the intuitionistic fuzzy LP technique. Subsequently, several researches has appeared on intuitionistic fuzzy linear programming (IFLP). Hussain and Kumar [19] worked on intuitionistic fuzzy transportation problem (IFTP). Ebrahimnejad and Verdegay [20] worked on a fully intuitionistic transportation problem (TP).

Several optimization models for sustainable production and consumption of renewable energy are constructed using
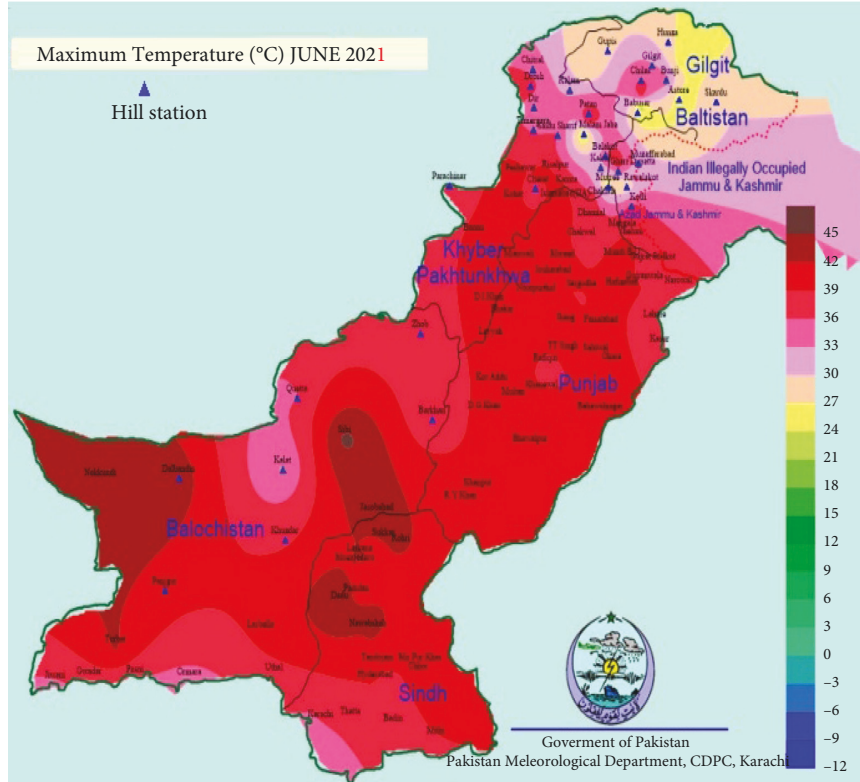
Figure 3: Temperature map of Pakistan (Pakistan Meteorological Department [12]).

bilevel programming [21], mixed integer linear programming [22], simulation-optimization modelling [23], linear programming [24–26], and goal programming [27] approach. Nematian and Farzi [28] considered the case study of energy recovered from urban solid waste in Iran and developed energy and environmental management model using fuzzy LP approach. Zhou et al. [29] used type-2 fuzzy chance-constrained fractional integrated modelling method for the management of energy system subjected to uncertainties and risks. Kouaissah and Hocine [30] and Hashemizadeh and Ju [31] introduced sustainable and renewable energy portfolios using a fuzzy interval goal programming technique. Khan et al. [32] examined and assessed the optimal cost system of electricity generation for the socio-economic sustainability of India by developing a sustainable and flexible electricity generation model using flexible fuzzy goal programming. Sustainable production in the textile industry depends upon the search of sustainable ways of energy management. Emeç and Akkaya [33] developed a fuzzy optimal renewable energy model (F-OREM) to solve the energy problem involving fuzzy parameters. Abbas et al. [34] systematically analyzed the potential of cotton crop waste to synergize industrial energy systems by integrating strategic and tactical decision models into an integrated model. The results indicate that cotton crop waste is a conducive and convenient source of sustainable energy supply for the textile industry. Techato et al. [35] presented a systematic review of the optimization models used in the textile industry, mostly established for cost minimization in logistics and production and optimized mainly with linear programming, integer programming, Markov chains, genetic algorithms, and multi-objective programming. The effectiveness of the present study in contrast with conventional LP is further discussed in upcoming sections where a fully intuitionistic fuzzy energy optimization model for Pakistan's textile industry is formed that results in optimal energy allocation that is further presented. For sustainability, the idea of replacing typical energy resources with solar energy generators along with its investing and working cost is also to be considered in reference to the existing self-sufficient textile industry of Pakistan.

3. Preliminaries

3.1. Intuitionistic Fuzzy Set. The intuitionistic fuzzy set was presented by Atanassov [16]; the degree of nonmembership and the degree of membership were expressed by the two characteristic functions. An intuitionistic fuzzy set (IFS) $\mathcal{N}$ in $\mathcal{Q}$ can be described as an element of the following form $\mathcal{N} = \{\langle s, \zeta_{\mathcal{N}}(s), \eta_{\mathcal{A}}(s)\rangle | s \in \mathcal{Q}\}$ where the functions $\eta_{\mathcal{N}}$: $\mathcal{Q} \longrightarrow [0, 1]$ and $\zeta_{\mathcal{N}}$: $\mathcal{Q} \longrightarrow [0, 1]$ represents the degree of nonmembership and the degree of membership of the component $s \in \mathcal{Q}$, respectively.

3.2. Triangular Intuitionistic Fuzzy Number. It is based on membership and nonmembership functions [36], considering $\widetilde{Q_l^{\gamma}} = (q_1, q_2, q_3; q_1, q_2, q_3)$ such that $q_1 \leq q_1 \leq q_2 \leq q_3 \leq q_3$ and its membership function is defined as

$$\mu_{\widetilde{Q_l^Y}} = \begin{cases} \dfrac{x - q_2}{q_2 - q_1} & q_1 \leq x < q_2 \\ \\ 1, & x = q_2 \\ \\ \dfrac{q_3 - x}{q_3 - q_2} & q_2 < x \leq q_3 \\ \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

The nonmembership function is as follows:

$$\nu_{\widetilde{Q_l^Y}} = \begin{cases} \dfrac{q_2 - x}{q_2 - \acute{q_1}} & \acute{q_1} \leq x < q_2 \\ \\ 0, & x = q_2 \\ \\ \dfrac{x - q_2}{\acute{q_3} - q_2} & q_2 < x \leq \acute{q_3} \\ \\ 1, & \text{otherwise.} \end{cases} \tag{2}$$

With restriction $0 \leq \mu_{\widetilde{Q_l^Y}} x + \nu_{\widetilde{Q_l^Y}} x \leq 1$. If q_1, q_2, q_3, $\acute{q_1}$, and $\acute{q_3}$ are nonnegative, then $\widetilde{Q_l^Y}$ will be a nonnegative intuitionistic fuzzy number. The geometrical interpretation of Triangular intuitionistic fuzzy number is presented in Figure 4.

3.3. Arithmetic Operations. Let $\widetilde{Q_l^Y} = (q_1, q_2, q_3; \acute{q_1}, q_2, \acute{q_3})$ and $\widetilde{P_l^Y} = (p_1, p_2, p_3; \acute{p_1}, p_2, \acute{p_3})$ are two triangular intuitionistic fuzzy numbers, then the algebraic operations between them are defined as follows:

(1) $\widetilde{Q_l^Y} \oplus \widetilde{P_l^Y} = (q_1 + p_1, q_2 + p_2, q_3 + p_3; \acute{q_1} + \acute{p_1}, q_2 + p_2, q_3 + p_3)$

(2) $\widetilde{Q_l^Y} \ominus \widetilde{P_l^Y} = (q_1 - p_3, q_2 - p_2, q_3 - p_1; \acute{q_1} - \acute{p_3}, q_2 - p_2, q_3 - \acute{p_1})$

(3) $\kappa \widetilde{Q_l^Y} = (\kappa q_1, \kappa q_2, \kappa q_3; \kappa \acute{q_1}, \kappa q_2, \kappa \acute{q_3}); \text{ for } \kappa \geq 0$

(4) $\kappa \widetilde{Q_l^Y} = (\kappa q_3, \kappa q_2, \alpha q_1; \kappa \acute{q_3}, \kappa q_2, \kappa \acute{q_1}); \text{ for } \kappa < 0$

(5) $\widetilde{Q_l^Y} \circledast \widetilde{P_l^Y} = (r_1, r_2, r_3; \acute{r_1}, r_2, \acute{r_3})$

Here,

$$\begin{aligned} r_1 &= \min\{q_1 p_1, q_1 p_3, q_3 p_1, q_3 p_3\} \\ r_3 &= \max\{q_1 p_1, q_1 p_3, q_3 p_1, q_3 p_3\} \\ r_2 &= q_2 p_2 = \acute{r_2} \\ \acute{r_1} &= \min\{\acute{q_1}\acute{p_1}, \acute{q_1}\acute{p_3}, \acute{q_3}\acute{p_1}, \acute{q_3}\acute{p_3}\} \\ \acute{r_3} &= \max\{\acute{q_1}\acute{p_1}, \acute{q_1}\acute{p_3}, \acute{q_3}\acute{p_1}, \acute{q_3}\acute{p_3}\}. \end{aligned} \tag{3}$$

In particular, if $\widetilde{Q_l^Y}$ and $\widetilde{P_l^Y}$ are two nonnegative triangular intuitionistic fuzzy numbers, then their product $\widetilde{Q_l^Y} \circledast \widetilde{P_l^Y}$ will be $\{q_1 p_1, q_2 p_2, q_3 p_3; \acute{q_1}\acute{p_1}, q_2 p_2, \acute{q_3}\acute{p_3}\}$.
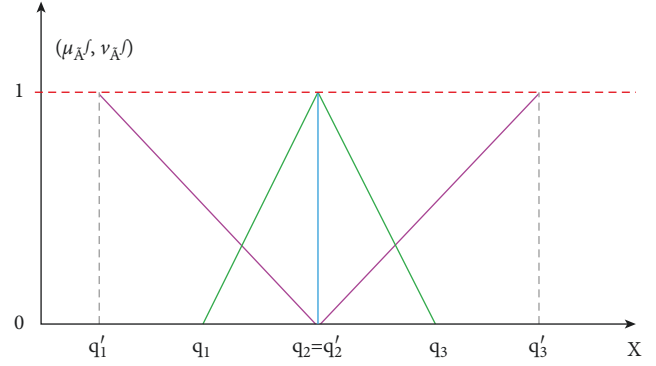
FIGURE 4: Graphical representation of a triangular intuitionistic fuzzy number.

For the conversion of triangular intuitionistic fuzzy number into crisp following accuracy function [20] is used.

$$I\left(\widetilde{Q_l^Y}\right) = \frac{1}{8}\left(q_1 + 4q_2 + q_3 + \acute{q_1} + \acute{q_3}\right). \tag{4}$$

4. Fully Intuitionistic Fuzzy Linear Programming Model

Mathematically, a general triangular fuzzy intuitionistic energy optimization model is expressed as follows:

$$\min \widetilde{\alpha^Y} = \sum_{j=1}^{m}\left\{\widetilde{C_j^Y} \circledast \widetilde{x_j^Y}\right\}$$

$$\min \widetilde{\alpha^Y} = \left(\sum_{j=1}^{m} c_{j,1} x_{j,1}, \sum_{j=1}^{m} c_{j,2} x_{j,2}, \sum_{j=1}^{m} c_{j,3} x_{j,3}; \sum_{j=1}^{m} \acute{c}_{j,1}\acute{x}_{j,1}, \right. \tag{5}$$

$$\left. \sum_{j=1}^{m} c_{j,2} x_{j,2}, \sum_{j=1}^{m} \acute{c}_{j,3}\acute{x}_{j,3}\right).$$

Subjected to triangular intuitionistic fuzzy energy optimization constraints,

$$\begin{array}{cccc} \widetilde{s_{21}^Y}\widetilde{x_1^Y} + \widetilde{s_{2j}^Y}\widetilde{x_j^Y} + \cdots + \widetilde{s_{2m}^Y}\widetilde{x_m^Y} & (\leq = \geq) & \widetilde{D_2^Y} \\ \vdots & (\leq = \geq) & \vdots \\ \widetilde{s_{n1}^Y}\widetilde{x_1^Y} + \widetilde{s_{nj}^Y}\widetilde{x_j^Y} + \cdots + \widetilde{s_{nm}^Y}\widetilde{x_m^Y} & (\leq = \geq) & \widetilde{D_n^Y} \\ \widetilde{s_{11}^Y}\widetilde{x_1^Y} + \widetilde{s_{1j}^Y}\widetilde{x_j^Y} + \cdots + \widetilde{s_{1m}^Y}\widetilde{x_m^Y} & (\leq = \geq) & \widetilde{D_1^Y}. \end{array} \tag{6}$$

Here, $\widetilde{x_j^Y} \geq \widetilde{0^Y}$. The values of $\widetilde{x_j^Y}$ will optimize the intuitionistic fuzzy energy cost objective function where all the $\widetilde{x_j^Y}$ will satisfy the nonnegativity condition and constraint equations. In the above objective function $\widetilde{C_j^Y} = (c_{j,1}, c_{j,2}, c_{j,3}; \acute{c}_{j,1}, c_{j,2}, \acute{c}_{j,3}) j = 1, 2, 3, \ldots$ represent the triangular intuitionistic fuzzy cost coefficients and s_{ij} are the technological coefficients representing the amount of i th resource consuming at rate of $\widetilde{x_j^Y}$ per unit where

$$\widetilde{D_i^Y} = \left(d_{j,1}, d_{j,2}, d_{j,3}; \acute{d}_{j,1}, d_{j,2}, \acute{d}_{j,3}\right) \tag{7}$$

refers to the total availability of the ith triangular intuitionistic fuzzy resource. For the conversion of intuitionistic fuzzy energy optimization model into linear programming problem, the accuracy function is used as defined in (4). The following equation presents the defuzzified objective function.

$$\min \sum_{j=1}^{m} \left(c_{j,1}x_{j,1}, c_{j,2}x_{j,2}, c_{j,3}x_{j,3}; c'_{j,1}x'_{j,1}, c'_{j,2}x_{j,2}, c'_{j,3}x'_{j,3} \right)$$

$$= \min \left\{ \frac{1}{8} \sum_{j=1}^{m} \left(c_{j,1}x_{j,1} + 4c_{j,2}x_{j,2} + c_{j,3}x_{j,3} + c'_{j,1}x'_{j,1} + c'_{j,3}x'_{j,3} \right) \right. $$

$$\tag{8}$$

The above defined set of triangular intuitionistic constraints (6) can be rewritten as follows:

$$\sum_{j=1}^{m} s_{ij}x_{j,1} \, (\leq \, = \, \geq) d_{i,1}.$$

$$\sum_{j=1}^{m} s_{ij}x_{j,2} \, (\leq \, = \, \geq) d_{i,2}$$

$$\sum_{j=1}^{m} s_{ij}x_{j,3} \, (\leq \, = \, \geq) d_{i,3} \tag{9}$$

$$\sum_{j=1}^{m} s_{ij}x'_{j,1} \, (\leq \, = \, \geq) d'_{i,1}$$

$$\sum_{j=1}^{m} s_{ij}x'_{j,3} \, (\leq \, = \, \geq) d'_{i,3}$$

The optimal solution of $x_{j,1}$, $x_{j,2}$, $x_{j,3}$, $x'_{j,1}$, and $x'_{j,3}$ will be obtained by solving objective function (7) under constraints of (4).

5. Energy Optimization Model for Textile Industry

The textile industry is usually based on five stages which are further subdivided into several others. The main five stages are spinning, sizing, rewinding, weaving, and dyeing. Each stage's product $(x'_i s)$ forward lessens quantity to the next stage, that is, input < output always. For example, 97% of stage 2 product is further processed for next weaving stage, and remaining 3% goes to the rewinding stage whose product again becomes the input of stage 2. Considering minimum waste up to 7% for weaving stage, the material processed further is 93% of total x_3, at the last stage, the wastage is approximately 4%. Considering a standard five stages textile model, the per month demand of stage x_1, x_4, and x_5 products 400 units, 600 units, and 20,000 units. Per month total working hours are 720 hours [37]. The electricity cost per unit is 20.62 PKR/kWh, fuel oil is 85.68 PKR/litre, and LPG cost at the rate of 19.4103 PKR/litre [38, 39]. Energy cost for each stage is presented in Table 1.

The formulation of above model in intuitionistic fuzzy environment provided triangular intuitionistic fuzzy cost coefficients that represent the cost spent during production period as follows:

Table 1: Energy cost in PKR.

Energy	Spinning	Sizing	Weaving	Rewinding	Dyeing
Electricity	51.55	5.155	41.24	15.465	20.62
LPG	0	19.0103	0	0	135.8721
Furnace fuel	0	0	0	0	21.40
Total energy cost	51.55	24.5653	41.24	15.465	177.892

$$\widetilde{C_j^{\curlyvee}} = \left\{ \widetilde{C_1^{\curlyvee}} = (46.55, 51.55, 56.55; 41.55, 51.55, 61.55), \right.$$

$$\widetilde{C_2^{\curlyvee}} = (19.5653, 24.5653, 29.5653; 14.5653, 24.5653, 34.5653),$$

$$\widetilde{C_3^{\curlyvee}} = (36.24, 41.24, 46.24; 31.24, 41.24, 51.24),$$

$$\widetilde{C_4^{\curlyvee}} = (10.465, 15.465, 20.465; 5.465, 15.465, 25.465),$$

$$\left. \widetilde{C_5^{\curlyvee}} = (172.892, 177.892, 182.892; 167.892, 177.892, 187.892) \right\}$$

$$\tag{10}$$

The monthly triangular intuitionistic fuzzy production demand in kg for three products and total availability of working hours in a month are stated as follows:

$$\widetilde{D_i^{\curlyvee}} = \left\{ \widetilde{D_1^{\curlyvee}} = (350, 400, 450; 300, 400, 500) \right.$$

$$\widetilde{D_2^{\curlyvee}} = (550, 600, 650; 500, 600, 700),$$

$$\widetilde{D_3^{\curlyvee}} = (19950, 20000, 20050; 19900, 20000, 20100),$$

$$\left. \widetilde{D_4^{\curlyvee}} = (670, 720, 770; 620, 720, 820) \right\}$$

$$\tag{11}$$

Mathematically, intuitionistic fuzzy energy optimization is framed as follows:

$$\min \widetilde{\alpha^{\curlyvee}} = \left(\sum_{j=1}^{m} \widetilde{C_j^{\curlyvee}} \circledast \widetilde{x_j^{\curlyvee}} \right) = \left((46.55, 51.55, 56.55; 41.55, 51.55, 61.55) \circledast \widetilde{x_1^{\curlyvee}} \right)$$

$$\oplus \left((19.5653, 24.5653, 29.5653; 14.5653, 24.5653, 34.5653) \circledast \widetilde{x_2^{\curlyvee}} \right)$$

$$\oplus \left((36.24, 41.24, 46.24; 31.24, 41.24, 51.24) \circledast \widetilde{x_3^{\curlyvee}} \right)$$

$$\oplus \left((10.465, 15.465, 20.465; 5.465, 15.465, 25.465) \circledast \widetilde{x_4^{\curlyvee}} \right)$$

$$\oplus \left((172.892, 177.892, 182.892; 167.892, 177.892, 187.892) \circledast \widetilde{x_5^{\curlyvee}} \right).$$

$$\tag{12}$$

Subject to

$$\widetilde{x_1^{\curlyvee}} \ominus \widetilde{x_2^{\curlyvee}} \geq (350, 400, 450; 300, 400, 500),$$

$$0.03\widetilde{x_2^{\curlyvee}} \ominus \widetilde{x_3^{\curlyvee}} \oplus 0.07\widetilde{x_4^{\curlyvee}} = (0,0,0;0,0,0),$$

$$0.97\widetilde{x_2^{\curlyvee}} \ominus \widetilde{x_4^{\curlyvee}} = (0,0,0;0,0,0),$$

$$0.93\widetilde{x_4^{\curlyvee}} \ominus \widetilde{x_5^{\curlyvee}}$$

$$0.96\widetilde{x_5^{\curlyvee}} \geq (19950, 20000, 20050; 19900, 20000, 20100),$$

$$0.007\widetilde{x_1^{\curlyvee}} \ominus 0.007\widetilde{x_2^{\curlyvee}} \oplus 0.013\widetilde{x_4^{\curlyvee}} \oplus 0.0062\widetilde{x_5^{\curlyvee}} \leq (670, 720, 770; 620, 720, 820),$$

$$\widetilde{x_i^{\curlyvee}} \geq (0,0,0;0,0,0), i = 1,2,3,4,5,$$

$$\tag{13}$$

According to the above intuitionistic model, this formulated intuitionistic fuzzy energy optimization model is defuzzified as follows:

$$
\begin{aligned}
\min = \frac{1}{8}&\left[46.55x_{1,1} + 4 \times 51.55x_{1,2} + 56.55x_{1,3} + 41.55\acute{x}_{1,1} + 61.55\acute{x}_{1,3}\right] \\
+ \frac{1}{8}&\left[19.5653x_{2,1} + 4 \times 24.5653x_{2,2} + 29.5653x_{2,3} + 14.5653\acute{x}_{2,1} + 34.5653\acute{x}_{2,3}\right] \\
+ \frac{1}{8}&\left[36.24x_{3,1} + 4 \times 41.24x_{3,2} + 46.24x_{3,3} + 31.24\acute{x}_{3,1} + 51.24\acute{x}_{3,3}\right] \\
+ \frac{1}{8}&\left[10.465x_{4,1} + 4 \times 15.465x_{4,2} + 20.465x_{4,3} + 5.465\acute{x}_{4,1} + 25.465\acute{x}_{4,3}\right] \\
+ \frac{1}{8}&\left[174.892x_{5,1} + 4 \times 177.892x_{5,2} + 182.892x_{5,3} + 167.892\acute{x}_{5,1} + 187.892\acute{x}_{5,3}\right].
\end{aligned}
\tag{14}
$$

Subject to

$$
\begin{aligned}
&x_{i,1} \geq 0, x_{i,2} \geq 0, x_{i,3} \geq 0, \acute{x}_{i,1} \leq 0, \acute{x}_{i,3} \leq 0 \, where \, i = 1, 2, 3, 4, 5 \\
&0.007x_{1,1} - 0.007x_{2,3} + 0.013x_{4,1} + 0.0062x_{5,1} \leq 670, \\
&0.007x_{1,2} - 0.007x_{2,2} + 0.013x_{4,2} + 0.0062x_{5,2} \leq 720, \\
&0.007x_{1,3} - 0.007x_{2,1} + 0.013x_{4,3} + 0.0062x_{5,3} \leq 770, \\
&0.007\acute{x}_{1,1} - 0.007\acute{x}_{2,3} + 0.013\acute{x}_{4,1} + 0.0062\acute{x}_{5,1} \leq 620, \\
&0.007\acute{x}_{1,2} - 0.007\acute{x}_{2,2} + 0.013\acute{x}_{4,2} + 0.0062\acute{x}_{5,2} \leq 820, \\
&x_{1,1} - x_{2,3} \geq 350, \\
&x_{1,2} - x_{2,2} \geq 400, \\
&x_{1,3} - x_{2,1} \geq 450, \\
&0.03x_{2,1} - x_{3,3} + 0.07x_{4,1} = 0, \\
&0.03x_{2,2} - x_{3,2} + 0.07x_{4,2} = 0, \\
&0.03x_{2,3} - x_{3,1} + 0.07x_{4,3} = 0, \\
&0.03\acute{x}_{2,1} - \acute{x}_{3,3} + 0.07\acute{x}_{4,1} = 0, \\
&0.03\acute{x}_{2,3} - \acute{x}_{3,1} + 0.07\acute{x}_{4,3} = 0, \\
&x_{1,1} - \acute{x}_{2,3} \geq 300, \\
&\acute{x}_{1,3} - \acute{x}_{2,1} \geq 500, \\
&0.97x_{2,1} - x_{4,3} = 0, \\
&0.97x_{2,2} - x_{4,2} = 0, \\
&0.97x_{2,3} - x_{4,1} = 0, \\
&0.97\acute{x}_{2,1} - \acute{x}_{4,3} = 0, \\
&0.97\acute{x}_{2,3} - \acute{x}_{4,1} = 0, \\
&0.93x_{4,1} - x_{5,3} \geq 550, \\
&0.93x_{4,2} - x_{5,2} \geq 600, \\
&0.93x_{4,3} - x_{5,1} \geq 650, \\
&0.93\acute{x}_{4,1} - \acute{x}_{5,3} \\
&0.93\acute{x}_{4,3} - \acute{x}_{5,1} \geq 700, \\
&0.96x_{5,1} \geq 19950, \\
&0.96x_{5,2} \geq 20000, \\
&0.96x_{5,3} \geq 20050, \\
&0.96\acute{x}_{5,1} \geq 19900, \\
&0.96\acute{x}_{5,3} \geq 20100.
\end{aligned}
\tag{15}
$$

The optimal solution of (13) is

$$
\begin{aligned}
\widetilde{x}_1^{\curlyvee} &= (24111.7, 24159.4, 24207.1; 24064, 24159.4, 24254.8), \\
\widetilde{x}_2^{\curlyvee} &= (23757.1, 23759.4, 23761.7; 23754.8, 23759.4, 23764), \\
\widetilde{x}_3^{\curlyvee} &= (874.161, 874.107, 874.054; 874.125, 874.107, 874), \\
\widetilde{x}_4^{\curlyvee} &= (23048.8, 23046.6, 23044.4; 23051.1, 23046.6, 23024.1), \\
\widetilde{x}_5^{\curlyvee} &= (20781.3, 20833.3, 20885.4; 20729.2, 20833.3, 20937.5).
\end{aligned}
\tag{16}
$$

By substituting the values of $\widetilde{x}_i^{\curlyvee}$'s in objective function (10), the minimal triangular intuitionistic fuzzy energy cost is calculated as follows:

$$
\begin{aligned}
\min \widetilde{\alpha}^{\curlyvee} = \left(\sum_{j=1}^{5} \widetilde{C_j^{\curlyvee}} \otimes \widetilde{x}_j^{\curlyvee}\right) = &(5463813.64, 5927615.1, 6403249.54; \\
&4979403.76, 5927615.1, 7880260.7)
\end{aligned}
$$

$$
\mu_{\widetilde{Q}_i^{\curlyvee}}\left(\widetilde{\alpha}^{\curlyvee}\right) = \begin{cases} \dfrac{x - 5927615.1}{5927615.1 - 5463813.64} & 5463813.64 \leq x < 5927615.1 \\[2mm] 1, & x = 5927615.1 \\[2mm] \dfrac{6403249.54 - x}{6403249.54 - 5463813.64} & 5927615.1 < x \leq 6403249.54 \\[2mm] 0, & \text{otherwise,} \end{cases}
$$

$$
\nu_{\widetilde{Q}_i^{\curlyvee}}\left(\widetilde{\alpha}^{\curlyvee}\right) = \begin{cases} \dfrac{5927615.1 - x}{5927615.1 - 4979403.76} & 4979403.76 \leq x < 5927615.1 \\[2mm] 0, & x = 5927615.1 \\[2mm] \dfrac{x - 5927615.1}{7880260.7 - 5927615.1} & 5927615.1 < x \leq 7880260.7 \\[2mm] 1, & \text{otherwise.} \end{cases}
$$

$$
\tag{17}
$$

5.1. Postoptimal Analysis.

Table 2 elaborates the flexibility of the presented model in intuitionistic fuzzy environment. Within the provided range for each variable, the model remains

TABLE 2: Optimal feasible range.

Variable	Final value	Reduced cost	Objective	Allowable increase	Allowable decrease	Range
$x_{1,1}$	24111.6857	0	5.819	1E + 30	5.819	5.819
$x_{1,2}$	24159.37627	0	25.775	1E + 30	25.775	25.775
$x_{1,3}$	24207.06684	0	7.069	1E + 30	7.069	7.069
$x_{1,1}'$	24063.99512	0	5.194	1E + 30	5.194	5.194
$x_{1,3}'$	24254.75742	0	7.694	1E + 30	7.694	7.694
$x_{2,1}$	23757.06684	0	2.46	1E + 30	12.2144187	12.2144187
$x_{2,2}$	23759.37627	0	12.283	1E + 30	46.3171348	46.3171348
$x_{2,3}$	23761.6857	0	3.696	1E + 30	10.9589062	10.9589062
$x_{2,1}'$	23754.75742	0	1.821	1E + 30	12.82117495	12.82117495
$x_{2,3}'$	23763.99512	0	4.321	1E + 30	10.33814995	10.33814995
$x_{3,1}$	874.1610548	0	4.53	1E + 30	365.2968733	365.2968733
$x_{3,2}$	874.107453	0	20.62	1E + 30	1258.959902	1258.959902
$x_{3,3}$	874.0538512	0	5.78	1E + 30	407.14729	407.14729
$x_{3,1}'$	874.2146565	0	3.905	1E + 30	344.6049983	344.6049983
$x_{3,3}'$	874.0002494	0	6.405	1E + 30	427.3724983	427.3724983
$x_{4,1}$	23048.83513	0	1.308	1E + 30	11.29784144	11.29784144
$x_{4,2}$	23046.59498	0	7.7325	1E + 30	47.74962351	47.74962351
$x_{4,3}$	23044.35484	0	2.558	1E + 30	12.59218423	12.59218423
$x_{4,1}'$	23051.07527	0	0.683	1E + 30	10.65788655	10.65788655
$x_{4,3}'$	23042.1147	0	3.183	1E + 30	13.21770613	13.21770613
$x_{5,1}$	20781.25	0	21.8615	1E + 30	35.40148304	35.40148304
$x_{5,2}$	20833.33333	0	88.946	1E + 30	140.2896812	140.2896812
$x_{5,3}$	20885.41667	0	22.8615	1E + 30	35.00971661	35.00971661
$x_{5,1}'$	20729.16667	0	20.9865	1E + 30	35.19908724	35.19908724
$x_{5,3}'$	20937.5	0	23.4865	1E + 30	34.94659306	34.94659306

TABLE 3: Limit report.

Variable	Final value	Lower limit	Objective value	Upper limit	Objective value
$x_{1,1}$	24111.6857	24111.6857	5933823.792	58264.74189	6132560.426
$x_{1,2}$	24159.37627	24159.37627	5933823.792	65363.31892	6995855.414
$x_{1,3}$	24207.06684	24207.06684	5933823.792	72461.89595	6274937.179
$x_{1,1}'$	24063.99512	24063.99512	5933823.792	51166.16486	6074592.462
$x_{1,3}'$	24254.75742	24254.75742	5933823.792	79560.47298	6359345.968
$x_{2,1}$	23757.06684	23757.06684	5933823.792	23757.06684	5933823.792
$x_{2,2}$	23759.37627	23759.37627	5933823.792	23759.37627	5933823.792
$x_{2,3}$	23761.6857	23761.6857	5933823.792	23761.6857	5933823.792
$x_{2,1}'$	23754.75742	23754.75742	5933823.792	23754.75742	5933823.792
$x_{2,3}'$	23763.99512	23763.99512	5933823.792	23763.99512	5933823.792
$x_{3,1}$	874.1610548	874.1610548	5933823.792	874.1610548	5933823.792
$x_{3,2}$	874.107453	874.107453	5933823.792	874.107453	5933823.792
$x_{3,3}$	874.0538512	874.0538512	5933823.792	874.0538512	5933823.792
$x_{3,1}'$	874.2146565	874.2146565	5933823.792	874.2146565	5933823.792
$x_{3,2}'$	874.0002494	874.0002494	5933823.792	874.0002494	5933823.792
$x_{4,1}$	23048.83513	23048.83513	5933823.792	23048.83513	5933823.792
$x_{4,2}$	23046.59498	23046.59498	5933823.792	23046.59498	5933823.792
$x_{4,3}$	23044.35484	23044.35484	5933823.792	23044.35484	5933823.792
$x_{4,1}'$	23051.07527	23051.07527	5933823.792	23051.07527	5933823.792
$x_{4,3}'$	23042.1147	23042.1147	5933823.792	23042.1147	5933823.792
$x_{5,1}$	20781.25	20781.25	5933823.792	20781.25	5933823.792
$x_{5,2}$	20833.33333	20833.33333	5933823.792	20833.33333	5933823.792
$x_{5,3}$	20885.41667	20885.41667	5933823.792	20885.41667	5933823.792
$x_{5,1}'$	20729.16667	20729.16667	5933823.792	20729.16667	5933823.792
$x_{5,3}'$	20937.5	20937.5	5933823.792	20937.5	5933823.792

optimal and feasible. Since $x_i^\gamma = (x_{i,1}, x_{i,2}, x_{i,3}; x_{i,1}', x_{i,2}, x_{i,3}')$; therefore, the allowable change in $x_{(i,j)}$ will vast the range for x_i^γ as well. This change will not impact the feasibility of the given model.

As shown in Table 3, the optimal results changed accordingly within the limits without impacting feasibility. The limit report of this model provided lower limit and upper limit for each decision variable along with its corresponding

TABLE 4: Sensitivity report.

Constraint LHS	Final value	Shadow price	Constraint RHS	Allowable increase	Allowable decrease
1	430.9286066	0	670	1E + 30	239.0713934
2	431.5724014	0	720	1E + 30	288.4275986
3	432.2161962	0	770	1E + 30	337.7838038
4	430.2848118	0	620	1E + 30	189.7151882
5	432.859991	0	820	1E + 30	387.140009
6	350	5.819	350	34153.0562	24111.6857
7	400	25.775	400	41203.94265	24159.37627
8	450	7.069	450	48254.82911	24207.06684
9	300	5.194	300	27102.16974	24063.99512
10	500	7.694	500	55305.71557	24254.75742
11	−1.7053E-13	−5.78	0	874.0538512	1E + 30
12	−1.13687E-13	−20.62	0	874.107453	1E + 30
13	3.97904E-13	−4.53	0	874.1610548	1E + 30
14	−2.84217E-13	−6.405	0	874.0002494	1E + 30
15	2.27374E-13	−3.905	0	874.2146565	1E + 30
16	−7.27596E-12	10.00247423	0	1E + 30	23044.35484
17	−3.63798E-12	39.87278351	0	1E + 30	23046.59498
18	−7.27596E-12	9.949381443	0	1E + 30	23048.83513
19	−3.63798E-12	10.00737113	0	1E + 30	23042.1147
20	−3.63798E-12	9.930051546	0	1E + 30	23051.07527
21	19950	36.87654483	19950	23197.95231	19950
22	20000	146.1350846	20000	13722.05904	20000
23	20050	36.4684548	20050	16418.68769	20050
24	19900	36.66571588	19900	26587.58462	19900
25	20100	36.4027011	20100	13029.05538	20100
26	550	12.14821661	550	17102.79968	21435.41667
27	600	51.34368119	600	20633.66667	21433.33333
28	650	13.53998304	650	24164.53365	21431.25
29	500	11.46009306	500	13571.93269	21437.5
30	700	14.21258724	700	27695.40064	21429.16667
31	24111.6857	0	0	24111.6857	1E + 30
32	24159.37627	0	0	24159.37627	1E + 30
33	24207.06684	0	0	24207.06684	1E + 30
34	24063.99512	0	0	24063.99512	1E + 30
35	24254.75742	0	0	24254.75742	1E + 30
36	23757.06684	0	0	23757.06684	1E + 30
37	23759.37627	0	0	23759.37627	1E + 30
38	23761.6857	0	0	23761.6857	1E + 30
39	23754.75742	0	0	23754.75742	1E + 30
40	23763.99512	0	0	23763.99512	1E + 30
41	874.1610548	0	0	874.1610548	1E + 30
42	874.107453	0	0	874.107453	1E + 30
43	874.0538512	0	0	874.0538512	1E + 30
44	874.2146565	0	0	874.2146565	1E + 30
45	874.0002494	0	0	874.0002494	1E + 30
46	23048.83513	0	0	23048.83513	1E + 30
47	23046.59498	0	0	23046.59498	1E + 30
48	23044.35484	0	0	23044.35484	1E + 30
49	23051.07527	0	0	23051.07527	1E + 30
50	23042.1147	0	0	23042.1147	1E + 30
51	20781.25	0	0	20781.25	1E + 30
52	20833.33333	0	0	20833.33333	1E + 30
53	20885.41667	0	0	20885.41667	1E + 30
54	20729.16667	0	0	20729.16667	1E + 30
55	20937.5	0	0	20937.5	1E + 30

objective values. As the main objective is to minimize the energy cost, so here the optimal result is totally based on the lower limits to obtain minimum output. Whereas the upper limit is providing the range for feasible solution. Table 4 presents the sensitivity report for the constraint equation. As there is not always perfect availability of resources and time so for each situation making new model is not possible. The following report allows us to predict the flexibility regarding

TABLE 5: Comparison of optimization techniques.

Variable	LP	FLP	FIF LP
x_1	24195.38	(24126.743, 24159.376, 24192.008)	(24111.69, 24159.38, 24207.07; 24064, 24159.38, 24254.76)
x_2	23759.38	(23736.743, 23759.376, 23782.008)	(23757.07, 23759.38, 223759.38; 23754.76,23759.38, 23764)
x_3	2326.043	(2323.872, 2326.0429, 2328.2586)	(874.1611, 874.1075, 874.0539; 874.2147, 874.1075, 874.0002)
x_4	23046.59	(23024.64, 23046.59, 23068.55)	(23048.84, 23046.59, 23044.35; 23051.08, 23046.59, 23042.11)
x_5	20833.33	(20822.92, 20833.33, 20843.75)	(20781.25, 20833.33333, 20885.42; 20729.15, 20833.33333, 20937.5)
Objective function	5987499	5980612.1	5927615.1
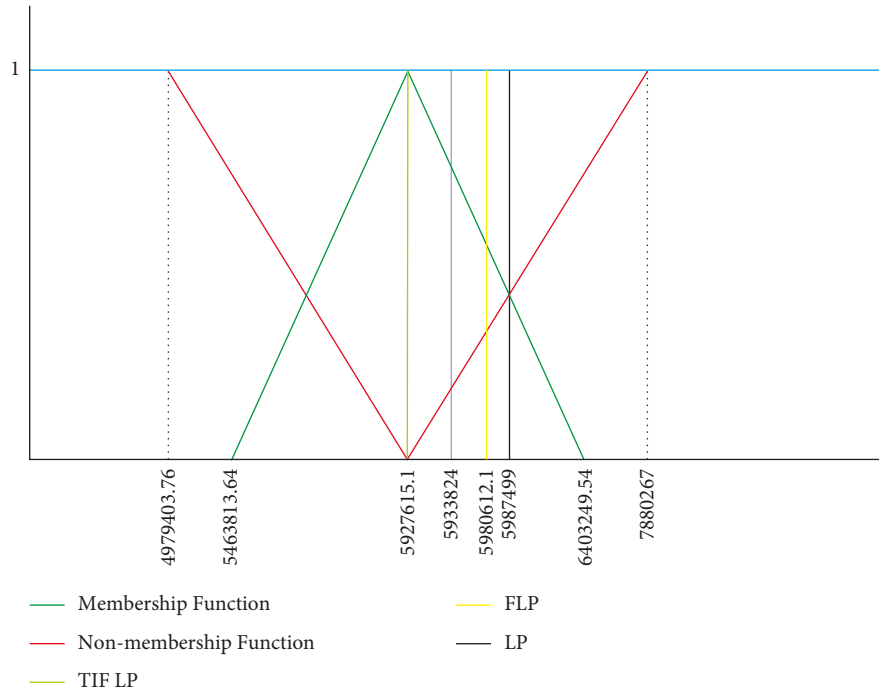
FIGURE 5: Graphical representation of LP, FLP, and FIF LP.

the optimal and feasible results we could have obtained if the availability of resources fluctuate between the allowable increase and decrease.

5.2. Comparison. The solution of linear optimization problem under consideration is carried out by already existing linear programming (LP) and fuzzy linear programming techniques (FLP). The objective function value for the proposed method is less than LP and FLP techniques. The results are presented in Table 5.

From Figure 5, the degree of acceptance (rejection) of energy cost per month increases (decreases) from 5463813.64 to 5927615.1 and decrease (increases) from 5927615.1 to 6403249.54, while 5927615.1 is considered as required value where the level of acceptance is fully satisfied and degree of rejection is fully zero. The optimal solution of linear programming and fuzzy linear programming whose resulting cost are PKR 5987499 per month and 5980612.1 per month, while for proposed method it is 5927615.1 per month. This is evident that the fully intuitionistic fuzzy optimization model is more effectively minimize the energy cost as compared to linear programming approach.

6. Conclusion

Since Pakistan has a huge textile industrial sector and is now facing so many uncertainties due to the unpredicted policy shift 2020–25, therefore economically improving this sector will help every person belonging to the textile industry's hierarchy from consumers, labourers to stakeholder, and government. In Pakistan, government suggests industries to have a local supply of gas, which will create an explosion in production costs. According to the executive director of all Pakistan mills associations (APTAMA), the textile industry will face a 50% increase in the production cost [40]. The best option for Pakistan's industrial sector is to become self-sufficient as soon as possible. For this purpose, they need to change their habit of being dependent on the government for their energy resources. It is only possible to shift their energy modes from nonrenewable to renewable ones for sustainable energy production. According to regional temperature conditions of Pakistan as shown in Figure 2, solar energy production refines option. Since Pakistan's Kohinoor Textile Mills (KTML) in Rawalpindi, Pakistan, converted to a 6 Mw solar power plant with Reon energy making it affordable, sustainable, and more competitive. The implementation of

this solar plant resulted in cost reduction and carbon footprint reduction. Because of its success, further three projects are also covered by KTML from 2017 to 2019, and their captive solar plant is saving almost 30% energy as well [41]. Therefore in Punjab, Pakistan, where there is a hub of textile industries should take a step forward towards the solar power plants.

The application of this solar plant project approximately requires 65 to 80 million in Pakistan. This project will contain a one Mw system that can produce 5000 units per day if minimum five hours of sunlight is considered. Monthly electricity unit production from this system will approximate between 150000 kWh and 180000 kWh. If LP solution is considered, the electricity unit per month consumed by all stages are 109108.6563 kWh and from FIFLP 106204.7783 kWh. In both cases, monthly electricity unit production through a solar plant is greater than need. The remaining electricity units can be sold out to the government or nearby industries that will accelerate to overcome the investing cost and the amount obtained from sold electricity can be further used for maintenance purpose as well. Decision making through FIFLP provide more flexible and optimal outputs. The less units industry will consume the more will be available for sale and soon the invested cost will overcome and providing profit. The conversion of few Pakistan's textile industry on solar power plants will help their neighboring industries, because the electricity cost per unit in solar generators is cheaper than other. That is, the government provide electricity at the rate of 20.62 Rs/Kwh while solar generating electricity cost is between 8 and 14 Rs/kWh.

The idea of intuitionistic fuzzy model helps in sufficient allocation of energy units that provides less wastage of energy. The implementation of this modelling can definitely lower down the production cost. For pilot study the implementation of sustainable energy optimization model on one of the industry from that region mentioned in Figure 1 can provide fruitful results along with the optimizing production cost of the neighboring industries as well.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that there were no conflicts of interest regarding the publication of this article.

Authors' Contributions

All authors contributed equally to the preparation of this manuscript.

References

[1] R. S. Blackburn, Ed., *Sustainable Textiles: Life Cycle and Environmental Impact*, vol. 235, pp. 369–380, Woodhead, Cambridge, 2009.

[2] M. R. Bhuiyan, M. M. Rahman, A. Shaid, M. Bashar, and M. A. Khan, "Scope of reusing and recycling the textile wastewater after treatment with gamma radiation," *Journal of Cleaner Production*, vol. 112, pp. 3063–3071, 2016.

[3] Redesigning Fashion's Future [Internet] E MacArthur Foundation, "A New Textiles Economy," 2021, https://www.ellenmacarthurfoundation.org/assets/downloads/publications/A-New-Textiles-Economy-Full-Report.pdf.

[4] A. K. R. Choudhury, *Environmental impacts of the textile industry and its assessment through life cycle assessment. In: Muthu S. S., editor. Roadmap to Sustainable Textiles and Clothing—Environmental and Social Aspects of Textiles and Clothing Supply Chain*, vol. 235, pp. 1–39, Springer Science-Business Media, Singapore, 2014.

[5] K. Nadeem, G. T. Guyer, B. Keskinler, and N. Dizge, "Investigation of segregated wastewater streams reusability with membrane process for textile industry," *Journal of Cleaner Production*, vol. 228, pp. 1437–1445, 2019.

[6] M. J. Dehghani and C. K. Yoo, *Modeling and extensive analysis of the energy and economics of cooling, heat, and power trigeneration (CCHP) from textile wastewater for industrial low-grade heat recovery*, Energy Conversion and Management, vol. 205, , 2020.

[7] A. R. Nagaraj, "Energy Management in Textile Industry," 2012, https://www.fibre2fashion.com/industry-article/6276/energy-management.

[8] B. Sudhakara Reddy and B. Kumar Ray, "Understanding industrial energy use: physical energy intensity changes in Indian manufacturing sector," *Energy Policy*, vol. 39, no. 11, pp. 7234–7243, 2011.

[9] A. S. M. Raja, A. Arputharaj, S. Saxena, and P. G. Patil, *Water requirement and sustainability of textile processing industries. In: Muthu S. S., editor. Water in Textiles and Fashion: Consumption, Footprint, and Life Cycle Assessment*, pp. 155–173, Woodhead, Cambridge, 2019.

[10] World of Garment-Textile-Fashion, "A facelift to textile machinery with renewable energy," 2013, Retrieved from http://fibre2fashion.com/sep 2013.

[11] Textile industry to adopt renewable energy solution: The Frontier Post, https://thefrontierpost.com/textile-industry-to-adopt-renewable-energy-solution/.

[12] GOP (Government of Pakistan), "Pakistan's Monthly Climate Summary," Pakistan Meteorological Department, June 2021, Retrieved from https://www.pmd.gov.pk/.

[13] L. V. Kantorovich, "Mathematical methods of organizing and planning production," *Management Science*, vol. 6, no. 4, pp. 366–422, 1960.

[14] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, p. 338, 1965.

[15] H. J. Zimmermann, "Fuzzy programming and linear programming with several objective functions," *Fuzzy Sets and Systems*, vol. 1, no. 1, pp. 45–55, 1978.

[16] K. T. Atanassov, "Intuitionistic fuzzy sets," *Fuzzy Sets and Systems*, vol. 20, no. 1, pp. 87–96, 1986.

[17] G. Negi, A. Kumar, S. Pant, and M. Ram, "Optimization of complex system reliability using hybrid Grey Wolf optimizer," *Decision Making: Applications in Management and Engineering*, vol. 4, no. 2, pp. 241–256, 2021.

[18] P. P. Angelov, "Optimization in an intuitionistic fuzzy environment," *Fuzzy Sets and Systems*, vol. 86, no. 3, pp. 299–306, 1997.

[19] R. J. Hussain and P. S. Kumar, "Algorithmic approch for solving intuitionistic fuzzy transportation problem," *Applied Mathematical Sciences*, vol. 80, no. 6, pp. 3981–3989, 2012.

[20] A. Ebrahimnejad and J. L. Verdegay, "A new approach for solving fully intuitionistic fuzzy transportation problems," *Fuzzy Optimization and Decision Making*, vol. 17, no. 4, pp. 447–474, 2018.

[21] H. T. Nguyen and F. A. Felder, "Generation expansion planning with renewable energy credit markets: a bilevel programming approach," *Applied Energy*, vol. 276, 2020.

[22] A. Cosic, M. Stadler, M. Mansoor, and M. Zellinger, "Mixed-integer linear programming based optimization strategies for renewable energy communities," *Energy*, vol. 237, 2021.

[23] M. Saeid Atabaki, M. Mohammadi, and V. Aryanpur, "An integrated simulation-optimization modelling approach for sustainability assessment of electricity generation system," *Sustainable Energy Technologies and Assessments*, vol. 52, 2022.

[24] K. Bio Gassi and M. Baysal, "Analysis of a linear programming-based decision-making model for microgrid energy management systems with renewable sources," *International Journal of Energy Research*, vol. 46, no. 6, pp. 7495–7518, 2022.

[25] F. Keck, S. Jütte, M. Lenzen, and M. Li, "Assessment of two optimisation methods for renewable energy capacity expansion planning," *Applied Energy*, vol. 306, 2022.

[26] M. Rabe, Y. Bilan, K. Widera, and L. Vasa, "Application of the linear programming method in the construction of a mathematical model of optimization distributed energy," *Energies*, vol. 15, no. 5, p. 1872, 2022.

[27] Z. Foroozandeh, S. Ramos, J. Soares, and Z. Vale, "Goal programming approach for energy management of smart building," *IEEE Access*, vol. 10, pp. 25341–25348, 2022.

[28] J. Nematian and Y. Farzi, "Future energy and environment management using fuzzy linear programming including energy recovered from urban solid wastes (Case study: Iran country)," *Environmental Progress & Sustainable Energy*, vol. 36, no. 3, pp. 953–966, 2017.

[29] C. Zhou, G. Huang, and J. Chen, "A type-2 fuzzy chance-constrained fractional integrated modeling method for energy system management of uncertainties and risks," *Energies*, vol. 12, no. 13, p. 2472, 2019.

[30] N. Kouaissah and A. Hocine, "Optimizing sustainable and renewable energy portfolios using a fuzzy interval goal programming approach," *Computers & Industrial Engineering*, vol. 144, 2020.

[31] A. Hashemizadeh and Y. Ju, "Optimizing renewable energy portfolios with a human development approach by fuzzy interval goal programming," *Sustainable Cities and Society*, vol. 75, 2021.

[32] M. F. Khan, A. Pervez, U. M. Modibbo, J. Chauhan, and I. Ali, "Flexible fuzzy goal programming approach in optimal mix of power generation for socio-economic sustainability: a case study," *Sustainability*, vol. 13, no. 15, p. 8256, 2021.

[33] S. Emeç and G. Akkaya, "Developing a new optimization energy model using fuzzy linear programming," *Journal of Intelligent and Fuzzy Systems*, vol. 40, no. 5, pp. 9529–9542, 2021.

[34] S. Abbas, L. H. Chiang Hsieh, L.-H. Chiang Hsieh, and K. Techato, "Supply chain integrated decision model in order to synergize the energy system of textile industry from its resource waste," *Energy*, vol. 229, 2021.

[35] M. B. Toledo, C. Torres Torres, J. C. Llivisaca, M. Peña, L. Siguenza-Guzman, and J. Veintimilla-Reyes, "Optimization models used in the textile sector: a systematic review," *Communications in Computer and Information Science*, vol. 1535 CCIS, pp. 202–213, 2022.

[36] D. Dubois and H. Prade, *Fuzzy Sets and Systems: Theory and Application*, Academic Press, Newyork, 1980.

[37] I. Kimutai, P. Maina, and A. Makokha, "Energy optimization model using linear programming for process industry: a case study of textile manufacturing plant in Kenya," *International Journal of Energy Engineering*, vol. 9, no. 2, pp. 45–52, 2019.

[38] NEPRA: Annual Reprot 2020-21. Retrieved from https://nepra.org.pk/publications/Annual%20Reports/Annual%20Report%202020-21.pdf.

[39] OGRA GOP, "Notified petroleum prices," 2021, Retrieved from https://ogra.org.pk/notified-petroleum-prices.

[40] K.Mustafa,Despite go-ahead by PM twice: Textile Policy 2020-25 in the doldrums,Jan 24(2021), https://Thenews.com.pk.

[41] https://www.texintel.com/eco-news/pakistans-kohinoor-textile-mills-change-to-6-mw-solar-power-with-reon-energy.

Research Article

Bus Single-Trip Time Prediction Based on Ensemble Learning

Haifeng Huang ⓘ,[1] **Lei Huang** ⓘ,[1] **Rongjia Song,**[2] **Feng Jiao,**[1] **and Tao Ai**[1]

[1]*Department of Information Management, School of Economics and Management, Beijing Jiaotong University,*
 Beijing 100044, China
[2]*Department of Information Management, School of Management, Hangzhou Dianzi University, Hangzhou 310018, China*

Correspondence should be addressed to Lei Huang; lhuang@bjtu.edu.cn

The prediction of bus single-trip time is essential for passenger travel decision-making and bus scheduling. Since many factors could influence bus operations, the accurate prediction of the bus single-trip time faces a great challenge. Moreover, bus single-trip time has obvious nonlinear and seasonal characteristics. Hence, in order to improve the accuracy of bus single-trip time prediction, five prediction algorithms including LSTM (Long Short-term Memory), LR (Linear Regression), KNN (K-Nearest Neighbor), XGBoost (Extreme Gradient Boosting), and GRU (Gate Recurrent Unit) are used and examined as the base models, and three ensemble models are further constructed by using various ensemble methods including Random Forest (bagging), AdaBoost (boosting), and Linear Regression (stacking). A data-driven bus single-trip time prediction framework is then proposed, which consists of three phases including traffic data analysis, feature extraction, and ensemble model prediction. Finally, the data features and the proposed ensembled models are analyzed using real-world datasets that are collected from the Beijing Transportation Operations Coordination Center (TOCC). Through comparing the predicting results, the following conclusions are drawn: (1) the accuracy of predicting by using the three ensemble models constructed is better than the corresponding prediction results by using the five sub-models; (2) the Random Forest ensemble model constructed based on the bagging method has the best prediction accuracy among the three ensemble models; and (3) in terms of the five sub-models, the prediction accuracy of LR is better than that of the other four models.

1. Introduction

Rapid economic growth is accompanied by increasingly serious traffic congestion on urban roads. The vigorous development of public transportation is considered as one of the effective means to alleviate traffic congestion. Many factors are considered as relevant to the acquisition of dynamic public transport information and the prediction of bus single-trip time, including the decision-making of passenger bus travel, the priority control of bus vehicles at intersections, and the intelligent scheduling and the real-time information dissemination of public transport [1]. Furthermore, the accurate prediction of bus single-trip time is the prerequisite and basis for achieving the bus priority and the intelligent scheduling, which could help to effectively reduce passenger waiting time and improve passenger

satisfaction, in order to further greatly impact the bus service level and bus travel attractiveness. More specifically, the role of accurate prediction of bus single-trip time is mainly reflected in the following aspects: providing real-time information such as bus arrival time for passengers, efficient management of bus fleets, improving bus service quality, and providing bus priority signals and information for traffic managers. Therefore, it is of great significance to explore the prediction method of bus single-trip time and improve the prediction accuracy of the bus running time, which could help to improve the attractiveness of public transport and promote the development of public transport [2].

The prediction of bus single-trip time has attracted extensive attention in the past decade or so, and various predicting methods have been proposed in the literature [3]. Among them, typical methods are mainly statistical models,

such as Auto Regressive Integrated Moving Average (ARIMA) models [4], Seasonal Auto Regressive Integrated Moving Average (SARIMA) model [5], the Grey Model (GM) [6], Kalman filtering [7], and spectral analysis. Although the above statistical methods perform well in capturing linear relationships, they have limited ability to capture nonlinear features [8]. In recent years, many scholars have also turned to some nonlinear methods, such as Support Vector Regression (SVR) [9] and Artificial Neural Networks (ANNs) [10]. Moreover, popular ANN models in this field include Back Propagation Neural Networks (BPNNs) [11], Extreme Learning Machines (ELMs) [12], and Deep Belief Networks (DBNs).

Bus operation could be affected by many random factors (e.g., weather, traffic congestion, and passenger flow change), which makes it very difficult to predict the running time of a bus accurately. In terms of predicting the bus running time by using a single prediction model, the most popular methods include time series models, artificial neural networks, and Kalman filtering. (1) Time series models mainly rely on the similarity between future information and historical information, and when the average situation of historical data changes, it would lead to an obvious deviation of the prediction results. Moreover, time series models also have significant lags when making real-time prediction [13]. (2) Kalman filtering technique, which is formed by introducing the state space into modern control theory, has been applied to short-term traffic demand and travel time prediction of highways [14]. Since the Kalman can fully adapt to irregular changes (Kalman gain), it is more suitable for single-step predicting, but its prediction accuracy decreases significantly in multi-step predicting [15]. (3) Artificial neural network is a model that explores the nervous system functions of the human brain by modeling and linking neurons (i.e., the basic units of the human brain) to simulate the functions of the human brain. The principle is an artificial system with intelligent information processing functions such as learning, association, memory, and pattern recognition. Its unique nonlinear adaptive information processing capability makes it an effective way to solve complex combinatorial optimization problems [16]. However, the generalization ability of neural network algorithms is limited due to the structure determination, over- and under-learning, and local convergence problems of neural network algorithms [17]. Among the existing prediction methods, the multiple regression model has poor applicability and low predicting accuracy. Neural network models can fit nonlinear systems well, but a large historical dataset need to be trained before use, which also has limit performance on under-learning and over-learning, as well as local optimality. Support Vector Machine (SVM) has strong learning ability and fault tolerance ability, and its generalization ability is better than neural network. However, similar to neural network models, pretraining is also needed for SVM, which makes it difficult to be used for real-time prediction. The Kalman filter model is applicable for predicting online and perform well, but it is difficult to guarantee the accuracy of nonlinear and non-Gaussian state models. To sum up, in the area of bus single-trip time

prediction, each single method has advantages and disadvantages. In order to further improve the prediction accuracy, this study adopts the idea of ensemble learning and concentrates the prediction advantages of several single prediction models by building an ensemble model.

Ensemble learning is a widely used approach in prediction using ensemble predictive models in machine learning. It is based on the principle of integrating different sets of learners for improving prediction accuracy [18]. The dominant area of research by scholars is currently designing ensemble models that enhance weak learners to strong learners and ensemble multiple learners generated by the same algorithm [18]. In ensemble learning, the prediction accuracy is greatly improved by combining multiple learners and the ensemble model performs better than each submodel. This result is due to the diversity among sub-models, which reduces the risk of using isolated models and meanwhile compensates the weakness of sub-models. In addition, ensemble models can solve many problems that individual models cannot solve. For instance, the transit running time of urban public transportation is dynamic and stochastic. It is difficult for a single model to fit its trend well, and ensemble learning can better compensate for this deficiency. Bus single-trip time prediction has attracted much attention in the recent years, and some major challenges about bus single-trip time prediction have been pointed out as follows [19]:

(1) It is necessary to predict the bus single-trip time based on dynamic spatiotemporal and weather conditions.

(2) Further exploration of ensemble learning in bus single-trip time prediction is needed.

(3) It is necessary to provide a highly universal bus single-trip time prediction method for public transport managers and passengers.

Toward meeting the above challenges, this study focuses on predicting the bus single-trip time using ensemble models. Based on the validation analysis using real-world data, we compare and analyze the prediction effects of five sub-models including LSTM (Long short-term memory), LR (Linear Regression), KNN (K-Nearest Neighbor), XGBoost (Extreme Gradient Boosting), and GRU (Gate Recurrent Unit). The ensemble learning models are constructed by using three ensemble methods, and a data-driven prediction framework is further proposed. The main contributions of this paper are summarized as follows:

(1) A data-driven predicting framework is proposed for bus runtime prediction, which consists of three phases: (a) data analysis-preparing bus runtime data from TOCC; (b) feature extraction-extracting key features for forecasting based on bus runtime time series and external data; (c) feature extraction-extracting key features required for predicting based on bus runtime time series and external data; and (d) prediction modeling-constructing three ensemble models for predicting.

(2) The ensemble learning method was introduced into the short-term transit runtime prediction model based on the three-model ensemble methods including bagging (Random Forest), boosting (Ada-Boost), and stacking (Linear Regression).

(3) Based on the real-world bus single-trip time data, after case analysis, the prediction results of LSTM, LR, KNN, XGBoost, and GRU and three ensemble models are compared and analyzed, and the advantages and disadvantages of these prediction models in bus single-trip time prediction are summarized.

This study significantly improves the bus single-trip time prediction accuracy by ensemble learning and provides a new modeling method for quantitative bus research, significant for theoretical guidance and methodological innovation.

This paper is organized as follows. Section 1 is the introduction, which mainly introduces the research background, motivation, and contributions. Section 2 is the literature review, which mainly sorts out the existing methods in the field of bus single-trip time prediction and points out the defects and deficiencies in current research. In Section 3, the methodology of this paper is proposed, which involves three ensemble models constructed based on five single prediction algorithms by using three-model ensemble methods. Furthermore, the case analysis is presented in Section 4, which includes real-world data description, experimental procedures by employing the methodology, and comparative analysis of prediction results. Then, the discussion comes in Section 5; according to the analysis results, the advantages and disadvantages of the three ensemble models and five single models are summarized. And the capability of applicability and generalization of the proposed methodology is discussed and elaborated. Finally, the paper is concluded in Section 5; the future perspectives are set as well.

1.1. Related Works. So far, scholars have proposed various operational and predictive models of public transportation. Popular models include regression models, time series (TS) models, ANN models, SVM models, traffic simulation and dynamic traffic simulation, and dynamic traffic assignment models [20]. Ghosh et al. [21] considered the effect of data fluctuations in the TS and established an autoregressive moving average TS model to predict the vehicle running time by data fitting and residual analysis. However, the large white noise in the data residual series negatively affects the prediction accuracy, indicating that the model lacks the treatment of the complexity and variability of urban traffic (Gu et al. [22]). A nonlinear regression model was developed considering factors such as road section length, traffic density, number of bus stops, and the vehicle turns. Agafonov and Yumaganov [23] developed a multiple linear regression model and an ANN model based on bus operation data in Samara and Russia by using real-time dynamic traffic data and historical data of bus stopping time. In the

analysis of the results, it was found that the ANN model possessed higher prediction accuracy and lower prediction error. Yu et al. [24] improved the SVM algorithm by introducing a decay factor to reduce the prediction error through the decay factor dynamically. Bie et al. [25] performed bus single-trip time prediction based on bus GPS data, but the model built did not consider the real-time status of traffic flow and roads. Dhivya Bharathi et al. [26] used the historical data averaging method and TS method for predicting the transit section running time but lacked influencing factors. Chang et al. [27] predicted the bus single-trip time using a regression algorithm based on historical data well, but the predictive model is quite complicated. Since bus single-trip time is greatly influenced by dynamic traffic conditions, Liu et al. [28] used a particle filtering algorithm to predict the bus arrival time with the nonlinear and non-Gaussian characteristics in real-time. Hua et al. [29] proposed a bus journey time prediction method based on SVR and interval upper and lower bound estimation methods, which considered the uncertain factors in the bus operation. However, the method only turned the point prediction values into prediction intervals and did not study the road traffic status as a variable. Wu S. proposed the Adaptive Boosting (AdaBoost) algorithm [30]. The Bagging algorithm uses sequential sampling with high operational efficiency and practical applications. Bagging algorithm uses self-sampling combined with a base learner, which was improved to Random Forest (RF) in 2001 [31]. Wolpert proposed the stacked generalization model, but the stacking algorithm only provides the ensemble idea, and the choice of its base learner is somehow subjective [31].

From above, it is known that there are mainly predicting methods based on statistics, intelligent algorithms, and combined models for bus single-trip time prediction. Based on the current mainstream algorithms in classification, this study divides three major categories into six subcategories, among which the statistical methods include historical contemporaneous and time series; intelligent algorithms include machine learning and deep learning; combinatorial models include ensembled models and combined models. More specifically, the advantages and disadvantages of the commonly used models in each category are compared in Table 1.

Although single models are studied by many researchers and proved to be suitable for many cases, some shortcomings still exist. For example, in the study of TS, Billings and Yang [51] used the ARIMA model to predict the arterial travel time with the GPS data collected from Minnesota State Highway 194. ARIMA is a very simple time series forecasting method. It can only capture linear relationships in nature, but its capture of nonlinear relationships is not accurate. Moreover, ARIMA requires that the time series data are stationary or are stable after differencing (mean and variance are stable). For the bus single-trip time of this study, under the influence of various factors, the mean and variance of the data will change greatly even after the nonlinear data are differentiated. Therefore, ARIMA is not selected as a sub-model in this paper. The bus travel time was also predicted using the SARIMA model in [5]. However, these models

TABLE 1: Comparison of advantages and disadvantages of popular models.

Category	Commonly used models	Advantages	Disadvantages	References
History of the same period	Smoothing method	Easy understandability, better results in normal conditions and with large time granularity	Excessive reliance on data patterns from historical data	Omkar and Kumar [32]
Time series	Kalman filtering AR(Auto regressive) ARIMA	Applicable to time series data and interpretability	Unsuitable for capturing nonlinear data patterns	Zhou et al. [33] Li et al. [34] Gummadi and Edara [35]
Machine learning	SVM SVR K-nearest neighbor Linear regression Decision tree Random forest	Suitable for learning nonlinear features in data	Low computational efficiency at high data volumes	Li and Xu [36] Sun et al. [37] Khiari and Olaverri-Monreal [38] Alajali et al. [39] Zhou et al. [40]
Deep learning	RNN LSTM GRU	Applicable for learning linear and nonlinear patterns with good data fitting capability	Low interpretability and low efficiency	Pang et al. [41] Agafonov and Yumaganov [23, 42] Shu et al. [43]
Ensembled model	AdaBoost Bootstrapped aggregation Stacked generalization Gradient boosting Machines, GBM Gradient boosted regression Trees, GBRT	Applicable to select the appropriate base model for ensemble according to the characteristics of different datasets	Prone to overfitting, low interpretability, and poor results when data are unbalanced	Zhou et al. [44] Vaish et al. [45] Sharma et al. [46] Monego et al. [47, 48] Chen et al. [49]
Combined model	Direct averaging, weighted averaging, and other combinations	High applicability with various sub-models and combinations	Subjective on choosing the combination method and sub-models	Yan et al. [50]

have not considered the dynamic factors that affect bus travel times, such as the traffic condition. The Kalman filter models can obtain accurate prediction results with limited historical data; however, the accuracy decreases with increasing time steps [52]. Moreover, the performance of the Kalman filter models becomes unstable if there is big difference in the predicted value between two consecutive time windows [53]. The regression model requires a large amount of historical data for training to achieve the expected prediction accuracy [54]. Although machine learning and deep learning methods are also quite popular at present, their low training efficiency and interpretability make them less utilized in the study of bus single-trip time [19, 55]. Alternatively, it is a much better way to fuse the results from different predictive methods. Combined models could achieve better accuracy compared to each single predictor. The ensemble learning models have been proved to be able to achieve much better performance in prediction accuracy than individual ones. Nowadays, ensemble learning has been used in many fields of traffic prediction, such as traffic sign detection and recognition [56], traffic speed [52], short-term traffic volume [57], and traffic incident detection [58]. The advantage of ensemble learning is that the benefits of multiple learners can be integrated to improve the accuracy of predictions. The general approach is to generate multiple individual learners first and then ensemble them for

predicting using specific ensemble strategy [59]. A type of ensemble is called "homogeneous" if the individual learners are of the same type and "heterogeneous" if else. Individual learners need to be accurate and diverse [60, 61].

More specifically, the advantages of ensemble learning are as follows. Overall, ensemble learning has a high accuracy rate and good resistance to noise, which makes the model less prone to overfit due to the introduction of randomness [59]. It is insensitive to outliers, so it can handle very high dimensional data without the necessity to select features. Ensemble models can handle both discrete and continuous data, and in addition, the dataset does not need to be normalized [62]. Meanwhile, the overall training speed is relatively impressive. However, current ensemble methods are not explicitly designed for dealing with spatiotemporal data. Therefore, how to effectively ensemble multiple models while utilizing the spatiotemporal information remains a challenging, especially for practical problem in the real world. In addition, a few studies focus on using ensemble learning methods to predict bus single-trip time [63].

In summary, each single predictive model has certain advantages and disadvantages. Therefore, in this paper, three ensemble learning methods including Random Forest, AdaBoost, and Linear Regression are used to fuse the prediction results of sub-models. The prediction results of the ensemble models could increase the prediction stability even

if there are quite large deviations among the prediction results of the sub-models.

1.2. Methodology

1.2.1. Ensemble Model Construction of RF-Bagging.
Bagging is a method for learning multiple weak learners independently. Bagging is known as bootstrap aggregating. Bootstrap randomly selects n training samples from n training samples, which allows the generation of a repeated set of samples that are slightly different from the original training set.

Random Forest (RF) is an evolved version of the Bagging algorithm; the idea is still bagging but with unique improvements [64]. First, Random Forest uses the CART decision tree as a weak learner. Secondly, Random Forest improves the decision tree building based on the use of decision trees. For a normal decision tree, we choose an optimal feature among all the n sample features on the node to do the left and right subtree divisions of the decision tree. But RF selects a part of the sample features on the nodes by randomly selecting a number less than n, which is assumed to be n_{sub} And then, among these randomly selected n_{sub} Then, among these randomly selected sample features, an optimal feature is selected to do the left and right subtree partitioning of the decision tree. This further enhances the generalization ability of the model. If $n_{sub} = n$, then there is no difference between the CART decision tree of RF and the ordinary CART decision tree at this time. n_{sub} The smaller the model is, the more robust it is, but of course the fit to the training set become worse. In other words, n_{sub}, the smaller the model, the smaller the variance of the model will be reduced, but the bias will be increased. A suitable value is usually obtained by cross-validating the tuning parameters in practical cases. n_{sub} value of the model. The structure of the ensemble model based on the random forest is shown in Figure 1.

(1) The input is a sample set of $D = \{x_1, y_1, (x_2, y_2), \ldots, (x_m, y_m)\}$, and the number of weak classifier iterations is T.

(2) For $t = 1, 2, \ldots, T$: randomly sample the training set for the t th time, randomly select m features for each training set, and obtain a sampling set $\mathcal{D}_{bs}$ containing m feature samples.

(3) Using the sampling set $\mathcal{D}_{bs}$ to train the t th decision tree model $G_t(x)$, when training the nodes of the decision tree model, select a portion of sample features among all the sample features on the node, and choose an optimal feature among these randomly selected partial sample features for the left and right subtree partitioning of the decision tree.

(4) The regression algorithm is used, and the value obtained by arithmetic averaging the regression results obtained from T weak learners is the final model output.

The entire algorithm flow is described in Algorithm 1.

Using Random Forest algorithm for model ensemble has the following advantages: the training can be highly parallelized, which is advantageous for the training speed of large samples in the era of big data; the introduction of two randomness makes Random Forest have good anti-noise ability and is insensitive to partial feature missing; due to the use of random sampling, the variance of the trained model is small, and the generalization ability is strong. Compared with Adaboost and GBDT based on Boosting, the Random Forest algorithm is relatively simple to implement; since the decision tree nodes can be randomly selected to divide the features so that the model can still be trained efficiently when the sample features are of high dimensionality; wrong predictions are made only when more than half of the base learners are in error: Random Forest is very stable, even if a new data point appears in the dataset, the whole algorithm will not be affected too much, it will only affect one decision tree, and it is not easy to affect all decision trees.

However, features with more value divisions are likely to have a greater impact on RF decision-making, thereby affecting the effect of the model.

1.2.2. Ensemble Model Construction of AdaBoost-Boosting.
Boosting is a class of algorithms that boosts weak learners to strong learners. It is a serial idea where serialization is performed [61]. The basic idea is that increasing the weights of the samples that the previous base learner incorrectly predicted makes the subsequent base learners pay more attention to these mislabeled samples and correct these errors as much as possible. Until T base learners are trained, eventually, these T base learners are weighted and combined.

As shown in Figure 2, boosting adaptively fits multiple base learners in sequence; the current model training is based on the training results of the previous base learner, and the current base learner increases the weight of the mis-estimated samples, which in turn reduces the prediction error rate. Therefore, unlike bagging, boosting mainly focuses on reducing the bias of the model, and usually, the base model is chosen to have high bias and low variance. If a decision tree is chosen as the base model, most decision trees with shallow depth are selected, reducing the computational cost of model fitting.

The loss function used by AdaBoost is the exponential loss function, so the weights and sample distribution of AdaBoost revolve around minimizing the exponential loss function. This study defines the ensemble learner as a linear weighting of the base learner, where α is the weight of the base learner:

$$H(x) = \sum_{t=1}^{T} \alpha_t h_t(x). \tag{1}$$

Also, the exponential loss function defined by AdaBoost is

$$loss_{exp}(h) = E_{x \sim \mathcal{D}}, y\left[e^{-yh(x)}\right]. \tag{2}$$

The specific AdaBoosting iteration, in three steps.

FIGURE 1: Structure of the ensemble model based on Random Forest.

Input: Training set $D = \{x_1, y_1, (x_2, y_2), \ldots, (x_t, y_t)\}$;
            Sub-models $L = \{CART\,1, CART\,2, \ldots, CART\,T\}$;
            Number of learning rounds T;
(1)   for $t = 1, 2, \ldots, T$ do
(2) $h_t = L(D, D_{bs})$,# D_{bs} is the sample distribution generated by self-sampling
(3)   end for
      Output: $H(x) = \underset{y\varepsilon\gamma}{\arg\max} \sum_{t=1}^{T} (h_t(x) = y)$

ALGORITHM 1: RF-bagging ensemble model.
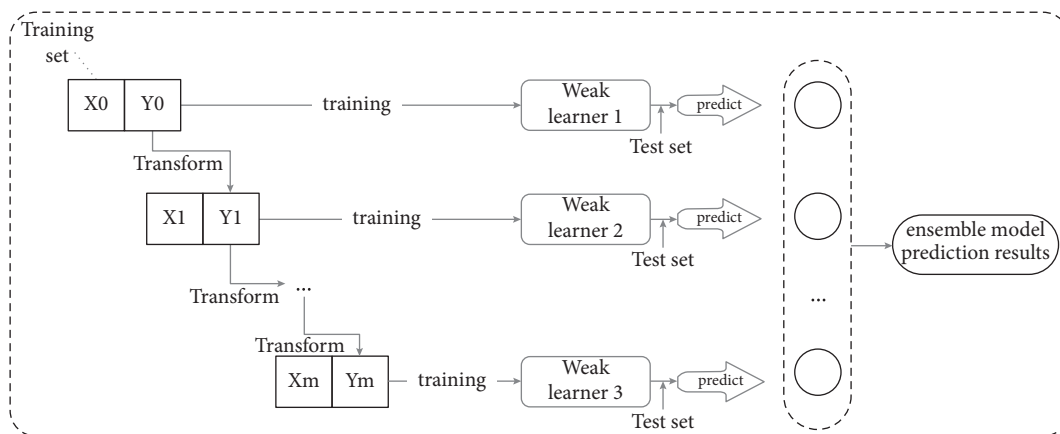
FIGURE 2: Structure of AdaBoost-based ensemble model.

(1) Initialize the weight distribution of the training data. If there are N samples, at first all samples are given the same weights: 1/N.

(2) Training the weak learner. In training, if a sample point has been learned accurately, the weight of that

sample point decreases in constructing the next training set; conversely, the weight of sample points that failed to be learned accurately increases.

(3) The weak learners obtained from each training are formed into strong learners. After each weak

Input: Training set $D = \{x_1, y_1, (x_2, y_2), \ldots, (x_m, y_m)\}$;
       Sub-models $\mathscr{L} = \{LSTM, LR, KNN, XGBoost, GRU\}$;
       Number of learning rounds T;
(1) $D_1(i) = 1/m$ # Initialize the weight distribution
(2)   for t $= 1, 2, \ldots, T$
(3) $h_t = \mathscr{L}(D, D_1)$, # Train a leaner h_t from D using # distribution D_t
(4) $\varepsilon_t = Pr_{x \sim D_t}, y[h_t(x) \neq y]$, # Measure the error of h_t
(5)   if $\varepsilon_t > 0.5$ then break
(6) $\alpha_t = 1/2 Ln(1 - \varepsilon_t/\varepsilon_t)$, # Determine the weight of h_t
(7) $D_{t+1}(i) = D_t(i)/Z_t \times \begin{cases} \exp(-\alpha_t) \, if \, h_t(x_i) = y_i \\ \exp(\alpha_t) \, if \, h_t(x_i) \neq y_i \end{cases} = D_t(i)\exp(-\alpha_t y_i h_t(x_i))/Z_t$,# Update the distribution, where Z_t is a nomolization factor that enables D_{t+1} to be distribution
(8)   End for
   Output: $H(x) = sign(\sum_{t=1}^{T} \alpha_t h_t(x))$

ALGORITHM 2: AdaBoost-Boosting ensemble model.

learner is trained, the weight of the weak learner with a small prediction error rate is increased (i.e., the base model that performs well and plays a larger role). The final linear combination of weak learners is formed.

The entire algorithm flow is described in Algorithm 2.

The advantage of the AdaBoost ensemble model is that the base learner can be constructed using various methods. Simple weak learners can be used without filtering the features and without overfitting. In addition, the AdaBoost ensemble model does not need to know the upper limit of the error rate of the weak learners in advance, and the accuracy of the final strong learner obtained depends on the accuracy of all the weak learners, which allows for digging deeper into the ability of the learners. AdaBoost can adjust the assumed error rate adaptively based on the feedback from the weak learners and performs efficiently.

However, it should be noted that boosting is very sensitive to the noise of the training data because the AdaBoost ensemble model focuses its attention on the error samples. If the training data contain a lot of noisy data, then the base learners will all focus on the noisy data for training, which will instead affect the effectiveness of the whole model.

1.2.3. Ensemble Model Construction of LR-Stacking. The model ensemble idea of stacking is completely different from bagging and boosting. Stacking aggregates base models using a model fusion approach, while bagging and boosting use specified strategies [65]. The idea of the stacking framework is to select heterogeneous base learners to be trained on the training set in parallel. All the trained base models are predicted on the prediction set, and the predicted values of all base models are trained as the training set of another fusion model. When a new dataset is an input, it is first predicted by the base model and then input to the fusion model for the final prediction output. For example, we fit a stacking ensemble model consisting of m base learners. First, we divide the training data into the training set and test set, train m base learners on the training set, use the trained base

learners to predict the prediction set to output m predictors, and m predictors are used as the training set of the fusion model to train the fusion model. We usually use K-fold cross-validation for model training in practical applications, and logistic regression is generally chosen for the fusion model.

As shown in Figure 3, based on the idea of stacking, a sub-base model for combining each other base model is trained. This is done by dividing the data into two parts, using one part to train five base models, using the other part of the data to test these base models, and using the output of the five base models as input to train the ensemble model. Instead of organizing the prediction results of the base models, this ensemble of models organizes the models. Theoretically, stacking can organize any model.

The Linear Regression ensemble model is constructed in four steps as follows:

(1) Prepare the training set and test set by dividing the training set into five parts: train1, train2, train3, train4, train5.

(2) Selected base models. Here, the five base models, LSTM, LR, KNN, XGBoost, and GRU, are selected as the base models. For example, in the XGBoost model part: train1, train2, train3, train4, and train5 are used as validation sets in turn, and the remaining four copies are used as training sets for 5-fold cross-validation for model training; then, the prediction is performed on the test set. This will result in five copies of predictions trained by the XGBoost model on the training set and one copy of predictions on the test set. The five copies are overlapped vertically and combined to obtain the training set for the LR model. The same is done for the rest of the base model.

(3) After the five base models are trained, the predicted values of the five models on the training set are used as five "features" for training using the LR model.

(4) Using the trained LR model, the final prediction is made based on the five "feature" values constructed from the previous predictions of the five base models on the test set.
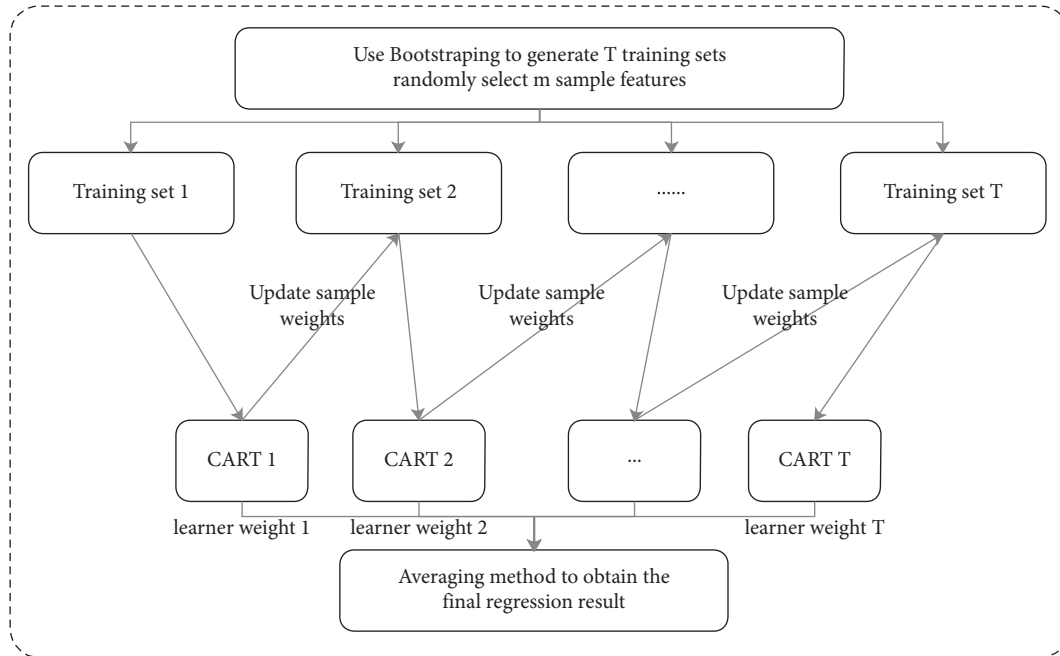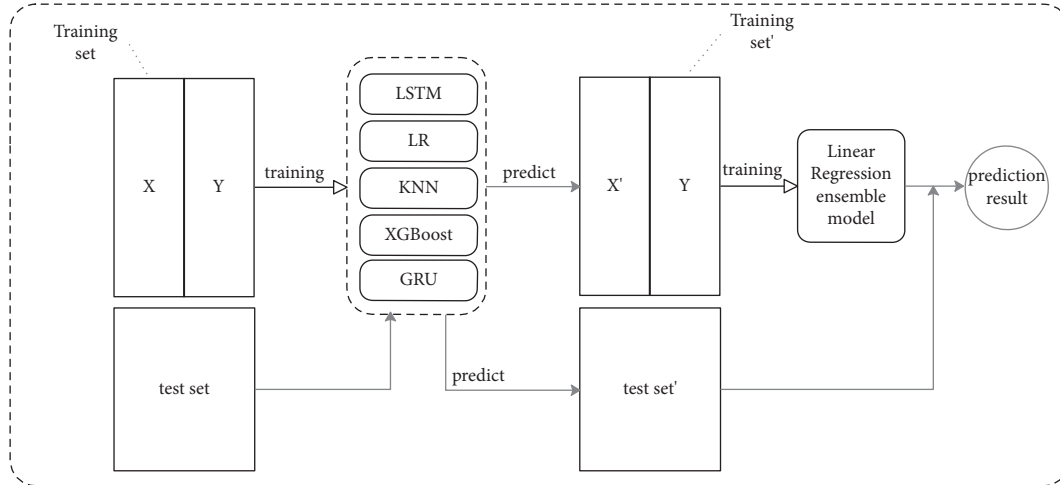
FIGURE 3: Structure of ensemble model based on Linear Regression.

Input: Training set $D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)\}$;
        Sub-mods $\mathscr{L} = \{LSTM, LR, KNN, XGBoost, GRU\}$;
        Ensemble-model $\mathscr{L}' = Linear\ Regression$;
        Number of learning rounds T;
(1)   for t = 1, 2, …, T do
(2) $h_t = \mathscr{L}_t(D)$
(3)   End for
(4) $D' \neq \varnothing$
(5)   for i = 1, 2, …, m do
(6)   for t = 1, 2, …, T do
(7) $z_{it} = h_t(x_i)$
(8)   end for
(9) $D' = D' \cup ((z_{i1}, z_{i2}, \ldots, z_{it}), y_i)$
(10)   end for
(11) $h' = \mathscr{L}'(D')$
        Output H(x) = $h'(h_1(x), h_2(x), \ldots, h_T(x))$
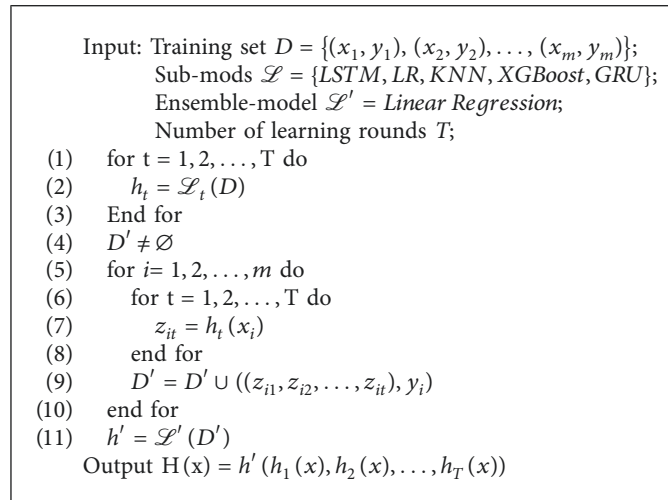
ALGORITHM 3: LR-Stacking ensemble model.

The entire algorithm flow is described in Algorithm 3.

The stacking framework is ensemble with different algorithms to make full use of different algorithms to make different observations on data from the different data space perspectives and data structure perspectives to take advantage of the strengths and weaknesses to optimize the results. Therefore, in addition to the same model with different parameters, the base model can also be of different kinds of base models, so this paper selects LSTM, LR, XGBoost, KNN, and GRU, which have their own characteristics. By aggregating different kinds of base models, we can fully learn the changing law of bus single-trip time, and the ensemble results will be more robust and accurate.

1.2.4. Selection of Sub-models. In this paper, the commonly used bus single-trip time prediction models are divided into types, and their advantages and disadvantages are compared and analyzed. Since the sub-models are selected to be as different as possible in terms of internal principles, the selection of sub-models is carried out in this paper from the broad category of model division. Five sub-models were selected from the broad categories of statistics, deep learning, machine learning, and ensemble learning. They are Multiple Linear Regression (LR), K-Nearest Neighbor (KNN), XGBoost, Long Short-Term Memory Network (LSTM), and Gate Recurrent Unit (GRU), and the internal structure and algorithmic principles of each model are briefly described below.

MLR: although ARIMA is one of the most commonly used time series models, it is more applicable to scenarios with single factor inputs [66]. The model principle is simple, as shown in Figure 4. The error between the predicted and true values is calculated, and a line (plane or hyperplane) is finally fitted by continuously seeking the optimal solution of the parameters utilizing gradient descent. The linear regression model can capture well the linear patterns present in the data but is insensitive to nonlinear patterns. MLR is very suitable when the dependent variable is affected by two or more features, and there is a linear relationship between
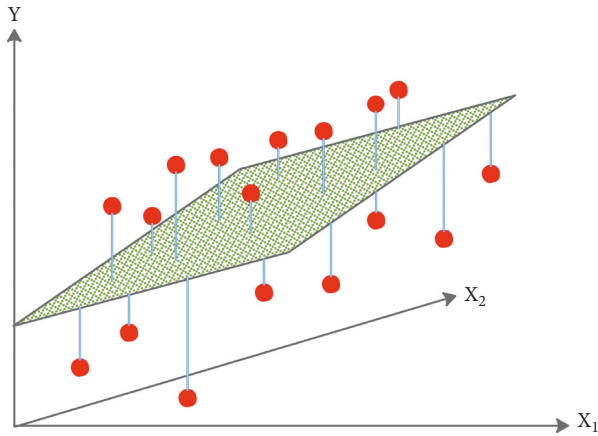
FIGURE 4: MLR schematic diagram.

FIGURE 5: KNN algorithm principle.

FIGURE 6: XGBoost principle.

multiple independent variables and the dependent variable. In this study, holidays, rainfall, visibility, and temperature are the data characteristics that affect the bus single-trip time, and they are independent and continuous, so multiple linear regression is selected as the sub-model.

KNN: KNN is a distance-based model, and the principle of its classification model algorithm is given in Figure 5. Based on the K-value, the nearby elements are selected and grouped into the class with the highest number of nearby element categories, the most similar class. The regression model is similar, based on the K-value size to select the nearby data and further processing to obtain the predicted value, which is usually processed by directly finding the mean value [67]. This study chooses KNN as the sub-model for the following two reasons. First of all, in terms of model, the principle and implementation of KNN are relatively simple. KNN does not need to estimate parameters and is suitable for dealing with regression problems. Secondly, in terms of data, KNN is not sensitive to the outliers of the data, and has good Lupin performance for the nonlinear data with large noise value such as the bus single-trip time.

XGBoost: XGBoost is a boosted tree model based on an ensemble learning boosting approach to regression tree as the base model [68]. It has an objective function. The objective function depends only on the first- and second-order derivatives of the data in the error function. In terms of implementation, it can be parallelized (sparse-aware algorithm is proposed) to speed up the training speed; a regular term is added to the objective function, which controls the complexity of the model and helps prevent overfitting; XGBoost supports columns sampling, can reduce overfitting, and reduce computation; it can handle missing values. Based on these advantages, this study selects XGBoost as a sub-model. The principle of the algorithm is shown in Figure 6.

LSTM: long short-term memory (LSTM) is a special RNN, mainly to solve the problem of gradient disappearance and gradient explosion during long sequence training. Simply put, LSTM can perform better in longer sequences than ordinary RNNs. As a deep learning model, CNN is not completely suitable for learning time series, so various auxiliary processing is required, and the effect is not necessarily good. For problems that are sensitive to time series, LSTMs are usually more suitable. LSTM is an excellent variant model of RNN. It inherits the characteristics of most RNN models and solves the vanishing gradient problem caused by the gradual reduction of the gradient back-propagation process. Therefore, LSTM is very suitable for dealing with time series [69]. So, this study chooses LSTM as one of the sub-models. The principle of LSTM is shown in Figure 7.
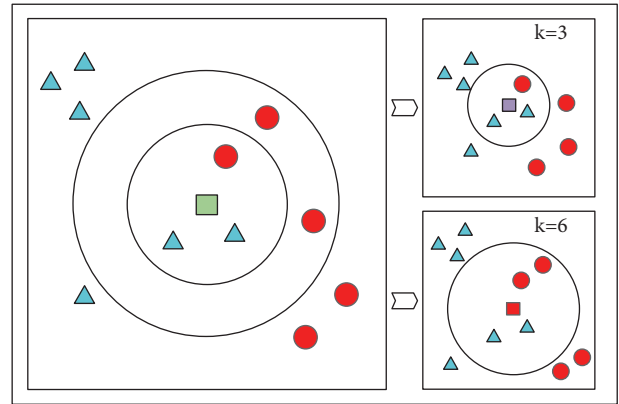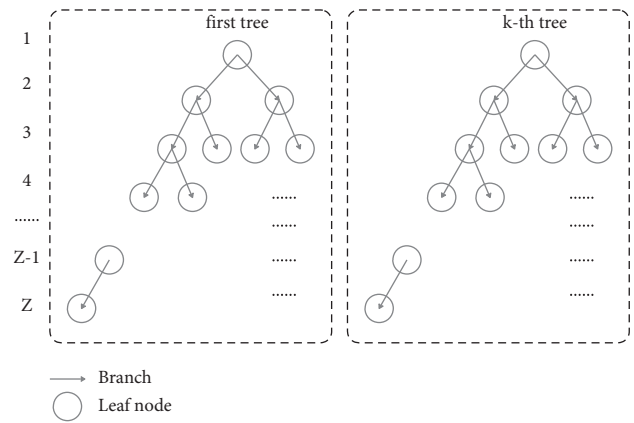
GRU: GRU is a very effective variant of the LSTM network, which is simpler and more effective than the structure of the LSTM network, so it is also a very streamlined network at present [70]. Three gate functions are introduced in LSTM: input gate, forget gate, and output gate to control the input value, memory value, and output value [71]. Moreover, there are only two gates in the GRU model: update gate and reset gate. The number of parameters of LSTM is 4 times that of RNN. If the number of parameters is too large, there is a risk of overfitting. GRU only uses two gated switches, which achieves results close to LSTM. In order to avoid overfitting of LSTM, GRU was selected as one of the sub-models for comparison in this study. The specific structure is shown in Figure 7.
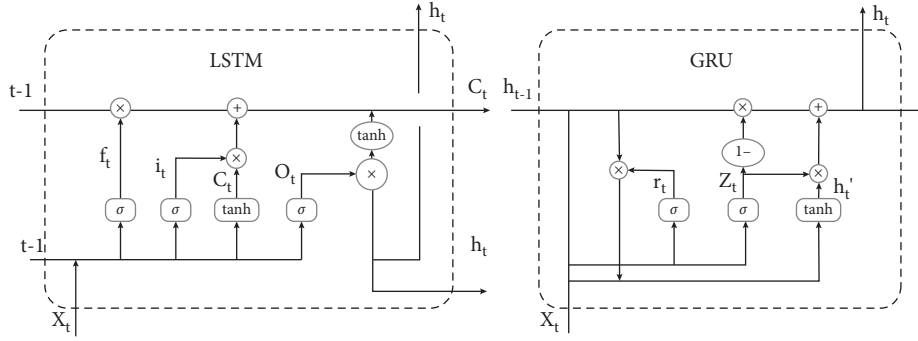
Figure 7: Comparison of the internal structure of LSTM and GRU.
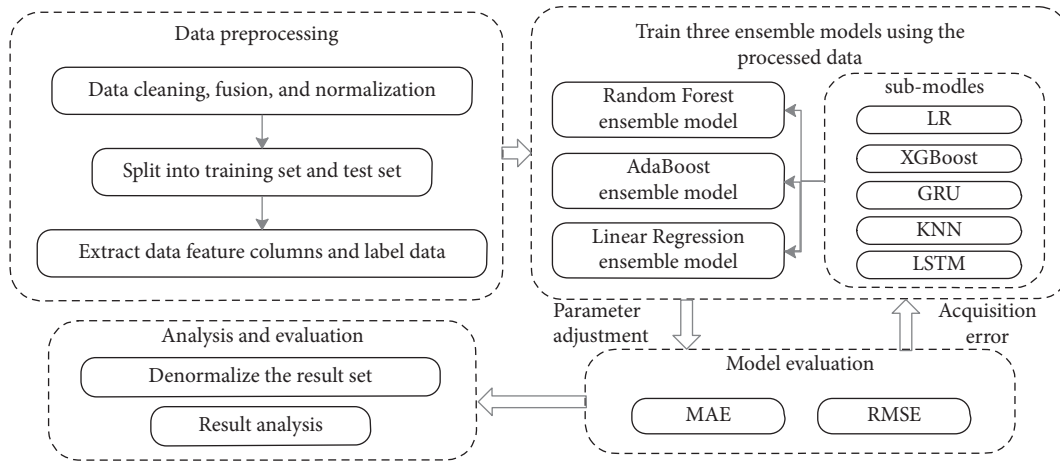
Figure 8: Prediction steps.

1.2.5. Model Construction and Prediction Steps. The specific prediction steps are shown in Figure 8. Firstly, the fused processed data are loaded into the dataframe, and the data are split in the ratio of 6 : 1 : 1 before feeding them into the model and divided into the training data set, the test data set, and the prediction data set.

Feature selection of the data is performed afterward to select the key features used to predict the target data. In this study, feature selection is performed using the method of removing low variance features to ensure that the complexity associated with redundant features is reduced while maximizing feature relevance. This step identifies the most meaningful subset of features by removing low variance features, including date, bus route up and down, whether a holiday, temperature, visibility, and precipitation.

Next, the organized training data set is input to each of the five sub-models for model training, and the model is trained to find the most suitable weights and optimal model parameters to minimize the error between predicted and actual values. The parameters in each algorithm are determined by the grid search method before using them for prediction.

The training dataset is fed into the ensemble model consisting of sub-models that have been tuned with hyperparameters to train the ensemble model separately, and the parameters of the ensemble model are tuned.

For RF-Bagging, firstly, a fixed number of samples are collected from the original training set by random sampling (bootstrap), but after each collection, the collected samples will be put back. Randomly collecting the same number of samples as the number of training samples m can make the number of samples in the sampling set and training set the same, but the sample content is different. Secondly, based on the gradient boosting tree, RF improves the establishment of the decision tree. We will select an optimal feature from all the n sample features on the node to divide the left and right subtrees of the decision tree. Finally, because we are studying the regression problem, we arithmetically average the obtained regression results to obtain the final model output.

For AdaBoost-Boosting, the training set of each round is unchanged, but the weight of each example in the training set in the classifier changes. The weights are adjusted according to the classification results of the previous round. The weights of the samples are continuously adjusted according to the error rate. The larger the error rate, the greater the weight. Individual prediction functions can only be generated sequentially, because the results of the previous model round are required for the latter model parameters.

For LR-Stacking, this stacked ensemble model is relatively simple, and the prediction results of the five sub-models are used as the data of the secondary learner, that is, as the training set of the LR model. The prediction result of

the trained LR model is the prediction result of the ensemble model.

Model validation is performed using a test dataset after the model is trained. The same parameters used for training the model are used for data validation, and the same feature values are selected on the estimated validation data used during model training. This step aims to verify the prediction accuracy of the model and minimize the output error of the validation data. Model parameters are tuned based on training and validation results to find the parameters that apply to the test data based on the features of different sub-models.

Finally, the prediction accuracy of the ensemble model is then validated using test data.

2. Case Analysis

2.1. Description of Usage Data. The data used in this study are the bus single-trip time data of Beijing 2 from April 1, 2020, to August 31, 2020, under the normal scenario, with a total of 3512600 entries. As shown in Figure 9, it can be seen from the figure that the data have obvious seasonal characteristics of time series. The unit of the *y*-axis in the figure is minutes.

Holiday and Beijing weather data in the same period as the bus single-trip time data are 1210 items with a time granularity of 3 hours. The temperature unit is Celsius, the visibility unit is a kilometer, and the precipitation unit is millimeters. The data analysis graph is shown in Figure 10.

As shown in Figure 11 and Figure 12, the bus single-trip time running time and weather data are combined to see that the data of bus single-trip time are mainly concentrated in the interval of 4 0 minutes to 60 minutes. The values of temperature are mainly distributed between 20 and 30. The distribution of visibility and rainfall is more scattered, and the values change more randomly.

According to the two-dimensional kernel density analysis plot in Figure 13, the kernel density of bus single-trip time and visibility has a bipartite distribution, while the kernel density and temperature have a single kernel distribution.

After fusing bus single-trip time data, weather data (precipitation, visibility, temperature), and holiday data, the training dataset, test dataset, and prediction dataset are set up in the ratio of 6 : 1 : 1. Data pre-processing is done using python language, and data noise reduction is done using scipy's own filter. In this study, the one-way operation data of Beijing Bus No. 2 are used as an example for prediction.

2.1.1. Model Evaluation Metrics. The root-mean-squared error (RMSE) and the mean absolute error (MAE) are the two most frequently used metrics to measure the accuracy of variables, and they are also two important yardsticks to evaluate models in machine learning. Therefore, these two indicators are selected to compare and analyze the prediction effects of different models in this paper. These two indicators mainly reflect the magnitude of error between the predicted and actual values, defined in Eq.

FIGURE 9: Bus single-trip time data trend graph (daily, weekly, monthly).

FIGURE 10: Weather data trend graph.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum \left(y^{true} - y^{pred}\right)^2}$$

$$\text{MAE} = \sqrt{\frac{1}{n} \sum \left|y^{true} - y^{pred}\right|}. \tag{3}$$

FIGURE 11: Data distribution.

FIGURE 12: Weather data box plot.

In the base model LR, fit_intercept is the default True. Since the dataset has been normalized, nomalize is false, and copy_X is set to True to avoid overwriting the original data. Set n_jobs to -1 to improve the operational efficiency. In the base model XGBoost, based on the amount of data, the number of iterations is set to 200, n_estimators = 200, and learning_rate = 0.4. The maximum depth of each tree Max_depth is 6, min_child_weight = 1, gamma is the default 0. In sampling, subsample = 1, because there are a few features, bytree, bylevel, and bynode are all 1. Because XGBoost has an advantage in training speed, so tree_method choose exact with higher precision. In the deep learning-based models LSTM and GRU, in order to prevent under-fitting or over-fitting, the grid search method is used to conduct multiple experiments to find the optimal epoch and batch_size. In LSTM, epoch = 200, batch_size = 16, neuron units = 4, dropout = 0.2. In GRU epoch = 150, batch_size = 4, neuron units = 4, dropout = 0.2. In KNN, since the distance weight needs to be considered, weight = 'distance', and the value of k is 4 according to the size of the data.

2.1.3. Comparison of Prediction Results. In this study, two feature sets containing weather conditions and holidays and only weather conditions were selected for model training. The granularity of bus single-trip time used in this study is hour. The original data counts the bus single-trip time from 0:00 to 12:00 per hour, while the time granularity of weather data is 4 hours. In the case that time-step is one hour, this study adopts 4 and 6 as the experimental time-step, which can fully study the influence of weather on the bus single-trip time, and can also fully reflect the periodic regularity of the bus single-trip time. The experimental results obtained in these cases can compare and analyze the performance and generalization ability of the model from different perspectives.

(1) Model evaluation considering weather conditions and holiday scenarios. The comparison of the predicted and true values of the three ensemble models at step sizes of 4 and 6, respectively, is shown in Figures 14 and 15. It can be found that the prediction results of the ensemble model based on Random Forest are better than those of the other two ensemble models in both step sizes, and the fit is better and closer to the true value. The ensemble model based on AdaBoost has the second-best fit, and the prediction result of the ensemble model based on Linear Regression has the worst fit among the three ensemble models. In general, all three ensemble models can learn the fluctuation pattern of the real data, and the prediction result of AdaBoost has the least fluctuation compared with the other two ensemble models. It is not sensitive to the response of data peaks and troughs.

From the evaluation indices of the ensemble model in Table 2, both mae and rmse of RF-Bagging are the smallest, indicating that it has the smallest error, while R^2 in the case of step size 4 is 0.909and in the case of step size 6 is 0.883; both are greater than 0.8, indicating that this model can learn

In the above equation, the y^{true} denotes the actual value, the $y^{pre\,d}$ MAE reflects the error between the model prediction and the real value, and RMSE is more sensitive to the outliers, reflecting the model's stability. The smaller the RMSE and MAE values of the prediction results, the closer the predicted value and the actual value are, and the higher the prediction accuracy of the model [72]. The smaller the RMSE and MAE values are, the closer the predicted and actual values are, and the higher the prediction accuracy of the model.

2.1.2. Experimental Parameter Settings. All algorithms are implemented using Python 3.7 running on a computer with a quad-core 2.6-GHz CPU and 16 GB random-access memory.

In terms of hyperparameter settings, based on factors such as error size, training efficiency, and degree of fit, a combination of grid search and cross-validation is used, and two timesteps of 4 and 6 are selected for training and prediction.

FIGURE 13: KDE plots (single-trip time-visibility, single-trip time-temperature).
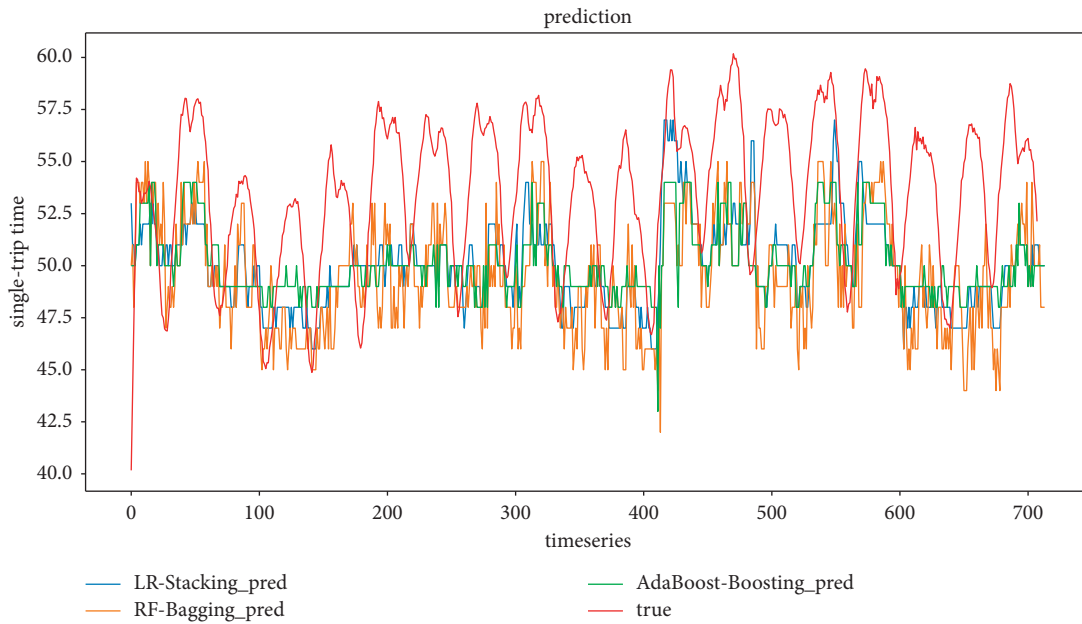
FIGURE 14: Ensemble model prediction results (timestep = 4).

the advantages of each sub-model and avoid its shortcomings, making the overall error lower and the fit higher, which is an excellent ensemble prediction model.

Figures 16 and 17 show the predicted and actual values of the bus single-trip time derived from the training of five single models and three ensemble models. It can be seen that the single models, LSTM, LR, GRU, KNN, and XGBoost, all have inferior prediction results than the ensemble models. Among the five sub-models, the prediction results of LSTM have the largest deviation, and the prediction results of LR and GRU are the closest to the real values. This is mainly due to the dynamic nature of neural networks and the advantage

of ensemble deep learning. In terms of model training time, LR, KNN, and XGBoost are much faster than LSTM and GRU and are more suitable for the short-time prediction of large data sets.

As shown in Table 3, among the sub-models, the training time of the LSTM model and the GRU model is quite different, mainly because LSTM and GRU require 200 and 150 rounds of learning, respectively, to achieve convergence. Among the five sub-models, LR completed the training and prediction in the shortest time and achieved the smallest error. In the ensemble model, RF-Bagging and LR-Stacking take much less time than AdaBoost-Boosting. On the whole,

FIGURE 15: Ensemble model prediction results (timestep = 6).

TABLE 2: Model evaluation indices.

Ensemble models	TimeStep = 4			TimeStep = 6		
	MAE	RMSE	R^2	MAE	RMSE	R^2
LR-stacking	4.317347	5.236942	0.155475	4.062111	4.808697	0.196314
RF-bagging	3.232055	4.313371	0.909258	3.895147	4.671306	0.883880
AdaBoost-boosting	4.086188	5.040191	0.489591	3.973029	4.724360	0.404368

RF-Bagging obtains the most accurate prediction results in a shorter time.

From the evaluation indices of all models in Table 4, all three ensemble models are smaller than the sub-models in terms of error, which can fully illustrate the ensemble advantages of the ensemble models. In contrast, among the five sub-models, except for the LSTM with larger error, the errors of the remaining four models are relatively close, among which LR and GRU are better than KNN and XGBoost in terms of error. It is noteworthy that in the case of step size 4, the prediction bias of LSTM is larger in the case of timestep size 6, while the error decreases in the case of step size 6, indicating that the adjustment of step size optimizes the prediction accuracy of LSTM.

Table 5 shows the results of model testing. To verify the superiority of RF-Bagging, Wilcoxon signed-rank test and Friedman test were performed on all models in this study, and the test results are shown in Table 5. In the Wilcoxon signed-rank test, the rank of the absolute value of the difference between the observed value and the center position of the null hypothesis is added according to different signs as its test statistic. It is suitable for pairwise comparisons in T-tests, but does not require that the differences in paired data follow a normal distribution. This test only requires a symmetrical distribution, so it is more suitable for the comparative test of the predicted value of the bus single-trip

time. Therefore, in order to verify the superiority of RF-Bagging, this study tested the predicted values of RF-Bagging with those of other seven models. Friedman test can take full advantage of all the information in the relevant sample. The prerequisites for using the Friedman test are (1) ordinal-level data, (2) three or more groups of data, and (3) randomly draw samples from the collocated values. Therefore, the Friedman test is also applicable to this study. From the test results, whether it is the Wilcoxon signed-rank test or the Friedman test, the ρ obtained is less than the significance level (0.05). This shows that RF-Bagging shows better prediction performance than other models when considering weather and holidays.

In summary, it can be seen that the three ensemble models have excellent prediction accuracy at both step sizes considering weather and holidays, and the ensemble model has a better fit than the sub-models. The Random Forest-based ensemble model constructed based on the Bagging ensemble idea has the best fit and prediction accuracy among the three ensemble models. This reflects the advantages of this model ensemble method: due to the use of random sampling, the variance of the trained model is small, and the generalization ability is strong; compared with the traditional decision tree model, it combines the results of multiple decision tree models, and the model ground effect will be better; in the case of large data fluctuations, the

FIGURE 16: All model prediction results (timestep = 4).
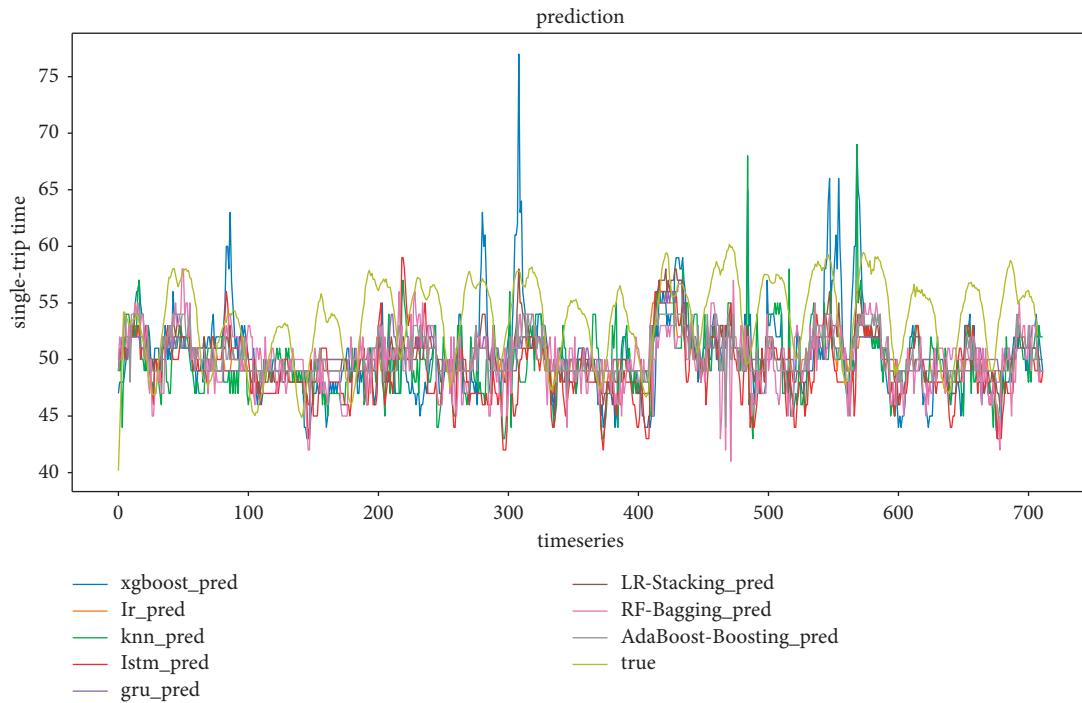
FIGURE 17: All model prediction results (timestep = 6).

ensemble model based on Random Forest can better integrate the advantages of each sub-model and the prediction accuracy higher prediction accuracy. GRU and Linear Regression have higher approximation accuracy and generalization ability among the five sub-models, and the prediction results are closer to the real values than other sub-models.

(2) Model evaluation considering only weather conditions scenarios. The comparison between the predicted and true values of the three ensemble models for the scenarios considering only rainfall, visibility, and temperature with step sizes of 4 and 6, respectively, is shown in Figures 18 and 19. It can be found that, similar to the prediction results of

TABLE 3: Model training and prediction time.

Models	TimeStep = 4		TimeStep = 6	
	Training	Prediction (ms)	Training	Prediction (ms)
LSTM	115 min	1257	121 min	1471
GRU	119 min	727	126 min	1103
LR	7 ms	2	16 ms	3
KNN	77 ms	63	98 ms	82
XGBoost	3619 ms	48	5192 ms	73
LR-stacking	58 ms	11	52 ms	56
RF-bagging	329 ms	79	417 ms	136
AdaBoost-boosting	2351 ms	92	3162 ms	141

TABLE 4: Model evaluation indices.

Models	TimeStep = 4		TimeStep = 6	
	MAE	RMSE	MAE	RMSE
LSTM	11.130382	11.977166	5.008087	5.851334
GRU	4.696761	5.694506	4.307110	5.269999
LR	4.324053	5.275761	4.038138	4.842730
KNN	4.931067	6.001118	4.690769	5.538258
XGBoost	5.112036	6.396894	4.604660	5.554114
LR-stacking	4.317347	5.236942	4.062111	4.808697
RF-bagging	3.232055	4.313371	3.895147	4.671306
AdaBoost-boosting	4.086188	5.040191	3.973029	4.724360

TABLE 5: Wilcoxon Single-Rank and Friedman test result.

Models	TimeStep = 4		TimeStep = 6	
	Wilcoxon single-rank test $(\rho \leq 0.05)$	Friedman test $(\rho \leq 0.05)$	Wilcoxon single-rank test $(\rho \leq 0.05)$	Friedman test $(\rho \leq 0.05)$
LSTM	0.00004		0.00251	
GRU	0.00891		0.00830	
LR	0.00916		0.00962	
KNN	0.00612	0.000298	0.00973	0.000463
XGBoost	0.00315		0.00988	
LR-stacking	0.00979		0.01374	
AdaBoost-boosting	0.01693		0.01421	
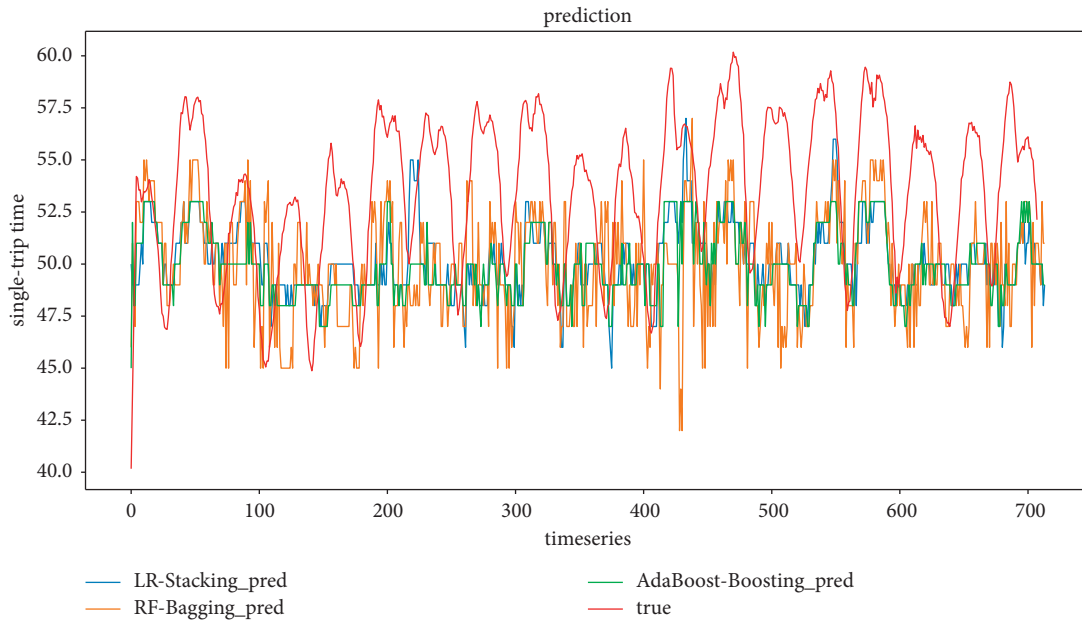
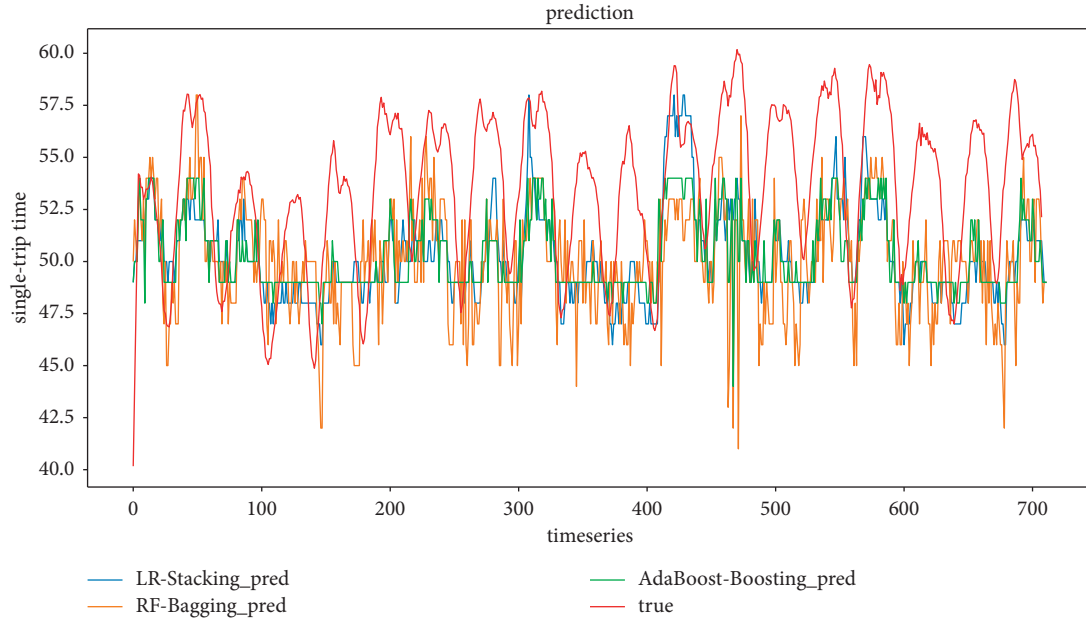FIGURE 18: Ensemble model prediction results (timestep = 4).
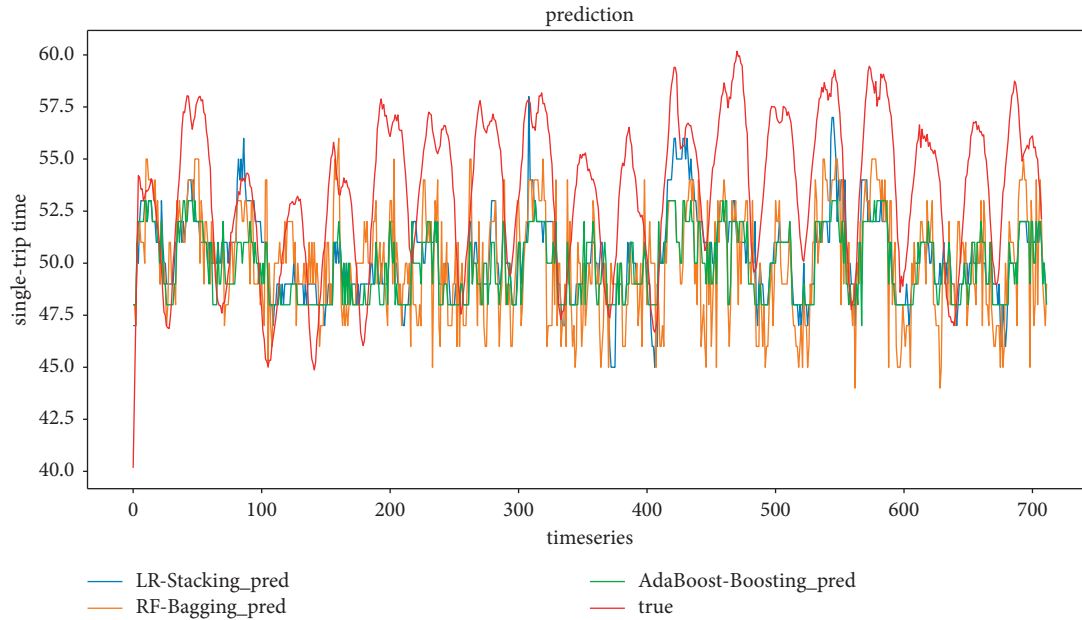
FIGURE 19: Ensemble model prediction results (timestep = 6).

TABLE 6: Model evaluation indices.

Ensemble model	TimeStep = 4			TimeStep = 6		
	MAE	RMSE	R^2	MAE	RMSE	R^2
LR-stacking	4.129265	5.092021	0.123339	4.098628	4.870725	0.205406
RF-bagging	4.146337	5.069491	0.836024	3.935297	4.683990	0.846075
AdaBoost-boosting	4.347092	5.273396	0.412592	4.139669	4.920567	0.390879

the scenarios considering holidays and weather, the prediction results of the ensemble model constructed based on Random Forest at two-step sizes still outperform those of the other two ensemble models, which are closer to the true values. The prediction result of the ensemble model based on AdaBoost is the second best, and the prediction result of the ensemble model based on Linear Regression is the worst fit among the three ensemble models.

From the evaluation indices of the ensemble model in Table 6, the mae and rmse of RF-bagging are basically the smallest, but the error is not large with the other two ensemble models in this scenario, while R^2 In the case of step size 4 is 0.836 and in step size 6 is 0.846; both are greater than 0.8. This indicates that this model can also learn the advantages of each sub-model under the condition of considering only the weather, which makes the overall error lower and the fit higher, and is a better prediction model in comparison.

Figures 20 and 21 show the predicted and actual values of bus single-trip time derived from the training of five single models and three ensemble models. It can be seen that LSTM, LR, GRU, KNN, and XGBoost are inferior to the ensemble models. It is worth noting that the prediction results of the XGBoost model in this context fluctuate more, indicating that the fit of XGBoost is not good in the case of

feature reduction. In terms of model training time, LR, KNN, and XGBoost are also trained much faster than LSTM and GRU, which are more suitable for short-time prediction of large datasets.

As shown in Table 7, the training time and prediction time of the model are overall longer than when the weather and holidays are considered, mainly because of the increase in data features. As far as the base model is concerned, LR is still the model with the shortest training and prediction time. And the model with the smallest error becomes LSTM, which reflects the advantages of deep learning in the case of increasing data features. In terms of ensemble models, the training and prediction times of the three ensemble models are not significantly different.

As shown in Table 8, from the evaluation indices of all models, all three ensemble models are smaller than the sub-models in terms of error, fully illustrating that the ensemble models have the same advantages of model ensemble when only weather conditions are considered. Compared with the scenarios considering weather and holidays, among the five sub-models, the prediction error of LSTM at step size 4 is the smallest, while the prediction results of LR and LSTM are the closest to the true value, and the prediction results of GRU and LR at step size 6 are the closest to the true value. Combining the prediction results of the two scenarios
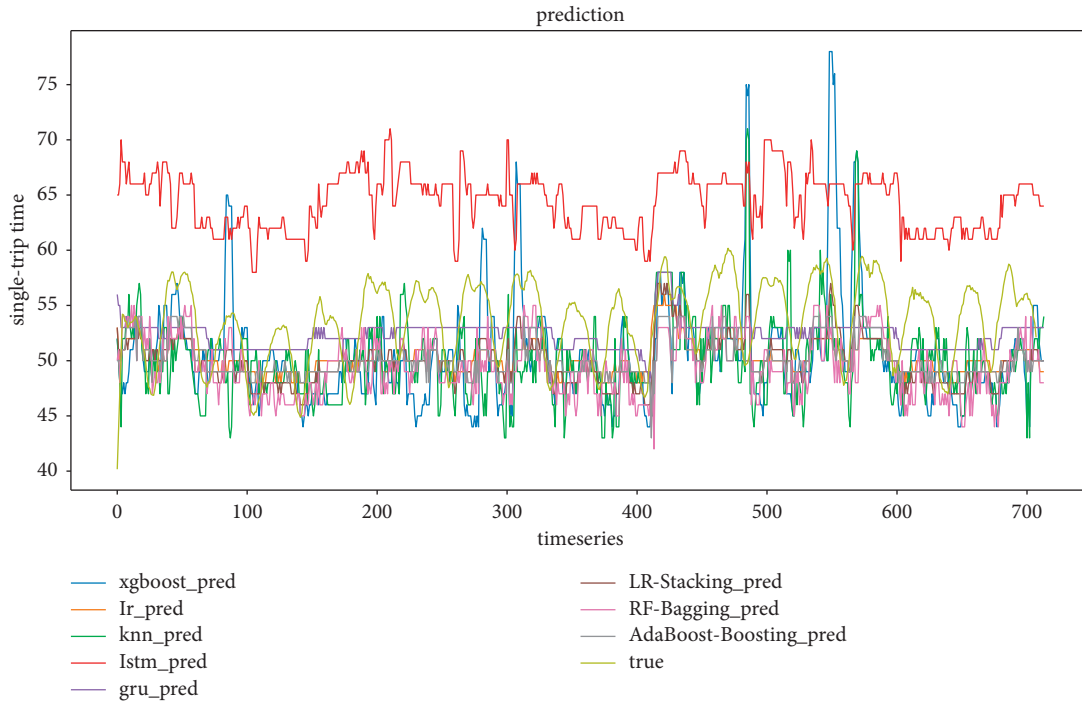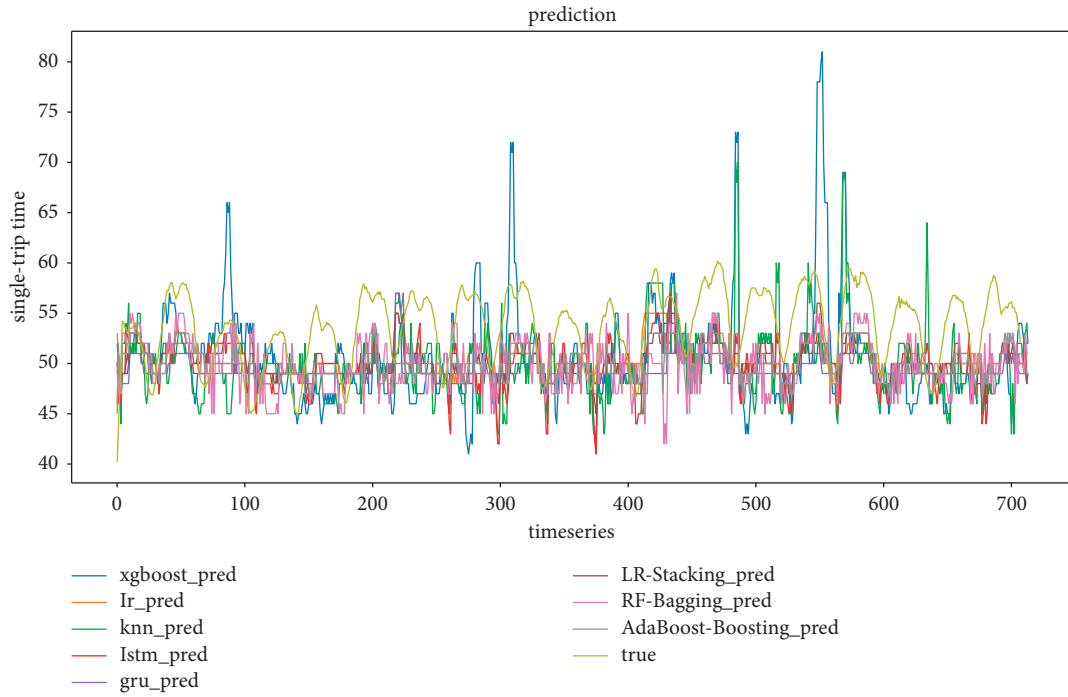
FIGURE 20: All model prediction results (timestep = 4).
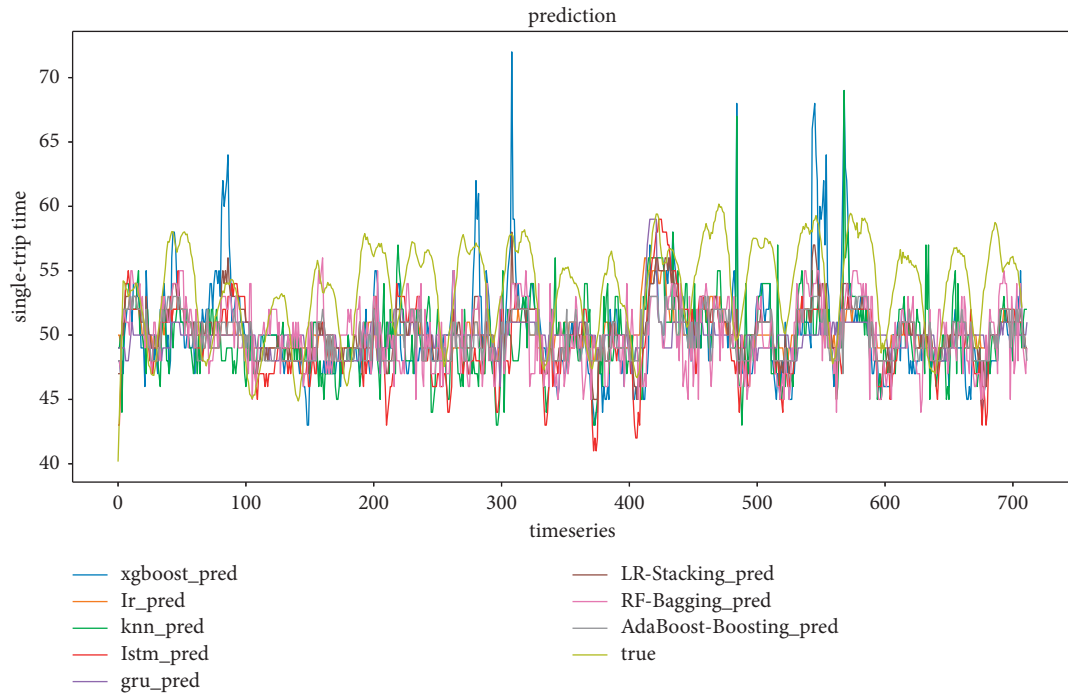
FIGURE 21: All model prediction results (timestep = 6).

illustrates that LR as a single prediction model has a more stable and high prediction accuracy when predicting bus single-trip time.

As shown in Table 9, the Wilcoxon signed-rank test and Friedman test results for models when only the weather is considered are shown in Table 8. It can be found that RF-

Bagging also maintains its superiority in this scenario, and the ρ in the test results are all less than the significance level (0.05).

In conclusion, it can be seen that the three ensemble models have excellent prediction accuracy at both step sizes when only weather is considered, and the ensemble models

TABLE 7: Model training and prediction time.

Models	TimeStep = 4		TimeStep = 6	
	Training	Prediction (ms)	Training	Prediction (ms)
LSTM	85 min	986	106 min	1326
GRU	93 min	645	114 min	735
LR	4 ms	1	6 ms	2
KNN	64 ms	52	71 ms	68
XGBoost	2639 ms	32	3283 ms	42
LR-stacking	21 ms	6	35 ms	21
RF-bagging	214 ms	51	257 ms	82
AdaBoost-boosting	1427 ms	72	1974 ms	102

TABLE 8: Model evaluation indices.

Models	TimeStep = 4		TimeStep = 6	
	MAE	RMSE	MAE	RMSE
LSTM	4.306787	5.199513	4.481952	5.300563
GRU	4.861043	5.786797	4.392949	5.170590
LR	4.539661	5.596035	4.361890	5.276523
KNN	4.954903	6.027883	4.684654	5.524954
XGBoost	5.214021	6.529260	4.698066	5.579520
LR-stacking	4.129265	5.092021	4.098628	4.870725
RF-bagging	4.146337	5.069491	3.935297	4.683990
AdaBoost-boosting	4.347092	5.273396	4.139669	4.920567

TABLE 9: Wilcoxon Single-Rank and Friedman test results.

Models	TimeStep = 4		TimeStep = 6	
	Wilcoxon single-rank test ($\rho \leq 0.05$)	Friedman test ($\rho \leq 0.05$)	Wilcoxon single-rank test ($\rho \leq 0.05$)	Friedman test ($\rho \leq 0.05$)
LSTM	0.00942		0.00887	
GRU	0.00725		0.00891	
LR	0.00932		0.00736	
KNN	0.00462	0.000327	0.00623	0.000409
XGBoost	0.00152		0.00693	
LR-stacking	0.00979		0.01729	
AdaBoost-boosting	0.00957		0.00932	

have a better fit than the sub-models. In comparing the prediction results of the ensemble models, the ensemble model constructed based on Random Forest has higher prediction accuracy and fit, which further illustrates the applicability of the parallelization approach based on the idea of bagging ensemble in bus single-trip time prediction. By fitting different learners individually and training them simultaneously, an ensemble model is generated that is more robust than a single model. Meanwhile, Random Forest supports multiple tree ensemble, which can form a powerful heterogeneous ensemble algorithm to randomly select samples and features, reducing the effect of outliers and reducing overfitting.

3. Discussion

In this study, five single predictive models and three ensemble models are validated with real-world data in order to find the best method for bus single-trip time prediction. And the sensitivity of the prediction accuracy of each model to the number of features is also verified by adding an extra feature with data of holidays. The optimal model for predicting the bus single-trip time is evaluated by the two error evaluation indices of MAE and RMSE combined with the training and prediction efficiency of the model. In terms of a single prediction model, this study selected a linear regression model from the traditional statistical category. The simplest and easy-to-implement KNN and XGBoost that can effectively prevent overfitting are selected from the machine learning category. LSTM that can learn long-term dependency information and GRU that are more efficient are selected from the deep learning category. In terms of the ensemble method, this research selects RF-bagging, Ada-Boosting, and LR-stacking from the three popular ensemble categories including bagging, boosting, and stacking. The purpose of this study is to compare the prediction accuracy, error, and efficiency of the ensemble predictive model and

the single predictive model under different characteristics and timesteps and to find an optimal predictive model suitable for the bus single-trip time dataset.

In the case of verification part, the parameters of the model are adjusted by continuous trial and error, so that each model achieves a better balance in terms of prediction accuracy and training efficiency. Finally, the comparison between the predicted value and the real value of each model is obtained, and the MAE and RMSE of each model are calculated. Based on the result comparison between ensemble models and single prediction models, whether or not the holiday data are considered, the prediction accuracy of the ensemble models is higher than that of the single predictive models. This is mainly due to the fact that ensemble learning combines multiple base learners to obtain superior generalization ability compared to a single learner. This shows that the idea of ensemble learning is suitable for the prediction of bus single-trip time. For nonlinear time series data ensemble models that combine multiple weather conditions and holidays, the advantages of basic learners can be well combined. This experimental result also well confirms the current academic description of the advantages of ensemble learning.

Looking into the prediction results of the three ensemble models selected in this study, when there are many data features, the prediction error of the ensemble model is lower. In both cases, the prediction error of the RF-Bagging model constructed based on the bagging ensemble method is the lowest. This reflects the difference between ensemble methods of bagging and boosting: bagging focuses on reducing the variance of the model (preventing overfitting), while boosting focuses on reducing the skewness of the model (preventing underfitting). The nonlinear characteristics of the bus single-trip time data set used in this paper are relatively obvious since Beijing public transport has been affected by the COVID-19 pandemic and various social activities during the time period in which the data set is located. The ensemble idea of bagging is more suitable for such noisy datasets. While RF-bagging is based on a decision tree model which introduces random samples and attribute selection based on bagging ensemble. It makes RF-bagging less prone to overfitting in case of using high-noise datasets. The LR-Stacking constructed by stacking is dependent on basic learners. Once too many basic learners are unsuitable for processing noisy datasets, the accuracy of LR-Stacking will decrease.

On the other hand, from the prediction results of the five single predictive models, LR has the highest accuracy, while KNN and XGBoost of machine learning and LSTM and GRU of deep learning have comparable performance. This is also probably due to the dataset's high noise and nonlinear characteristics. Relatively complex machine learning and deep learning models are more seriously affected by data noise, and the degree of overfitting is higher than that of simpler LR.

From the perspective of algorithm complexity, among sub-models, LSTM and GRU, which belong to the deep learning category, are the most complex. In this study, LSTM and GRU have six layers of hidden nodes. The complexity of

the two data dimensions is relatively high. They occupy most of the system resources and consume a long time for model training. However, the training and prediction time of other models is in the millisecond level, and the resource consumption is not high. Among them, LR has the fastest resource occupation and training speed. This is mainly due to the fact that LR only needs to store the eigenvalues of each dimension, so relatively speaking, the resource occupancy is small, and the calculation amount is only the number of features. Relevant, this study has fewer features, so the calculation speed is relatively fast. In ensemble models, LR-Stacking, which also uses LR as a secondary base learner, has the fastest resource occupancy and training speed. In the process of each iteration, Ada-Boosting gradually approaches the expected value from the two aspects of the detection rate and the misrecognition rate of positive samples to construct a cascaded classifier, which can only be achieved after iterative training generates a large number of weak classifiers, construction process. From this, it takes more time to train the classifier with a circular approximation, so its complexity is the highest among the three ensemble models.

The theoretical significance of this study is to verify the effectiveness of ensemble learning in the prediction of bus single-trip time. In ensemble learning, the RF-bagging model constructed by bagging is the most suitable for predicting the bus single-trip time. In addition, the RF-bagging model is versatile when dealing with nonlinear and noisy datasets, which can effectively prevent model overfitting. The practical significance of this study is that when the influence of many external factors makes the bus single-trip time become irregular, the ensemble model proposed in this research can provide the public transport managers and passengers with accurate bus single-trip time predictions. It provides convenience for passengers' travel and also provides a basis for managers to assist in decision-making. Based on the characteristics of the datasets used in this research, in future, more solutions will be proposed for excessive data noise.

4. Conclusion

Predicting traffic demand is a central issue in the organization of any transportation system, and the predictive demand could help to plan a reasonable supply in advance. From the perspective of public transportation, the distribution of bus single-trip time is needed in real time for travel planning, operation strategy formulation and adjustment, and contingency planning. This paper proposes a methodology for constructing a multi-model ensemble bus single-trip time prediction model based on the public transportation data, holiday data, and external weather data. The empirical analysis is conducted by using the proposed methodology and comparatively predict bus single-trip time based on the single models and the ensemble models. The specific research work done and research results obtained in this paper are mainly as follows:

(1) The bus single-trip time data, holiday data, and external weather data are cleaned separately, including

data redundancy, data gap filling, and abnormal data processing and noise reduction. Moreover, the cleaned datasets are fused with multiple sources to provide data support for the further bus single-trip time prediction model building.

(2) A data-driven bus single-trip time prediction framework is constructed, including three steps of data analysis, feature extraction, and prediction modeling.

(3) Three ensemble models of bagging (Random Forest), boosting (AdaBoost), and stacking (Linear Regression) were constructed for predicting bus single-trip time based on three-model ensemble methods.

(4) A case analysis was conducted using real data of Beijing bus Line No. 2. The advantages and disadvantages of the five base models and the three ensemble models were compared and analyzed. The ensemble model for bus single-trip time prediction is constructed, and the constructed ensemble model is used to make short-time predictions of bus single-trip time under normalization. The real values are used as the baseline for detailed comparison with the prediction results of the constructed model from the perspective of the single model and the ensemble model. The results of case analysis show the following: (1) in general, the prediction results of the ensemble model are commonly better than those of the sub-models, regardless of whether the nonlinear time-series data are volatile or regular, which reflects the strong benefits of ensemble learning; (2) the ensemble model of Random Forest built by the method of bagging ensemble the advantages and disadvantages of the five sub-models. The overall prediction results are smoother and better than those of the five sub-models and is closer to the real value. Since the ensemble model fully learns the laws between the independent variables and the prediction results of the sub-models instead of simply integrating the prediction results directly. The overall prediction effect is better, which brings out the optimal solution for the prediction model in each scenario. (3) Among the single predictive model, LR is the best model with high prediction accuracy but with not high computational cost that is also easy to implement. It can be applied to distributed data and handle large data with fewer resources. In addition, LR is robust to a small noise in the data and does not suffer from slight multicollinearity, making it an optimal solution for a single prediction model.

Further research on the bus single-trip time prediction problem can be done in the following two areas:

(1) The selection of sub-models and the number of selecting rounds mainly rely on historical experience, and the subsequent optimization algorithm can be considered to make the selection more intelligent and reasonable.

(2) This paper only predicts the bus single-trip time from a theoretical point of view and provides data support for the subsequent development of emergency strategies. The subsequent development of emergency plans, the arrangement of travel plans, and the practical application of bus connections still need further discussion with relevant staff in the field.

Data Availability

The bus single-trip time data used to support the findings of this study may be released upon application to the Beijing Transportation Operations Coordination Center (TOCC), who can be contacted at http://jtw.beijing.gov.cn/. Holiday data can be obtained at http://www.gov.cn/shuju/index.htm. Weather data can be obtained at https://www.wunderground.com/history/.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] B. A. Kumar, L. Vanajakshi, and S. C. Subramanian, "Pattern-based time-discretized method for bus travel time prediction," *Journal of Transportation Engineering, Part A: Systems*, vol. 143, no. 6, p. 04017012, 2017 UNSP.

[2] D Subhe, "Network-scale traffic modeling and forecasting with graphical lasso and neural networks," *Journal of Transportation Engineering*, vol. 138, no. 11, p. 23, 2022.

[3] G. Chen, X. Yang, J. An, and D. Zhang, "Bus-arrival-time prediction models: link-based and section-based," *Journal of Transportation Engineering*, vol. 138, no. 1, pp. 60–66, 2012.

[4] L. Jianhui, S. Yang, and S. Zhenhao, "A method of road traffic flow prediction based on ARIMA model," in *Proceedings of the Aeit 2012: 2012 2nd International Conference on Aerospace Engineering and Information Technology Hong Kong Education Soc: Tseung Kwan O*, M. Zhu, Ed., vol. Vol 2pp. 95–100, 2012.

[5] N. K. Chikkakrishna, C. Hardik, K. Deepika, and N. Sparsha, "SHORT-TERM TRAFFIC PREDICTION USING SARIMA AND FbPROPHET," in *Proceedings of the 2019 Ieee 16th India Council International Conference (Ieee Indicon 2019)*IEEE, New York, 2019.

[6] S. Mao, X. Xiao, M. Gao, X. Wang, and Q. He, "Nonlinear fractional order Grey model of urban traffic flow short-term prediction," *Journal of Grey System*, vol. 30, pp. 1–17, 2018.

[7] A. Kalman, *Filter Approach to Dynamic OD Flow Estimation for Urban Road Networks Using Multi-Sensor Data - Lu - 2015 - Journal of Advanced Transportation*, Wiley Online Library, Available at: https://onlinelibrary.wiley.com/doi/10.1002/atr.1292 accessed on 3 January 2022, 2022.

[8] X. Pan, W. Zhou, Y. Lu, and N. Sun, "Prediction of network traffic of smart cities based on DE-BP neural network," *IEEE Access*, vol. 7, pp. 55807–55816, 2019.

[9] D. Satrinia and G. A. P. Saptawati, "Traffic speed prediction from GPS data of taxi trip using support vector regression," in *Proceedings of the 2017 International Conference on Data and Software Engineering (icodse)*, IEEE, New York, 1-2 November 2017.

[10] M. Parida, K. Kumar, and V. K. Katiyar, "Prediction of urban traffic noise using artificial neural network approach," *Environ. Eng. Manag. J.* vol. 13, no. 4, pp. 817–826, 2014.

[11] W. Yanqiu, L. Qiang, Z. Jian, M. Lifeng, and W. Yu, "The city traffic flow prediction based on BP neural network," in *Proceedings of the 2008 Chinese Control and Decision Conference*, pp. 2550–2552, IEEE, Yantai, Shandong, 02-04 July 2008.

[12] X. Ban, C. Guo, and G. Li, "Application of Extreme learning machine on large scale traffic congestion prediction," in *Proceedings of the Proceedings of Elm-2015, Vol 1: Theory, Algorithms and Applications*, J. Cao, K. Mao, J. Wu, and A. Lendasse, Eds., vol. 6, pp. 293–305, Springer International Publishing Ag, Cham, 2016.

[13] H. Yu, R. Xiao, Y. Du, and Z. He, "A bus-arrival time prediction model based on historical traffic patterns," in *Proceedings of the 2013 International Conference on Computer Sciences and Applications (csa)*, pp. 345–349, IEEE, New York, 14-15 December 2013.

[14] F. Pili, A. Olivo, and B. Barabino, "Evaluating alternative methods to estimate bus running times by archived automatic vehicle location data," *IET Intelligent Transport Systems*, vol. 13, no. 3, pp. 523–530, 2019.

[15] S. O. Crudden and S. Berrebi, "An open-source framework to implement kalman filter bus arrival predictions," *Networks and Spatial Economics*, vol. 776, p. 21, 2018.

[16] Z. P. Tang, Z. X. Chen, J. P. Sun, Y. T. Hu, and M. Zhao, "Noise prediction of traction gear in high-speed electric multiple unit," *International Journal of Simulation Modelling*, vol. 18, no. 4, pp. 720–731, 2019.

[17] S. Mousavian, J. Valenzuela, and J. Wang, "Real-time data reassurance in electrical power systems based on artificial neural networks," *Electric Power Systems Research*, vol. 96, pp. 285–295, 2013.

[18] X. Wang, L. Huang, H. Huang, B. Li, Z. Xia, and J. Li, "An ensemble learning model for short-term passenger flow prediction," *Complexity*, vol. 2020, pp. 1–13, 2020.

[19] G. Zhong, T. Yin, L. Li, J. Zhang, H. Zhang, and B. Ran, "Bus travel time prediction based on ensemble learning methods," *IEEE Intelligent Transportation Systems Magazine*, vol. 14, no. 2, pp. 174–189, 2022.

[20] P. Balasubramanian and R. Rao, "An adaptive long-term bus arrival time prediction model with cyclic variations," *Journal of Public Transportation*, vol. 18, no. 1, p. 1, 2015.

[21] B. Ghosh, B. Basu, and M. O'Mahony, "Multivariate short-term traffic flow forecasting using time-series analysis," *IEEE Transactions on Intelligent Transportation Systems*, vol. 10, no. 2, pp. 246–254, 2009.

[22] W. Gu, V. V. Gayah, M. J. Cassidy, and N. Saade, "On the impacts of bus stops near signalized intersections: models of car and bus delays," *Transportation Research Part B: Methodological*, vol. 68, pp. 123–140, 2014.

[23] A. Agafonov and A. Yumaganov, "Bus arrival time prediction with LSTM neural network," in *Proceedings of the Advances in Neural Networks - Isnn 2019, Pt I*, H. Lu, H. Tang, and

Z. Wang, Eds., vol. 11554, pp. 11–18, Springer International Publishing Ag, Cham, 2019.

[24] B. Yu, W. H. K. Lam, and M. L. Tam, "Bus arrival time prediction at bus stop with multiple routes," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 6, pp. 1157–1170, 2011.

[25] Y. Bie, D. Wang, and H. Qi, "Prediction model of bus arrival time at signalized intersection using GPS data," *Journal of Transportation Engineering*, vol. 138, no. 1, pp. 12–20, 2012.

[26] B. Dhivya Bharathi, B. Anil Kumar, A. Achar, and L. Vanajakshi, "Bus travel time prediction: a log-normal autoregressive (ar) modelling approach," *Transportmetrica: Transportation Science*, vol. 16, no. 3, pp. 807–839, 2020.

[27] H. Chang, D. Park, S. Lee, H. Lee, and S. Baek, "Dynamic multi-interval bus travel time prediction using bus transit data," *Transportmetrica*, vol. 6, no. 1, pp. 19–38, 2010.

[28] J. Liu and G. Xiao, "Efficient bus arrival time prediction based on spark streaming platform," in *Proceedings of the Proceedings of the 2019 Ieee 23rd International Conference on Computer Supported Cooperative Work in Design (Cscwd)*, W. M. Shen, H. Paredes, J. Luo, and J. P. Barthes, Eds., pp. 416–421, IEEE, New York, 2019.

[29] X. Hua, W. Wang, Y. Wang, and M. Ren, "Bus arrival time prediction using mixed multi-route arrival time data at previous stop," *Transport*, vol. 33, no. 2, pp. 543–554, 2017.

[30] S. Wu and H. Nagahashi, "Parameterized AdaBoost: introducing a parameter to speed up the training of real AdaBoost," *IEEE Signal Processing Letters*, vol. 21, no. 6, pp. 687–691, 2014.

[31] S. Kotsiantis, "Combining bagging, boosting, rotation forest and random subspace methods," *Artificial Intelligence Review*, vol. 35, no. 3, pp. 223–240, 2011.

[32] G. Omkar and S. V. Kumar, "Development of simple exponential smoothing model for traffic flow prediction under heterogeneous traffic conditions," in *Proceedings of the Urbanization Challenges in Emerging Economies: Energy and Water Infrastructure; Transportation Infrastructure; and Planning and Financing*, U. P. Singh, B. R. Chahar, H. R. P. Yadav, and S. K. Vij, Eds., pp. 659–666, Amer Soc Civil Engineers, New York, 2018.

[33] T. Zhou, D. Jiang, Z. Lin, G. Han, X. Xu, and J. Qin, "Hybrid dual kalman filtering model for short-term traffic flow forecasting," *IET Intelligent Transport Systems*, vol. 13, no. 6, pp. 1023–1032, 2019.

[34] Z. Li, H. Yan, C. Zhang, and F. Tsung, "Long-short term spatiotemporal tensor prediction for passenger flow profile," *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 5010–5017, 2020.

[35] R. Gummadi and S. R. Edara, "Analysis of passenger flow prediction of transit buses along a route based on time series," in *Proceedings of the Information and Decision Sciences*, S. C. Satapathy, J. Tavares, V. Bhateja, and J. R. Mohanty, Eds., vol. 701, pp. 31–37, Springer International Publishing Ag, Cham, 2018.

[36] C. Li and P. Xu, "Application on traffic flow prediction of machine learning in intelligent transportation," *Neural Computing & Applications*, vol. 33, no. 2, pp. 613–624, 2021.

[37] B. Sun, W. Cheng, P. Goswami, and G. Bai, "Short-term traffic forecasting using self-adjusting k-nearest neighbours," *IET Intelligent Transport Systems*, vol. 12, no. 1, pp. 41–48, 2018.

[38] J. Khiari and C. Olaverri-Monreal, "Boosting algorithms for delivery time prediction in transportation logistics," in *Proceedings of the 20th Ieee International Conference on Data Mining Workshops (Icdmw 2020)*, G. DiFatta, V. Sheng,

A. Cuzzocrea, C. Zaniolo, and X. Wu, Eds., , pp. 251–258, Ieee Computer Soc: Los Alamitos, 2020.

[39] W. Alajali, W. Zhou, S. Wen, and Y. Wang, "Intersection traffic prediction using decision tree models," *Symmetry*, vol. 10, no. 9, p. 386, 2018.

[40] X. Zhou, P. Lu, Z. Zheng, D. Tolliver, and A. Keramati, "Accident prediction accuracy assessment for highway-rail grade crossings using random forest algorithm compared with decision tree," *Reliability Engineering & System Safety*, vol. 200, p. 106931, 2020.

[41] J. Pang, J. Huang, Y. Du, H. Yu, Q. Huang, and B. Yin, "Learning to predict bus arrival time from heterogeneous measurements via recurrent neural network," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 9, pp. 3283–3293, 2019.

[42] J. F. Ren, C. M. Ye, and F. Yang, "A novel solution to jsps based on long short-term memory and policy gradient algorithm," *International Journal of Simulation Modelling*, vol. 19, no. 1, pp. 157–168, 2020.

[43] W. Shu, K. Cai, and N. N. Xiong, "A short-term traffic flow prediction model based on an improved gate recurrent unit neural network," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–12, 2021.

[44] T. Zhou, G. Han, X. Xu et al., "δ -agree AdaBoost stacked autoencoder for short-term traffic flow forecasting," *Neurocomputing*, vol. 247, pp. 31–38, 2017.

[45] J. Vaish, S. S. Datta, and K. Seethalekshmi, "Short-term load forecasting using bootstrap aggregation based ensemble method," in *Proceedings of the 2021 7th International Conference on Electrical Energy Systems (icees)*, pp. 245–249, IEEE, Chennai, India, 11-13 February 2021.

[46] S. R. Sharma, B. Singh, and M. Kaur, "A novel approach of ensemble methods using the stacked generalization for high-dimensional datasets," *IETE Journal of Research*, vol. 78, pp. 1–16, 2022.

[47] V. S. Monego, J. A. Anochi, and H. F. de Campos Velho, "South America seasonal precipitation prediction by gradient-boosting machine-learning approach," *Atmosphere*, vol. 13, no. 2, p. 243, 2022.

[48] P. Tampubolon and A. S. Girsang, "Classification of attacks through the type of protocol using data mining," *JSMS*, vol. 88, p. 667, 2021.

[49] Y. Chen, Y. Lv, P. Ye, and F. Zhu, "Traffic-condition-awareness ensemble learning for traffic flow prediction,"vol. 53, pp. 582–587, in *Proceedings of the Ifac Papersonline*, vol. 53, pp. 582–587, Elsevier, Amsterdam, 2020.

[50] P. Yan, L. Zhang, Z. Feng, and J. Zhang, "Research on logistics demand forecast of port based on combined model," in *Proceedings of the 2018 International Conference on Computer Information Science and Application Technology*, vol. 1168, p. 032116, Iop Publishing Ltd, Bristol, June 2 2019.

[51] D. Billings and J.-S. Yang, "Application of the ARIMA models to urban roadway travel time prediction - a case study," in *Proceedings of the 2006 Ieee International Conference on Systems, Man, and Cybernetics*, p. 2529, Taipei, Taiwan, 08-11 October 2006.

[52] L. Rapant, "Traffic speed prediction using ensemble kalman filter and differential evolution," in *Proceedings of the 2018 6th International Conference on Traffic and Logistic Engineering (Ictle 2018) E D P Sciences: Cedex A*, R. Montemanni, Ed., vol. 259p. 02001, 2019.

[53] P. He, G. Jiang, S.-K. Lam, and D. Tang, "Travel-time prediction of bus journey with multiple bus trips," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 11, pp. 4192–4205, 2019.

[54] C. Bai, Z.-R. Peng, Q.-C. Lu, and J. Sun, "Dynamic bus travel time prediction models on road with multiple bus routes," *Computational Intelligence and Neuroscience*, vol. 2015, pp. 1–9, 2015.

[55] S. Bidwai, U. Wali, and S. Patil, "A comparative study of Markov chain and deep learning predictive models in spectrum sensing," *JSMS*, vol. 23, p. 8789, 2021.

[56] A. Vennelakanti, S. Shreya, R. Rajendran, D. Sarkar, D. Muddegowda, and P. Hanagal, "Traffic sign detection and recognition using a CNN ensemble," *Proceedings of the 2019 Ieee International Conference on Consumer Electronics (Icce)*, IEEE, , 11-13 January 2019.

[57] J. Xiao, Z. Xiao, D. Wang, J. Bai, V. Havyarimana, and F. Zeng, "Short-term traffic volume prediction by ensemble learning in concept drifting environments," *Knowledge-Based Systems*, vol. 164, pp. 213–225, 2019.

[58] J. Xiao, "SVM and KNN ensemble learning for traffic incident detection," *Physica A: Statistical Mechanics and Its Applications*, vol. 517, pp. 29–35, 2019.

[59] Y. Guo, X. Wang, P. Xiao, and X. Xu, "An ensemble learning framework for convolutional neural network based on multiple classifiers," *Soft Computing*, vol. 24, no. 5, pp. 3727–3735, 2020.

[60] B. Tang, Q. Chen, X. Wang, and X. Wang, "Reranking for stacking ensemble learning," in *Proceedings of the Neural Information Processing: Theory and Algorithms, Pt I*, K. W. Wong, B. S. U. Mendis, and A. Bouzerdoum, Eds., vol. 6443, pp. 575–584, Springer-Verlag Berlin, Berlin, 2010.

[61] H. Jiang, W. Zheng, L. Luo, and Y. Dong, "A two-stage minimax concave penalty based method in pruned AdaBoost ensemble," *Applied Soft Computing*, vol. 83, p. 105674, 2019.

[62] H. Huang, L. Huang, F. Jiao, R. Song, and J. Li, "Data-driven prediction of one-way bus running time: an integrated model," in *Proceedings of the 2021 the 5th International Conference on Management Engineering, Software Engineering and Service Sciences*, pp. 156–163, Association for Computing Machinery, New York, NY, USA, January 8 2021.

[63] B. Yu, H. Wang, W. Shan, and B. Yao, "Prediction of bus travel time using random forests based on near neighbors," *Computer-Aided Civil and Infrastructure Engineering*, vol. 33, no. 4, pp. 333–350, 2018.

[64] Z. Feng, L. Mo, and M. Li, "A random forest-based ensemble method for activity recognition," in *Proceedings of the 2015 37th Annual International Conference of the Ieee Engineering in Medicine and Biology Society (embc)*, pp. 5074–5077, IEEE, Milan, Italy, 25-29 August 2015.

[65] N. Rooney, D. Patterson, and C. Nugent, "Non-strict heterogeneous stacking," *Pattern Recognition Letters*, vol. 28, no. 9, pp. 1050–1061, 2007.

[66] Z. Liao, M. Su, G. Ning, Y. Liu, T. Wang, and J. Zhou, "A novel stacked generalization ensemble-based hybrid PSVM-PMLP-MLR model for energy consumption prediction of copper foil electrolytic preparation," *IEEE Access*, vol. 9, pp. 5821–5831, 2021.

[67] S. Shi, Y. Liu, Y. Huang, S. Zhu, and Y. Liu, "Active learning for KNN based on bagging features," in *Proceedings of the Icnc 2008: Fourth International Conference on Natural Computation, Vol 7*, M. Z. Guo, L. Zhao, and L. P. Wang, Eds., , p. 61, Ieee Computer Soc: Los Alamitos, 2008.

[68] J. Tang, L. Zheng, C. Han, F. Liu, and J. Cai, "Traffic incident clearance time prediction and influencing factor Analysis

using Extreme gradient boosting model," *Journal of Advanced Transportation*, vol. 2020, pp. 1–12, 2020.

[69] Y. Kim and W. Soh, "A study on character recognition of Korean vehicle license plates based on deep learning," *JSMS*, vol. 73, p. 231, 2021.

[70] R. Fu, Z. Zhang, and L. Li, "Using LSTM and GRU neural network methods for traffic flow prediction," in *Proceedings of the 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (yac)*, pp. 324–328, IEEE, Wuhan, China, 11-13 November 2016.

[71] P. J. Wu and D. Yang, "E-commerce workshop scheduling based on deep learning and genetic algorithm," *International Journal of Simulation Modelling*, vol. 20, no. 1, pp. 192–200, 2021.

[72] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," *Climate Research*, vol. 30, pp. 79–82, 2005.

Research Article

Dynamic Collaboration Model of Production Network Based on Cloud Service Bus

Yuan Zhao [ID] **and Shifeng Liu** [ID]

School of Economics and Management, Beijing Jiaotong University, Beijing 100044, China

Correspondence should be addressed to Yuan Zhao; 20113057@bjtu.edu.cn

The IT system of manufacturing enterprises usually has many problems, such as complex industrial software, different development languages, diverse communication protocols, and complex operation environment. Cloud service bus (CSB) technology based on service model encapsulates various applications existing in enterprises by means of the integration of cloud service bus and micro services, which can realize the rapid cloud migration and deployment of heterogeneous industrial application software. Theoretically, after the production system is connected to CSB, it can be arranged arbitrarily by service choreography technology to produce any possible products. However, in the context of industrial Internet, the production system connected to CSB corresponds to the equipment, materials, personnel, and other resources on the production line one by one. These production nodes need to consider the production capacity and cost of production service nodes and the output value of the whole production network, and cannot be combined arbitrarily. Therefore, when integrating production service nodes, we should not only consider the technical integration, but also consider whether the production conditions support this integration. To solve the problem of production node integration in CSB, a dynamic collaboration model of production network based on cloud service bus is proposed in this paper. The model takes the capacity, cost, and production relationship of production nodes as constraints, and the overall efficiency of production network as the optimization goal. The model can calculate the new creation, modification, deletion, and other scheduling operations of production line services in real time and give the production plan with the highest resource utilization and the greatest value in the current production network. The model can improve the rationality and economy of service choreography and give full play to the value of production network. Taking an enterprise with 11 production nodes and 5 production lines as an example, this paper discusses in detail how to use this model to calculate the optimal production organization scheme and the maximum output value of the enterprise.

1. Introduction

There are many problems in the IT system of manufacturing enterprises, such as complex industrial software, different development languages, diverse communication protocols, and complex operation environment. These software have great differences in connection form and management relationships. In the context of Industrial Internet, it is difficult for the production systems of these enterprises to access cloud computing. In view of the problems of various industrial communication protocols and complex operating environment of industrial software in manufacturing enterprises, CSB technology supports the service access, opening and conversion of multiple protocols, supports flexible and customized data transformation, supports the connection and integration of heterogeneous industrial software, and realizes the connectivity between enterprise organizations. CSB provides many adapter connectors, which can connect different types of applications and data services on local and cloud, effectively reducing the complexity of application system integration. The service choreography capability provided by CSB effectively solves the problems of choreography and unified application of heterogeneous industrial software after accessing the cloud level. At the same time, CSB uses the service choreography technology to combine and arrange multiple production

service nodes according to the production process flow to form a production line. Multiple production lines interweave to form a production network. The production service node connected to CSB is mapped to the manufacturing resources on the production line. Therefore, any service choreography of production service nodes should consider the production capacity, cost, and other constraints of the actual production line. Only in this way can the production network realize the integration of online and offline and improve the efficiency of the whole production network.

The production capacity provided by the production node connected to the bus is scarce and limited. Moreover, enterprises pursue the maximization of overall interest when integrating production nodes. Therefore, how to optimize the overall efficiency of the production network under the condition of limited capacity and cost is the key problem to be solved in the research field of heterogeneous system integration. This paper introduces a dynamic collaboration model of production network based on CSB. This model applies complex network technology to systematically analyze the production network constructed by service scheduling technology, and puts forward a dynamic collaboration model of production node resources to maximize the overall efficiency of the production network under the condition of limited production node capacity. Based on the online and offline integration of node lines in CSB, the model makes decisions on whether the new production line can be arranged, and gives the maximum production scheme of the production network. This method can make full use of existing resources and dynamically optimize the production network.

The rest of this paper is organized as follows. Section 2 reviews some related work, introduces the solution of CSB technology to integrate heterogeneous systems and analyses the problems faced by this solution in the actual production environment. Section 3 investigates the heterogeneous system integration and service choreography based on CSB, and analyses the characteristics of industrial Internet using CSB integration. Section 4 constructs a dynamic cooperative control model of production network based on CSB. In Section 5, through an example, this paper introduces how to apply the model to the dynamic cooperation of production resources to maximize the overall efficiency of the production network. Section 6 concludes the whole paper.

2. Related Work

2.1. Heterogeneous System Integration.
In the field of enterprise heterogeneous system integration, service-oriented architecture (SOA) is a common and mature method to solve enterprise heterogeneous system integration. Enterprise service bus (ESB) is the core component of SOA. It extends middleware by connecting heterogeneous components and systems and provides integration services. In addition, SOA also provides service choreography, service management, and other functions [1]. From the perspective of SOA architecture, it mainly solves the problems of heterogeneous system, technical access, and service quality management access. Service nodes in SOA mainly guarantee

the quality of service through quality-of-service system (QoS), but QoS mainly manages the nonfunctional requirements of access services. The main evaluation indicators of QoS include response time, reliability, and availability.

2.2. Cloud Manufacturing.
Cloud manufacturing is widely used in cross-enterprise integration in the manufacturing field. Based on the concept of manufacturing as a service, cloud manufacturing adopts new technologies such as cloud computing and Internet of Things to integrate the manufacturing resources connected to the network and provide high value-added, low-cost, and global product manufacturing services. Zhao et al. [2] combined with the characteristics of cloud manufacturing, discussed the connotation of manufacturing cloud service adaptation, including definition, characteristics, and content, and proposed a manufacturing cloud service adaptation technology framework composed of data source layer, data perception layer, data analysis and decision-making layer, and action execution layer. Xu et al. [3] proposed an adaptive bat algorithm (Saba) to solve the problem that optimizing and selecting appropriate services to complete manufacturing tasks is manufacturing services. Yin et al. [4] established a multi-agent-based manufacturing resource cloud service encapsulation adapter for the needs of distributed, heterogeneous and different conditions of manufacturing resource service encapsulation, and studied the adapter knowledge representation and accumulation method based on semantic directed graph.

2.3. Cloud Manufacturing Service Collaboration Based on Service Choreography.
At present, many scholars have studied manufacturing service collaboration based on service choreography technology. Through the orderly choreography of different services connected to cloud services, the efficiency of cloud manufacturing service system can be effectively improved. Maythaisong and Songpan [5] proposed mutation-based harmony search algorithm to select web services and to compose with minimum defects. This algorithm can help find appropriate solutions to compose services based on business plans. Yin et al. [6] proposed a cloud manufacturing service scheduling method based on WS-CDL. By realizing the standardized description of cloud manufacturing service collaboration, it promotes the effective communication between cross-organizational services and achieves the purpose of improving the resource utilization of service providers. Kopp and Leymann [7] proposed a modelling choreography method based on WS-BPEL to solve the problem that there is no agreed standard to describe choreography. You [8] constructed the technical framework of service-oriented composition evolution method for large-grained web services and proposed the determination method of evolution type and influence scope for large-grained web service composition. Ojstersek and Buchmeister [9] and Grznar et al. [10] applied simulation technology to study the production optimization strategy of enterprises under resource constraints. Hwang et al. [11]
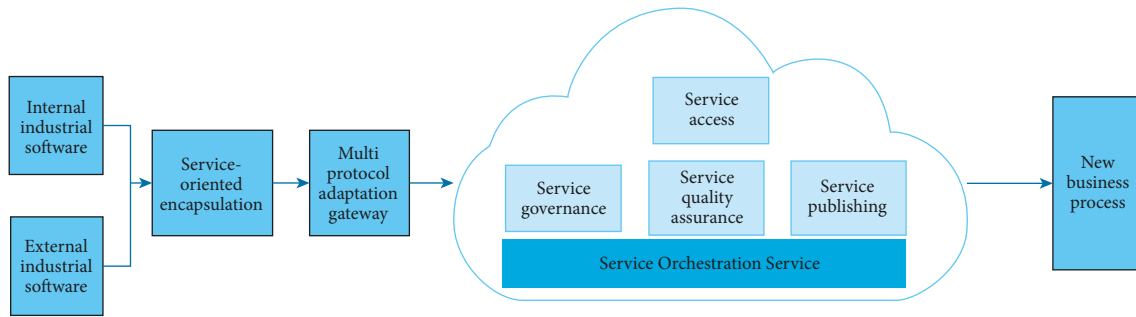
FIGURE 1: Structure of CSB.

studied the performance monitoring of message server and system based on MQTT. One of the core functions of CSB is to realize the message transmission between components stably and with low delay.

2.4. The Application of Complex System Theory. After the production service node is connected to CSB or cloud manufacturing system, it forms an intertwined mesh production system through service choreography, showing the characteristics of the complex network. Pan et al. [12] proposed a feasible architecture of multi-agent manufacturing system in a discrete manufacturing workshop to solve the problem of order uncertainty caused by the increase of customer personalized demand. Huang et al. [13] proposed a real-time dynamic scheduling mechanism of manufacturing system based on event triggering from the aspects of scheduling method, multi-agent coordinated control, and scheduling control algorithm. Li et al. [14] studied the multi-objective planning using ant colony algorithm, and verified through engineering practice that the optimized ant colony algorithm can avoid the low efficiency of optimal solution search and the shortage of initial pheromone.

To sum up, the research on heterogeneous system integration in the context of Industrial Internet mainly focuses on the following three aspects: firstly, SOA and cloud manufacturing technologies are used to solve the access integration problem at the technical level of heterogeneous services. Secondly, service choreography is used to solve the composition problem of discrete services. Finally, QoS is used to solve the quality assurance problem of access service. At present, there are few studies on the capacity of the connected production service node and the overall optimization of the production network.

3. Heterogeneous System Integration and Service Choreography Based on CSB

3.1. Heterogeneous System Integration Based on CSB. CSB technology realizes the encapsulation of heterogeneous systems through service encapsulation technology. Based on multi-protocol connection adaptation technology, CSB realizes the integration, choreography, governance, and other operations of heterogeneous systems, and finally provides services for internal and external applications of enterprises

through a unified service gateway. The structure of CSB is shown in Figure 1:

In Figure 1, internal industrial software and external industrial software represent the internal and external software systems used by the enterprise in production. The function of service-oriented encapsulation component is to combine different types of production software into standardized service modules in a technical way. The multi-protocol adaptation gateway component is mainly used to adapt the industrial software with different communication protocols, so that the software with different communication protocols can communicate with each other.

To solve the problems of poor compatibility and inconsistent architecture of industrial software in manufacturing enterprises, CSB provides a unified and highly compatible heterogeneous software packaging solution to realize the service-oriented packaging of enterprise heterogeneous software. In view of the problems of diverse communication protocols and complex operating environment of industrial software in manufacturing enterprises, CSB provides multi-protocol adaptation gateway to support the service access, opening and conversion of common protocols, and flexible and customized data transformation, to support the connection and integration of heterogeneous industrial software and realize the connectivity between enterprise organizations. At the same time, CSB provides rich adapter connectors, which can connect different types of applications and data services on local and cloud, effectively reducing the complexity of application system integration. After accessing CSB, the service quality of production service nodes is guaranteed by QoS system. QoS can provide quality assurance for nodes, including reliability and response speed.

3.2. CSB-Based Service Choreography. After the heterogeneous software is connected to the service, it is managed by the service management function provided by CSB. The service choreography function provided by CSB will connect multiple production nodes of CSB, effectively, integrate the business process according to the operation process of product production, realize the reasonable choreography of various services, and form a new production line. In this way, enterprises can reuse existing heterogeneous assets, quickly solve the production problems of new products, reduce the cost of production line construction, and improve
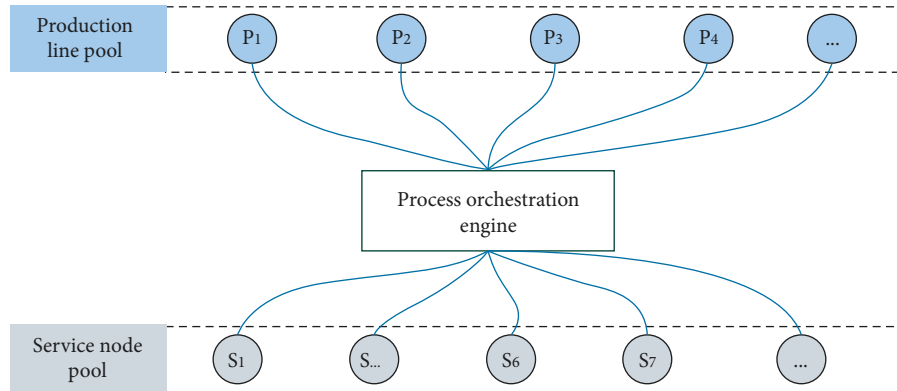
FIGURE 2: Structure of the production network.

the ability of enterprise applications to market changes. The production node connected to CSB will be called by multiple production processes, and the processes of multiple production lines will be intertwined to form a complex production network system. The structure of the production network is shown in Figure 2.

As shown in Figure 2, the production line pool is the collection of all production nodes connected to the CBS. Service node pool is the collection of production control systems integrated into the system through CBS. The production line in the production line pool is the production line formed by the process engine connecting and combining multiple production nodes according to the product production scheme. Multiple production nodes and multiple production lines call and interweave each other to form a complex production network.

3.3. Characteristics of Industrial Internet Using CSB Integration.

Under the condition of Industrial Internet, there is a big difference between the integration of heterogeneous systems using CSB and the integration that only provides computing, storage, and other services. The main differences are as follows:

(1) Strong mapping relationship between online production nodes and offline production resources. Since the production nodes connected to CSB have a unique mapping relationship with the offline production system and resources, the reorganization of the offline production line corresponding to the assembly of online services needs to consider various factors, such as production capacity, cost, effective utilization of production resources, and maximization of the value of production products. This strong correspondence between online and offline resources has more restrictions on service integration under the CSB mode.

(2) Production nodes are limited by many factors. Each production node has various resource constraints such as capacity and cost. Since each node corresponds to a unique production resource one by one, the capacity constraints of each node need to be

considered. Therefore, production nodes cannot be accessed without restrictions.

(3) The choreography of the production line is related to the production decision of the enterprise. Since the production line in CSB is the mapping of physical production lines, the profit of the enterprise should be considered in the combination of online production services and the production plan of each production line. To achieve the optimal efficiency of the production network, it is necessary to calculate and dynamically adjust the production strategy of products in the production network in real time, make full use of production resources, and maximize the profits of enterprises.

At present, there is no restriction strategy for production nodes in the cloud service integration architecture, and there is no strategy to achieve the optimal efficiency of the production line. Therefore, this paper proposes a dynamic collaboration model of production network based on CSB, which mainly solves the problems of service choreography and production planning after heterogeneous systems are connected to CSB. On the basis of fully considering the production constraints, the optimal combination of service choreography and production strategy is realized.

4. Dynamic Cooperative Control Model of Production Network Based on CSB

4.1. Structure of the Model.

To realize the dynamic system control of production network, it is necessary to redesign the structure of the original production network and add the dynamic collaboration model. Based on the production network structure of CSB shown in Figure 2, the model is upgraded, and the structure is shown in Figure 3.

As shown in Figure 3, the function of the "Dynamic collaboration model" component is to dynamically control the service orchestration according to certain rules to ensure the availability and rationality of the production line. After the production control model is added to the production network, the operation of the service choreography engine can be controlled according to the agreed rules. The dynamic collaboration model of production network is an event-
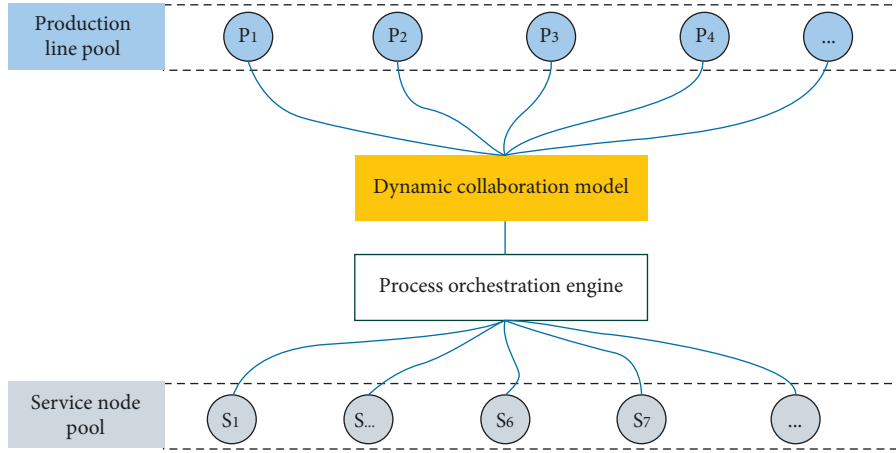
FIGURE 3: Service choreography and control model.
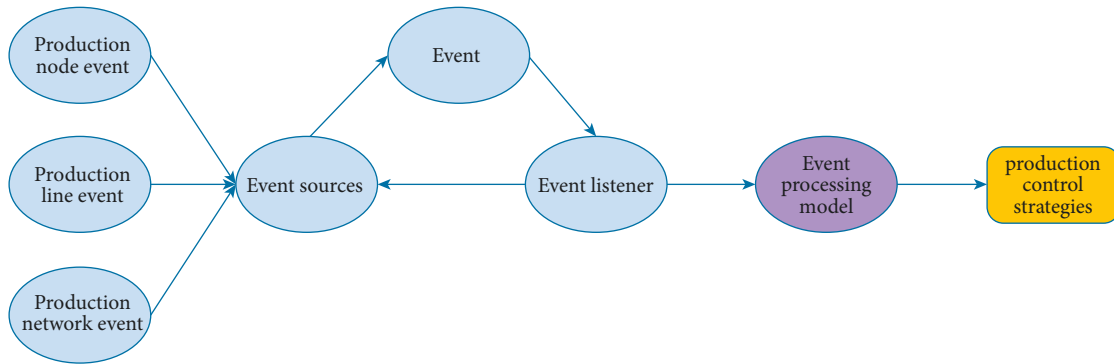
FIGURE 4: Dynamic cooperative control model based on event processing.

driven model, which can respond to the change events of the production network and realize the dynamic collaboration control of the production network, to achieve the optimal configuration of the network.

The operation mode of the collaborative control model is that the event processing center processes the events generated by the production node, production line, and production network according to the predefined processing rules and optimization objectives, and outputs the overall control strategy of the model. Its structure is shown in Figure 4.

As shown in Figure 4, the dynamic cooperative control model adopts the classical event processing model to realize the dynamic cooperative control. Event sources include: production node events, production line events, and production network events. Node events include production node addition, removal, capacity change, and cost change. Production line events include changes in the arrangement of production nodes, changes in product prices, and changes in the production priority of the production line. Production network events include changes in optimization objectives and restrictions. After the model change event occurs, it will be propagated to the registered listeners. The listener then passes the event to the event processing model for processing and outputs the control strategy. The control strategy mainly

includes the change decision of nodes and production lines and production planning strategy.

4.2. Dynamic Collaborative Control Model of Production Network. This paper analyzes the network structure of heterogeneous systems based on CSB by using the theory and method of complex network systems.

According to the network structure in Figure 3, the production network is composed of production nodes and production line nodes. The production node represents the system that can provide production services connected to the CSB. The production line node is composed of the production line generated by the service node through choreography. Products need to be produced in a certain order, so the edges between nodes are directed. Edge connection indicates that there is a production combination relationship between nodes. Because the production of each product needs to consume the resources of the node, so each edge has a weight.

Let G be an abstract representation of the production network based on CSB, and define $G(S, P, E, W)$ as a directed and weighted network, where: $S = \{s_1, s_2, s_3, \ldots, s_m\}$, S represents the collection of all production systems connected to CSB, m is the total number of production nodes connected in CSB, and each s_i represents the production

resources responsible for completing certain business functions; $P = \{p_1, p_2, p, \ldots, p_n\}$, P represents the set of production line nodes. Each production line is composed of multiple production service nodes according to a certain production order; $E = \{e_1, e_2, e_3, \ldots, e_l\}$, $(s_1, s_2) \in E$, E represents the set of production relations between production nodes, the directed edge of nodes s_1 to s_2 in the network, and the production dependency of s_1 to s_2 in the production process. Since the same node can be combined into multiple production lines, there are multiple edges derived from a specific node; $W = \{w_1, w_2, w_3, \ldots, w_k\}$, W is a set describing the weight of each edge. w means that the production of a specific product p needs to consume the capacity of the corresponding production node.

There are multiple production lines in production network $G(S, P, E, W)$. The quantity and price of products produced by each production line are different, so the output value of production network G in unit time can be defined as:

$$V = \sum_{i=1}^{n} q_i \times v_i, \tag{1}$$

where V represents the total output value of the production network, n represents the number of production lines, and q_i and v_i represent the output and price of line p_i.

In the same way, the cost, consumption, and profit calculation functions of the production network can be defined. According to these functions, the optimization objectives of the model, such as maximum output value, maximum profit, and minimum cost, can be defined.

The constraints of the model mainly include node capacity, cost, and industrial logic requirements of upstream and downstream nodes. Next, the node capacity constraints are analyzed. In production network G, multiple production lines can share the same production node, that is, the same production node can provide services for multiple production lines. Let O_{s_j} represent the production capacity of node s_j, and $w_{p_i s_j}$ represent that the production line p_i needs to consume the production capacity of production node s_j. In the production network, the capacity consumed by any production line using the same production node cannot exceed the capacity of that node. This relationship can be defined as:

$$\sum_{i=1}^{n} w_{p_i s_j} \leq O_{s_j}. \tag{2}$$

Based on the above analysis, the optimal production plan of production network G can be described as the optimal production plan to realize the output value, cost, and profit of production network under the constraints of capacity and cost of each node. If the maximum output value is taken as the optimization objective, the objective function of the model is as follows:

$$\max(V) = \max\left(\sum_{i=1}^{n} q_i \times v_i\right). \tag{3}$$

There are two main factors that affect the decision-making of an enterprise. First of all, the production line arrangement affects the product category decision of the enterprise. The products produced by different production lines bring different profits to the enterprise, so the enterprise needs to make decisions on which production lines to arrange. Second, the production line arrangement affects the output decision of each production line. Because production lines need to consume the capacity of production nodes, enterprises need to make decisions on the output of each production line under the condition of limited capacity.

According to the business needs of the enterprise, the user can establish the set of constraints in the way of equation (2), establish the optimization objective function according to the requirements of equation (3), and finally form the optimization control model.

4.3. Application of the Production Network Dynamic Collaboration Model.
Combined with the service choreography capability provided by CSB, the production network dynamic collaboration model can realize the following functions:

(1) Dynamic control of production nodes. After the production node is changed, the model realizes resource control and early warning by monitoring whether the resources of the production node are within the limit.

(2) Dynamic capacity expansion and adjustment of production nodes. The model can calculate the resource consumption of production network nodes in real time, and enterprises can find the bottleneck nodes. Enterprises usually face the decision of expanding production nodes. Combined with the decision-making objectives of the model, the model can calculate the costs and benefits of capacity expansion nodes and provide a basis for enterprises to make capacity expansion decisions.

(3) Dynamic control of production line editing. Managers need to judge the impact of these operations on the production line and the whole network when adding, modifying, and deleting the production line. This model evaluates the impact of production line changes on the overall income of the enterprise by calculating the income of the production line itself and the network, and provides decision-making basis for managers.

(4) Dynamic adjustment of production strategy. In the production network, if the production capacity and price of products change, and the optimal target rules of the model are adjusted, the model can control which production lines are produced and how many are produced.

5. Model Validation and Application

In the context of Industrial Internet, production enterprises can use the heterogeneous system integration, service management, service choreography, service control, and other functions provided by CSB to carry out the flexible combination of production lines according to the market

demand, to realize the rapid integration of internal and external resources of enterprises.

Taking a production enterprise as an example, this part explains the operation mechanism of the production network dynamic collaboration model based on CSB through the control of nodes, production lines, and networks. There are 11 internal and external production nodes integrated by the enterprise, which are combined into 5 production lines through the service choreography group function provided by CSB. The structure of the production service nodes included in the five production lines is shown in formula (1).

$$P = \begin{bmatrix} p_1\left(S_1, S_2, S_3, S_4, S_5\right) \\ p_2\left(S_1, S_2, S_3, S_4, S_6\right) \\ p_3\left(S_7, S_8, S_9\right) \\ p_4\left(S_1, S_2, S_7, S_{10}\right) \\ p_5\left(S_3, S_8, S_{11}\right) \end{bmatrix}. \tag{4}$$

The value of the production line in the production network can be expressed as: $V = \{v_1, v_2, v_3, v_4, v_5\}$ where v_i is the value generated by the production line p_i.

According to the characteristics that the production network G conforms to the bisection network [15], the projection method is used to analyse the production network G, and the structure diagram of the production network can be obtained, as shown in Figure 5.

In Figure 5, there are 11 nodes, 15 interactive edges, and 5 production lines to produce 5 different products. The price of each product is different. Since the same production node can serve multiple production lines, there are multiple edges between two production nodes. The weight $w_{p_k i}$ on each edge indicates that production line p_k needs to consume the capacity of production node S_i. The weight matrix of the production network is as follows:

$$w = \begin{bmatrix} w_1\left(w_{p_1 1}, w_{p_1 2}, w_{p_1 3}, w_{p_1 4}, w_{p_1 5}\right) \\ w_2\left(w_{p_2 1}, w_{p_2 2}, w_{p_2 3}, w_{p_2 4}, w_{p_2 6}\right) \\ w_3\left(w_{p_3 7}, w_{p_3 8}, w_{p_3 9}\right) \\ w_4\left(w_{p_4 1}, w_{p_4 2}, w_{p_4 7}, w_{p_4 10}\right) \\ w_5\left(w_{p_5 3}, w_{p_5 8}, w_{p_5 11}\right) \end{bmatrix}. \tag{5}$$

The dynamic collaboration model of production network proposed in this paper aims at maximizing the output value of the production network. According to the above analysis, the optimal production function of this enterprise can be expressed as:

$$\text{Max}\,(p) = V * Q, \tag{6}$$

where V is the value set of each product, and Q is the quantity produced by each production line. One production node can be used by multiple production nodes. The total capacity of a node consumed by multiple production lines cannot exceed the total capacity of the node. The capacity of any production node S_i should meet the conditions expressed in formula (2).
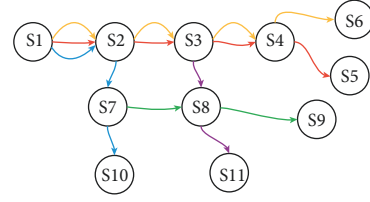
Figure 5: Production network diagram.

Table 1: Production information of network.

Category	P1	P2	P3	P4	P5	Capacity
S1	16.8	30.4	0	25.3	0	800
S2	15.9	25.4	0	29.3	0	400
S3	30.5	0	0	0	40.2	500
S4	42.3	15.6	0	0	0	600
S5	22.1	0	0	0	0	700
S6	0	20.3	0	0	0	500
S7	0	0	32	16.3	0	300
S8	0	0	24.3	0	20.5	250
S9	0	0	33.4	0	0	350
S10	0	0	0	25.5	0	300
S11	0	0	0	0	39.5	350
Price	130	105	200	155	130	—

The production node capacity, consumption of each production line, and product price of the production enterprise are shown in Table 1.

By substituting the data in Table 1 into formula (6) the production dynamic control model of the enterprise can be obtained. This paper uses the linear programming method of operations research to solve it. The results of model calculation are shown in Tables 2 and 3.

Because the production data in Table 1 corresponds to the nodes and edges of the production network in Figure 5, the changes of the network model will be transferred to the model parameters. This change triggers the calculation of the model through the event transmission mechanism to realize the purpose of dynamic collaborative control of network change. Managers can make decisions based on the results calculated by the model. The specific applications of the model are shown as follows.

(1) Real-time calculation of the optimal production plan of the production network. Bring the production data in Table 1 into formula (2) and use the linear programming method of operational research to find that when the product output $q = \{11, 8, 9, 0, 1\}$, $v = \{1430, 840, 1800, 0, 130\}$, the maximum output of the production network can be 4200 under the condition of meeting the constraints of the production node.

(2) Collaborative analysis of production node capacity. The model can calculate the capacity consumption of each production node when the output value of the production network is maximized. Managers can make scientific decisions according to the capacity consumption of production nodes. For example, it can be seen from the calculation results that the

TABLE 2: Consumption of production node capacity.

Category	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11
Capacity	800	400	500	600	700	500	300	250	350	300	350
Consumption	428	378.1	375.7	590.1	243.1	162.4	288	239.2	300.6	0	39.5

TABLE 3: Output and output value of production line.

Category	P1	P2	P3	P4	P5
Price	130	105	200	155	130
Quantity	11	8	9	0	1
Value	1430	840	1800	0	130

consumption of production node $S4$ is close to the production capacity. This may be a key node restricting the capacity of the entire production network. For nodes with large capacity surplus, enterprise managers can fully consume the remaining capacity by developing new products, or share the capacity on the Industrial Internet to realize cross-organizational capacity optimization; for production nodes that tend to be saturated with capacity, enterprise managers can expand the capacity of nodes to expand production and increase enterprise income.

(3) Collaborative decision analysis of production line. The model can provide decision support for the change of production line. It can be seen from the calculation results that the output of $P4$ is 0, that is, no production is carried out. The manager can decide whether to remove production line P4. Furthermore, if the maximum output value of the production network is reduced after the new production line is added, the production line cannot join the production network. Otherwise, you can join the production network. In production practice, it is necessary to realize the optimization of production network under the condition of product output limitation. The model proposed in this paper also supports this scenario and only needs to add the output limit condition of the production line to the limit condition of the model.

6. Conclusion

Manufacturing enterprises use the Industrial Internet to realize cross-organizational business integration. The enterprise encapsulates the existing capacity into services and deploys them in the cloud service environment. With the integration capability provided by CSB, the integration of internal and external production services is realized to meet the requirements of flexible production and intelligent manufacturing. In this paper, we systematically introduce the technical system of production service integration and choreography based on CSB, and propose a dynamic collaboration model of production network based on CSB. Then, we use the theory and method of complex network to model the production network based on CSB, and establish a dynamic collaboration model, which provides support for the realization of dynamic collaboration control of service choreography. Finally, an example is used to introduce the application method and effect of the model in detail.

By adding the dynamic collaboration model to CSB system, this paper solves the problem of dynamic collaboration control in CSB model and improves the rationality and economy of production network. In the process of model building, this paper does not consider the time consumed by the node to produce unit products, which may lead to the production line unable to produce as planned. Next, we will do further research in application scenarios such as multi-factor constraints and multiobjective collaboration to optimize the model and improve the generalization ability of the model.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] M. P. Papazoglou and W. J. Van den Heuvel, "Service oriented architectures: approaches, technologies and research issues," *The VLDB Journal*, vol. 16, no. 3, pp. 389–415, 2007.

[2] Q. Zhao, L. Wei, and H. Shu, "Adaptive technology framework of manufacturing cloud service in cloud manufacturing environment," *Journal of Chengdu University of Information Technology*, vol. 36, no. 1, pp. 45–50, 2021.

[3] B. Xu, J. Qi, X. Hu, K. S. Leung, Y. Sun, and Y. Xue, "Self-adaptive bat algorithm for large scale cloud manufacturing service composition," *Peer-to-Peer Networking and Applications*, vol. 11, no. 5, pp. 1115–1128, 2018.

[4] H. Yin, C. Yin, W. Wang, and M. Wang, "Research on cloud service encapsulation method and adapter of manufacturing resources based on multi agent," *Computer Integrated Manufacturing Systems*, vol. 20, no. 12, pp. 3182–3188, 2014.

[5] E. Maythaisong and W. Songpan, "Mutation-based harmony search algorithm for hybrid testing of web service composition," *Computational Intelligence and Neuroscience*, vol. 2018, Article ID 4759405, 2018.

[6] H. Yin, C. Yin, W. Wang, and M. Wang, "Cloud manufacturing service collaboration based on service

choreography," *China Mechanical Engineering*, vol. 32, no. 7, pp. 849–859, 2021.

[7] O. Kopp and F. Leymann, "Choreography design using WS-BPEL," *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*, vol. 31, no. 3, pp. 31–34, 2008.

[8] D. L. You, *Research on Composition Evolution Methods and Key Technologies for Large-Granularity Web Service of Service-Oriented Choreography*, Yanshan University, Qinhuangdao, China, 2014.

[9] R. Ojstersek and B. Buchmeister, "Simulation based resource capacity planning with constraints," *International Journal of Simulation Modelling*, vol. 20, no. 4, pp. 672–683, 2021.

[10] P. Grznar, M. Gregor, M. Gaso, G. Gabajova, M. Schickerle, and N. Burganova, "Dynamic simulation tool for planning and optimisation of supply process," *International Journal of Simulation Modelling*, vol. 20, no. 3, pp. 441–452, 2021.

[11] K. Hwang, J. M. Lee, and I. H. Jung, "Performance monitoring of MQTT-based messaging server and system," *Journal of Logistics, Informatics and Service Science*, vol. 9, no. 1, pp. 85–96, 2022.

[12] J. Pan, D. Tang, Z. Zhang, J. Wu, and J. Feng, "Research on multi-agent manufacturing system scheduling based on rule adjustment," *Machine Building & Automation*, vol. 50, no. 5, pp. 160–163, 2021.

[13] T. Huang, D. Tang, H. Zhang, and C. Xu, "Research on real-time and dynamic scheduling mechanism of intelligent manufacturing system based on agent," *Machine Building & Automation*, vol. 46, no. 6, pp. 164–168, 2017.

[14] Y. Li, S. Wang, and Y. He, "Multi-objective optimization of construction project based on improved ant colony algorithm," *Tehnicki Vjesnik-Technical Gazette*, vol. 27, no. 1, pp. 184–190, 2020.

[15] S. P. Borgatti and M. G. Everett, "Network analysis of 2-mode data," *Social Networks*, vol. 19, no. 3, pp. 243–269, 1997.

Review Article

Improved Fitness-Dependent Optimizer for Solving Economic Load Dispatch Problem

Barzan Hussein Tahir [iD],[1] **Tarik A. Rashid** [iD],[1] **Hafiz Tayyab Rauf** [iD],[2] **Nebojsa Bacanin** [iD],[3] **Amit Chhabra** [iD],[4] **S. Vimal** [iD],[5] **and Zaher Mundher Yaseen** [iD][6,7,8]

[1]*Department of Computer Science and Engineering, University of Kurdistan Helwer, Erbil, Iraq*
[2]*Centre for Smart Systems, AI and Cybersecurity, Staffordshire University, Stoke-on-Trent, UK*
[3]*Singidunum University, Danijelova 32, Belgrade 11000, Serbia*
[4]*Department of Computer Engineering and Technology, Guru Nanak Dev University, Amritsar, India*
[5]*Department of Artificial Intelligence and Data Science, Ramco Institute of Technology, North Venganallur Village, Rajapalayam 626 117, Virudhunagar District, Tamilnadu, India*
[6]*Department of Earth Sciences and Environment, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, Bangi 43600, Selangor, Malaysia*
[7]*Adjunct Research Fellow, USQ's Advanced Data Analytics Research Group, School of Mathematics Physics and Computing, University of Southern Queensland, Toowoomba, QLD 4350, Australia*
[8]*New Era and Development in Civil Engineering Research Group, Scientific Research Center, Al-Ayen University, Thi-Qar 64001, Iraq*

Correspondence should be addressed to Zaher Mundher Yaseen; yaseen@alayen.edu.iq

Economic load dispatch depicts a fundamental role in the operation of power systems, as it decreases the environmental load, minimizes the operating cost, and preserves energy resources. The optimal solution to economic load dispatch problems and various constraints can be obtained by evolving several evolutionary and swarm-based algorithms. The major drawback to swarm-based algorithms is premature convergence towards an optimal solution. Fitness-dependent optimizer is a novel optimization algorithm stimulated by the decision-making and reproductive process of bee swarming. Fitness-dependent optimizer (FDO) examines the search spaces based on the searching approach of particle swarm optimization. To calculate the pace, the fitness function is utilized to generate weights that direct the search agents in the phases of exploitation and exploration. In this research, the authors have used a fitness-dependent optimizer to solve the economic load dispatch problem by reducing fuel cost, emission allocation, and transmission loss. Moreover, the authors have enhanced a novel variant of the fitness-dependent optimizer, which incorporates novel population initialization techniques and dynamically employed sine maps to select the weight factor for the fitness-dependent optimizer. The enhanced population initialization approach incorporates a quasi-random Sabol sequence to generate the initial solution in the multidimensional search space. A standard 24-unit system is employed for experimental evaluation with different power demands. The empirical results obtained using the enhanced variant of the fitness-dependent optimizer demonstrate superior performance in terms of low transmission loss, low fuel cost, and low emission allocation compared to the conventional fitness-dependent optimizer. The experimental study obtained 7.94E−12, the lowest transmission loss using the enhanced fitness-dependent optimizer. Correspondingly, various standard estimations are used to prove the stability of the fitness-dependent optimizer in phases of exploitation and exploration.

1. Introduction

After the development of computers, the main objective was to investigate unknown solutions and find the best possible solution. During World War II, Alan Turing broke a cipher of Germany named Enigma by using an algorithm used for searching [1]. Many challenges arose in solving problems in real life due to the improvements in working methods and the exciting acceleration in the extent of computations. Hence, techniques based on numerical programming and conventional logic emerged to overcome the drawbacks of instantly and capably resolving complicated problems [2]. Various algorithms like optimization problems have been designed to manage these limitations. The best possible solution was gained through the optimization method by studying its parameter. All the possible values of present solutions were expressed as a set one of which is the fittest solution. Usually, problems of optimization are solved to design algorithms of optimization [3]. Optimization algorithms are classified into two groups: stochastic algorithms and deterministic algorithms [4, 5]. Deterministic algorithms generate a group of related answers when the iterations are started by using an introductory initial point all this happened by using inclination [6]. On the other hand, stochastic algorithms constantly generate distinct answers with related values in the absence of inclination. Diversely, concluding values have a slight difference. There are two categories of stochastic algorithms: metaheuristic and heuristic [7, 8].

Heuristic algorithms use the trial-and-error method to find a solution, and it is supposed that these algorithms will consume reasonable time to reach a solution [9, 10]. Moreover, the aim of heuristic algorithms is to utilize various methods in local examinations and techniques of randomization [11]. Further analysis and advancements were made in heuristic algorithms and converted to metaheuristic algorithms [12, 13]. The novel collections of algorithms have better performance than heuristic algorithms; accordingly, the affix of "meta" that means "far off" was linked with these algorithms.

Recently, available problems of the real world have turned complicated in considerations of cost, time, and space; it is not possible to traverse all credible solutions. Hence, fast and low-cost techniques are required [14, 15]. Thus, scientists studied the natural events and behaviors of animals to resolve these issues, like how ants select their paths; how fish, flies, or birds chase their prey; and the working of gravity. So, all the algorithms inspired by nature are called nature-inspired algorithms [16]. FDO algorithm that is also known as the fitness-dependent optimizer was introduced by Jaza Abdullah and Tarik Rashid in 2019. The FDO algorithm studies the bee swarms' reproduction practices and follows the activities of swarms. This algorithm finds out the best solution among the pool of solutions [17]. Intelligent computations are rising in various areas of research because of their capability to integrate with large complex and interconnected systems with high speed and accuracy.

Economic load dispatch (ELD) is the most vital and significant field of power system planning and operation [18, 19]. The chief aim of ELD was to list a group of real power provided by resources of online generation to satisfy the lacked demand whenever needed under a group of limitations [20], regarding system and unit technical constraints with the least production cost. In general, the ELD problem can be validated as an extremely nonlinear and nonsmooth stifled problem of optimization usually for huge systems. The cost of fuel is concerned with the varying costs of the generation of electricity.

The main purpose of addressing the ELD problem is to determine the necessary output power to satisfy the system's requirements in such a way that the cost is limited to its possible minimum value and limits, such as prohibited operating zone (POZ) and valve-point effects (VPEs) [20]. There is a need for efficient ways of producing electricity. Increasing the cost of fuel makes the method of power production expensive. Advanced systems are therefore expected to propose an economical generation, delivery, and transmission method while keeping electrical limitations in mind. The total device requirement is divided into several units by ELD, which reduces the total cost of generation. Several complications exist while obtaining the ELD problem's global best solution; the results are less accurate due to the nonlinear nature of classical methods, confined to convergence issues, and the best local solutions [6].

The major drawback of some techniques like evolutionary computing is premature convergence [21]. In heuristics, exploitation, and exploration perform a significant role. The capacity of an algorithm to hunt globally is called exploration, and the ability to search locally is known as exploitation [22]. The stability of exploration and exploitation highly affects the swarm-based algorithms' performance. Smaller exploration and extreme exploitation lead to premature convergence; and on the other hand, more limited exploitation and higher exploration can cause barriers to gaining the best solution [23].

In this research, by reducing fuel expense, emission allocation, and transmission loss, the authors have implemented FDO to solve the ELD problem. Similarly, the authors have enhanced a new FDO variant that consolidates specific techniques of population initialization and manipulates sinus maps to pick the FDO weight factor dynamically. The enhanced population initialization method combines a quasi-random Sabol sequence to generate the initial solution for the multidimensional search space. A typical 24-unit device is implemented with different power demands for preliminary evaluation. The observational results obtained by implementing the enhanced FDO variant illustrate the outstanding performance compared to the standard FDO in low transmission loss, low fuel cost, and low emission allocation.

2. Related Work

In this section, various studies have been reported similar to the enhanced approach but with different evolutionary and

swarm-based algorithms, that is, PSO, BA, DE, GA, BCO, and other combinations of the recent state-of-the-art algorithms.

The authors carried out a study to suggest a method for the solution of ELD problems and to contrast it with various available solutions [24]. Their enhanced method owns the following characteristics, managing issues of non-differentiability, all the problems caused by PBC are also resolved, and problems of multi-objective nature of PBC are also resolved. They implemented their approach on 5 generation systems, and the achieved outcomes proved that more effective Pareto-curve is gained. An enhanced algorithm inspired by nature, that is, Bat algorithm that offers firm convergence and excellent computational performance, was conducted by [7]. Yang enhanced the BAT algorithm in 2010 after inspiring by bat's echolocation behavior. They stated that this bat quality enables the bat to locate the prey, that is, various insects, even in the absence of light. Their enhanced method aimed at reducing the total cost of generation in the case of a thermal power plant.

In another research, an enhanced version of GA by utilizing mutation and crossover for the solution of CHPED problems was adopted by [25]. They stated that primary GA is grown in three phases. In their first phase, they did not include the process of selection to bypass population diversity loss, while in the second phase, they utilized two various crossover operations to dig data about parents and produce possible children. In the third phase, they used the operation of mutation to substitute children with children of other parents. They proved that their enhanced algorithm is the best substitute for the CHPED problem. Another study [26] was carried out to propose a novel quantum bat algorithm (QBA) based on quantum computing; its main purpose was to solve the problem of multi-objective combined economic emission dispatch (CEED). To minimize the system's nonlinearities, they represented CEED by utilizing the function of the cubic criterion. Their primary concern was the eruption of CO2, NOx, and NOx and load dispatch. Thus, it is known as a multi-objective problem of optimization. Their outcomes proved that QBA is the best solution for the problem of CEED as compared to different available solutions. An improved self-adaptable differential evolution algorithm integrating with multiple mutation strategies (ADE-MMSs) as a solution to ELD problems [27]. They suggested a strategy to improve and explore of basic DE problem. Their enhanced method has 3 expansions of DE. Furthermore, they suggested an approach to manage constraints of equality of problems of ED. Their algorithm enhances the speed of convergence as well as maintains a balance between exploration and exploitation. ADE-MMS is proved by them to be the most suitable solution.

Novel differential evolution algorithm was enhanced in a study [28], for the solution of simultaneous power flow OARPD problems for renewable generators. Their enhanced algorithm, that is, DEa-AR, utilized a combination of arithmetic crossover and performed scaling based on Laplace distribution. For the evaluation of their approach, they utilized the IEEE 57-bus system in various situations. The outcomes of their simulations verified that the suggested

method could be used for solving OARPD problems with sources that are efficiently renewable and can provide optimum solutions. Qiao and Liu [29] have carried out to propose a combined framework of EVs and wind farms (WEVs) that reduced the over and underestimation of wind power by utilizing the discharging and charging capability of EVs. They designed a dynamic economic emission dispatching based on the WEV system (WE_DEED). They utilized their algorithm for the solution of complicated problems of WE_DEED. While they handle the limitations of WE_DEED through their enhanced algorithm. They verified their algorithm on various 10-unit systems.

The authors enhanced a novel technique [30], for thermal plants' dispatch generating powers based on motion optimization algorithm (IMA). They achieved ELD being an objective function through implementing IMA. For the testing phase, multiple instances of various units of thermal plants were utilized to examine the execution of their algorithm. Their preceding outcomes were matched with various approaches.

A detailed description of related ELD applications concerning different evolutionary approaches is given in Table 1.

3. Methodology

3.1. Problem Formulation.
Emission can be included in economic dispatch's formulation in a variety of methods. Combined economic and emission dispatch (CEED) is one of the methods that are expressed as a problem of multi-objective optimization used to reduce emission and fuel costs to satisfy demand and avoid losses [38].

3.1.1. Combined Emission and Economy Dispatch (CEED).
CEED problem can be expressed as [39]

$$\sum_{i=1}^{N} P_i - P_l - P_d, \tag{1}$$

$$FC = \sum_{i=1}^{N} a_i P_i^2 + b_i P_i + c_i, \tag{2}$$

$$EC = \sum_{i=1}^{N} a_i P_i^2 + b_i P_i + c_i \mid \text{Cost function}, \tag{3}$$

$$P_l = \sum_{i=1}^{N} \sum_{j=1}^{N} B_{ij} P_j P_i + \sum_{j=1}^{N} B0_i P_i + B_{00}, \tag{4}$$

where P_d represents total load demand, P_l shows total transmission loss, and P_i is the power produced by ith generator. Total fuel cost is denoted by FC, and total emission is denoted by EC. In (2), a_i, b_i, and c_i represent fuel cost coefficients. From (3), the cost function depends on the problem nature, and it can be quadratic, square, sinusoidal, or any other function.

Referred to (4), B_{ij} coefficients or load flow can be utilized to find transmission losses denoted as P_l, where $B0_i$

TABLE 1: Detailed description of related ELD applications concerning different evolutionary approaches.

Sr.	Ref.	Proposed technique	Dataset
1	[7]	BAT algorithm	—
2	[26]	Quantum bat algorithm (QBA)	—
3	[30]	Artificial bee colony algorithm	—
4	[31]	Bat algorithm (BA) and artificial bee colony (ABC) with chaotic-based self-adaptive (CSA) search strategy (CSA-BA-ABC)	23 benchmark function and three CHPED problems
5	[25]	Improved genetic algorithm using novel crossover and mutation (IGA-NCM)	—
6	[32]	Learner nondominated sorting genetic algorithm (NSGA-RL)	10 famous multi-objective functions
7	[33]	Chaotic-crisscross differential evolution (CCDE)	Generalized test functions and two practical hydrothermal system problems
8	[34]	Differential evolution algorithm (DEA)	IEEE-30 bus system
9	[35]	Dynamic economic emission dispatching based on WEV system (WE_DEED)	10 unit systems.
10	[27]	Self-adaptable differential evolution algorithm integrating with multiple mutation strategies (ADE-MMS)	4 DE algorithms are tested on the ten ELD problems with diverse complexities
11	[28]	Differential evolution the algorithm denoted as DEa-AR	IEEE 57-bus system
12	[31]	Modified crow search algorithm (MCSA)	Five different well-known test systems
13	[36]	Multi-objective multi-verse optimization algorithm	140 bus system
14	[24]	Multi-objective economic and environmental dispatch problem (EEDP)	Five generation systems
15	[37]	Coyote optimization algorithm (COA)	Power system consisting thermal generator
16	[29]	Motion optimization algorithm (IMA)	Several cases of different units of thermal plants

is the coefficient vector of B_{ij} and a value B_{00}. The price penalty factor is utilized to transform the problem of multi-objective optimization into a single-objective optimization problem as follows:

$$f(\text{FC, EC}) = \text{Minimize (FC + EC)}. \qquad (5)$$

Each plant or price penalty factor can be found for a specific demand as follows:

(i) The ratio between the average fuel cost and the average emission of the maximum power capacity of that plant is obtained as follows:

$$b_i = \frac{\text{FC}_i(U_i)}{\text{EC}_i(U_i)}, \quad i = 1, 2, n, \qquad (6)$$

where (U_i) is the ith unit of plant capacity.

(ii) Plants are arranged in ascending order based on the value of the price penalty factor.

(iii) Each unit (U_i) maximum capacity is added one at a time, beginning from the lowest value of b_i, unit until $\sum P_i \geq P_d$.

(iv) At this point b_i, linked with the last unit of the process, is the price penalty factor "b," Rs/Kg for the provided demand of the load.

3.1.2. Emission-Controlled Economic Dispatched (ECED).
Emission-controlled economic dispatch (ECED) is another method to reduce the economy related to a specific limit of emission concerning a specific demand. ECED problem's primary concern is to discover the cost-effective placement of plants while fulfilling the losses and demand and keeping the permissible limit of emission; FC needs to be reduced directed to (1), that is, power balance constraint and emission limit constraint expressed as
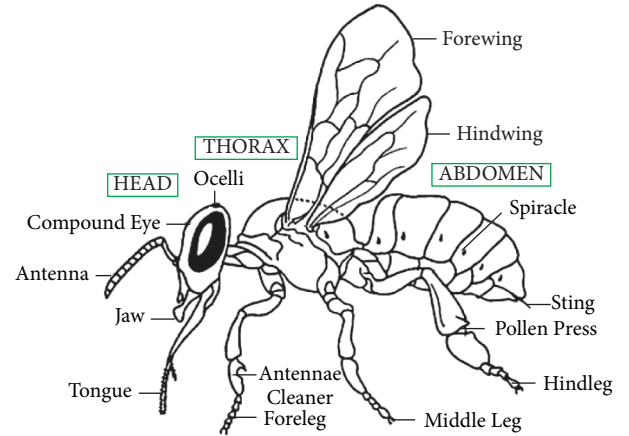
FIGURE 1: Honey bee anatomy [17].

$$f(\text{FC}) = \sum_{i=1}^{N} P_i - P_l - P_d, \; P_l \leq P_i \leq P_d, \quad \text{EC} \leq E_{\text{limit}}. \qquad (7)$$

Here, the system's total emission limit is denoted by E_{limit}.

3.2. Bee Swarming.
This extraordinary insect is one of the most remarkable creatures since the old times. Honeybees have been the topic of scientific research. Moreover, multiple books and experiments have been carried out about honeybees; for instance, Ribbands published "Behavior and the Social Life of Honeybees" in 1953. "Anatomy of the Honey Bee" was written by Snodgrass in 1956, and Thomas D. Seeley wrote "the wisdom of hive" in 1995. The anatomy of a bee is shown in Figure 1. A process known as swarming is carried out to form new honeybee colonies:
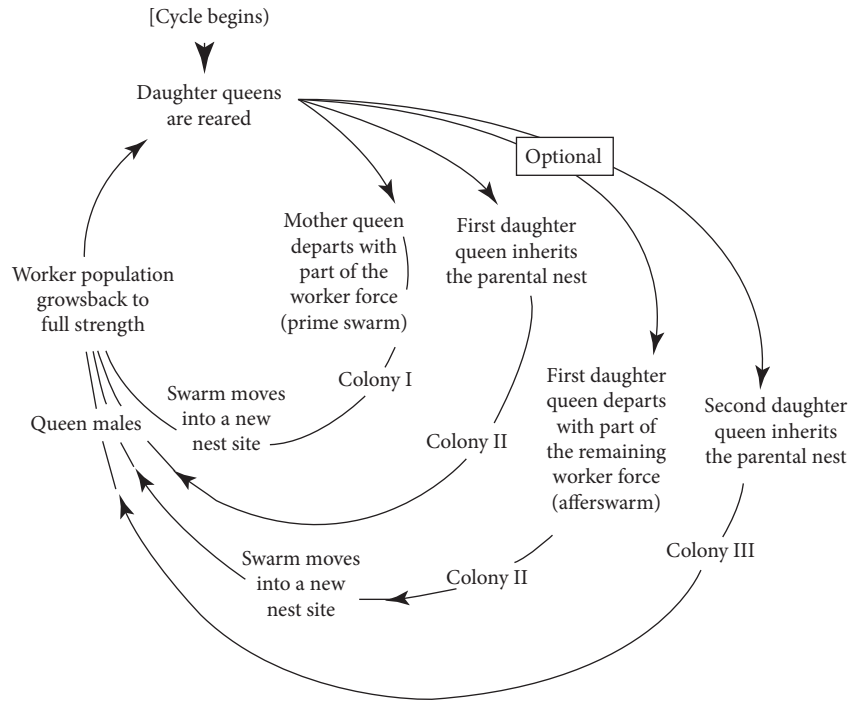
Figure 2: Bee swarming process cycle [17].

(i) The old colony is left by queen bees among some workers and scout bees.

(ii) A swarming cycle is shown in Figure 2. A collection of thousands to tens of thousands of bees make up a swarm.

(iii) They make a cluster around the queen in some branch or a tree, and twenty-fifty scouts are sent out to discover some new proper hives.

(iv) Finally, under the supervision of the scout's bees, all other bees fly to the new hive.

Scout bees check a hive for various standards to meet, such as it must be wide enough to hold the entire swarm, and the entrance must be small and must be at the bottom of the hive, and must get a particular amount of heat from sunlight [40]. Processes of decision-making of scout bees are the source of inspiration. When they find various proper hives, they select the best among them. The source of communication between scouts is the movement of their wings and legs, which is called bee dance. The new hive is selected after the agreement of 80% of the scouts [40].

In terms of algorithms, every hive utilized by a scout demonstrates a feasible solution of an artificial search agent, while the fittest hive expresses a global optimum solution, which is represented in Table 2. The characteristics of the hive, like its size, size of entrance, and location of the entrance, can be viewed as a solution's fitness function. The process of collective decision-making of scouts is expressed as fitness weight (fw) in the algorithm.

3.3. Fitness-Dependent Optimizer.
The reproduction process of swarm bees is replaced by this algorithm. The major

Table 2: FDO-related bee biological characteristics.

Sr.	Nature	Algorithm
1	Selected hive	Global solution
2	Scout collective decision	Objective weight
3	Hive specification	Objective function
4	Hive	Solution found
5	Scout bee	Search agent

portion of the algorithm is obtained from hive exploring the process of scout bees from a pool of suitable options. The algorithm starts with the random initialization of the artificial scout population within the search space of $X_i \, (i = 1, 2, \ldots n)$; the position of every scout expresses a recently recognized hive. Scout bees keep on finding the more suitable hive; once they find a better hive, they neglect the previous better hive; the same is the case with the algorithm. Whenever it discovers a new, more suitable solution than the earlier determined solution is neglected. If they cannot find any other better solution than the previous one, they will consider the current solution as the best solution.

A mechanism of fitness weight and random walk is used to randomly explore the landscape by artificial scouts in this algorithm. The following equation expresses the movement of artificial scout bees:

$$X_{i,t+1} = X_{i,t} + \text{pace}, \tag{8}$$

where i denotes the current search agent, x denotes an artificial scout bee (search agent), pace denotes the direction and movement rate, and t denotes the current iteration of the artificial scout bee. Pace usually depends on fw, that is, fitness weight, whereas pace's direction fully depends on a

random mechanism. Therefore, minimization problems' fw can be measured as

$$fw = \left| \frac{x^*_{i,t\,\text{fitness}}}{x_{i,t\,\text{fitness}}} \right| - wf. \qquad (9)$$

The current best global solution's fitness function value is denoted by $x^*_{i,t\,\text{fitness}}$, current solution's fitness function value is denoted by $x_{i,t\,\text{fitness}}$, and the weight factor is expressed as wf, which can have only 0 or 1 value and is used to control fw. If $wf = 1$, then it shows a low possibility of coverage and a high level of convergence. But if it is equal to 0, then it will not have any effect on equation (3.9), so it can be ignored, and if the variable is $wf = 0$, it will present us a more stable search. However, it reverses as the value of the fitness function entirely depends on the optimization problem. But, the value of fw must be in the range of [0, 1]; still, in some situations, when $fw = 1$, for instance, if the recent solution is the global best solution or global best solution and the recent solution are the same or hold similar fitness value. Furthermore, a possibility exists when $fw = 0$, if $x^*_{i,t\,\text{fitness}} = 0$. Lastly, it must bypass the chances to divide a number with 0 in case $x_{i,t\,\text{fitness}}$, so it must follow the following rules:

$$\left\{ \begin{array}{l} fw = 1 \text{ or } fw = 0 \text{ or } x_{i,t\,\text{fitness}} = 0, \text{ pace} = x_{i,t} * r \\ fw > 0 \text{ and } fw < 1 \left\{ \begin{array}{l} r < 0, \text{ pace} = \left(x_{i,t} - x^*_{i,t} \right) * fw * -1 \\ r \geq 0, \text{ pace} = \left(x_{i,t} - x^*_{i,t} \right) * fw \end{array} \right\} \end{array} \right\}, \qquad (10)$$

where a random number with a range of [−1, 1] is denoted by r. The random walk can be implemented in a variety of ways, but here Levy flight is selected as its good distribution curve offers stable movements [22]. According to FDO mathematical complexity: its time complexity for every iteration is $O(p * n + p * \text{CF})$; here, p denotes the size of the population, problem dimensions are denoted by n, and cost of the objective function is CF. Space complexity for every iteration is $O(p * \text{CF} + p * \text{pace})$; here, pace denotes the best previous paces stored. From this point, the time complexity of FDO is proportional to the number of iterations. However, space complexity will remain identical throughout the sequence of iterations. For the calculation of objective value, FDO owns a simple tool for calculations; it only calculates one random number and fitness weight for every agent [41]. Similarly, DA alignment, attraction, separation, some random values, and distraction are required to be calculated, while a majority of them are accumulative and the value of one depends on the value of others making the calculations complicated [42].

3.3.1. Single-Objective Optimization-Based FDO.
FDO with single-objective optimization problems (FDOSOOPs) starts with the initialization of artificial scouts on random locations of search landscape by utilizing lower and upper boundaries. For each iteration, it selects the global best solution, after that each artificial scout bee is computed by

using (9). Then, the value of fw is examined to decide whether it is 1 or 0 and if $x_{i,t\,\text{fitness}} = 0$. pace is generated by utilizing (10). But, a random number denoted by r of range [−1, 1] will be generated if $fw > 0$ and $fw < 1$. For calculation of *pace*, (10) will be utilized if the value of r is less than 0 and value of fw will have a negative sign; however, for $r \geq 0$, the pace will be calculated with the help of (10) and fw will have a positive sign. Random selection of signs for fw will ensure the random search of artificial bees in all directions.

In FDO, direction and size of pace are controlled by the randomization method; however, only the direction of pace is usually controlled by this method; in such situations, the pace's size depends on fw. Whenever scout bees find a new solution, it is compared with the current solution to determine whether it is better or not based on a fitness function. The earlier solution is neglected if the better latest solution is obtained. Similarly, if it is not better than the previous value of pace, it will be used by the scout bee to continue. On the other hand, if a better solution cannot be achieved by utilizing the previous value of pace, then the current solution will be continued by FDO to the next iteration. In FDO, whenever a solution is acquired, the value is saved for utilization in the next iteration. Two minor alterations are required for the implementation of FDO in maximization problems. (9) should be replaced by (11) as it is the inverse variant of (9):

$$fw = \left| \frac{x_{i,t\,\text{fitness}}}{x^*_{i,t\,\text{fitness}}} \right| - wf. \qquad (11)$$

Then, the criteria for the selection of the best solution must be altered. The statement "if $(X_{t+1,i}\text{fitness} < X_{t,i}\text{fitness})$" needs to be replenished with "if $(X_{t+1,i}\text{fitness} > X_{t,i}\text{fitness})$."

3.3.2. Multi-Objective Optimization-Based FDO.
FDO with multi-objective optimization problems (FDOMOOPs) begins with the initialization of artificial scouts into two-dimensional search space (X_i, Y_i). Each scout bee in the search space of (X_i, Y_i) can be defined as $X_i (i = 1, 2, \ldots n)$ and $Y_i (i = 1, 2, \ldots n)$. Then, the value of fw is examined to decide whether its 1 or 0 and if $x_{i,t\,\text{fitness}} = 0$ or $y_{i,t\,\text{fitness}} = 0$. In both cases, the *pace* can be generated as $fw = |x_{i,t\,\text{fitness}}/x^*_{i,t\,\text{fitness}}| - wf$ and $fw = |y_{i,t\,\text{fitness}}/y^*_{i,t\,\text{fitness}}| - wf$.

3.4. Enhanced Method.
Multidimensional and multi-objective optimization algorithms tend to perform better for solving the linear and nonlinear constraint problems than the single-dimensional and single-objective optimization algorithms, especially in the case of ELD, when both the emission rate and fuel cost need to be minimized to approach the total loss and power demand. FDO is a multi-objective metaheuristic algorithm and, therefore, best suitable for solving constrained ELD problems. It can be implied to complex problems with nonlinear approximation. In this thesis, the authors have carried out FDO to solve the ELD

problem by minimizing fuel cost, emission allocation, and transmission loss. Besides, the authors have employed a novel variant of FDO, which incorporates novel population initialization techniques and employed sine maps to select the weight factor for FDO dynamically.

3.4.1. Population Initialization.

The swarm or group of swarms needs to be fired randomly to obtain their initial fitness solution in the optimization process. The entire process is called population initialization. The most conventional method to assign an initial location to each individual is through a random number generator following the normal distribution. However, the major drawback to using the random number generator is premature convergence and abnormal exploration and exploitation. The random number generator developed a random number between the internal of 0 and 1. The probability of obtaining an optimal solution in the case of local minima is reduced when the initial locations are directed far from the solution, and each individual requires more steps and iterations to seek the entire solution. The swarm can be stuck into premature convergence during the searching process and lead to poor exploration. Similarly, as opposed to this, the probability of obtaining a global solution in the case of global minima is diminished when the primary positions are delivered too near around the search space while the solution is out of search space; hence, each individual requires more steps and iterations to seek the entire solution and can be stuck into the premature convergence and leads to poor exploitation.

3.4.2. Quasi-Random Sequence Initialization.

Quasi-random is a distinction of n-rows that occupies n-dimensional search space. It is also called a low-disparity sequence. However, the usual standard quasi-random sequences and odd numbers all give consistently suitable sequences. There is a significant distinction between these two patterns, aside from the standardized manner. An identical arbitrary generator on (0, 1) will deliver sequences, so every preliminary has a similar likelihood of producing a point on equivalent subintervals, for instance [(0, 1/2), (1, 1/2)]. In this manner, it is attainable for n preliminaries to inadvertently all extend in the top half of the range, while the $(n + 1)$ points fall inside the other of the two parts with a likelihood of 1/2. While this is not the situation with the quasi-random sequences, the generated sequences are obliged by a low-inconsistency prerequisite that has a net impact on centers being created in a profoundly connected way. To avoid the premature convergence problem in FDO, the authors have carried out one of the quasi-random sequences called the Sobol sequence for the population's initialization.

3.4.3. Quasi-Random Sequence Initialization.

Sobol sequence is a low discrepancy sequence that was first enhanced by mathematicians in Russia in 1967 [43]. It mimics the

random distribution by appropriating a base of two to shape progressively better uniform edges of the required interval and afterward reorder the directions in each measurement. Following are prime steps to generate the Sobol sequences S^d:

(i) Let S^d be the hypercube with the interval of $S^d = [0, 1]^d$ and d-dimensional. The approximation function f^{opr} is integrated over the hypercube S^d.

(ii) The Sobol sequence termed as Sobol $[x, y]$ can be generated using the following equation over the nonlinear approximation of S^d:

$$\lim_{x \to \infty} \frac{1}{x} \sum_{i=1}^{x} f(S_i) = \int_{S^d}^{i} f. \tag{12}$$

(iii) It is a notable pattern against each dimensional vector that for the whole to reach towards the indispensable points S^d. Furthermore, the second great feature would be that the forecasts of x in the low range of the dimensioned face of S^d cover most of the search area in terms of optimization.

(iv) Subsequently, the comparable center of S^d does not meet the criteria because in lower measurements, numerous focuses will be at a similar spot, in this way unnecessary for the vital estimation.

The comparison of FDO population initialization with random numbers following the Sobol distribution and the uniform distribution is presented in Figures 3 and 4 respectively. The following equation is used in the standard FDO for the swarm to select their initial locations to seek the entire optimal solution:

$$\text{Random}_i (i = 1, 2, \ldots n) \therefore [0, 1]. \tag{13}$$

In the enhanced FDO, the authors have selected the interval of [0, 1] for generating both sequences uniform and Sobol sequences in the process of FDO population initialization. The enhanced equation for initializing the swarm in FDO is presented as follows:

$$\text{Sobol}_i (i = 1, 2, \ldots n) \therefore [0, 1], \tag{14}$$

where [0, 1] in (13) and (14) represents the standard limits of both generated sequences. The uniform random positions can be seen in Figure 3 with very random locations and ill-patterned sequences, which may lead to poor exploitation. As compared to the uniform random, the Sobol sequence comes up with a well-patterned sequence in Figure 4, which may lead the swarm to converge maturely.

3.4.4. Enhanced Approach for Updating Weight Factor.

The weight factor is revealed as wf, which can have particularly 0 or 1 utility and is utilized to control wf. If wf belongs to absolute 1, then it confers a low probability of coverage and a high level of convergence. However, if the weight factor wf belongs to 0, then it will not have any influence on (15), so it can be ignored; if the variable is $wf = 0$, it will present us a
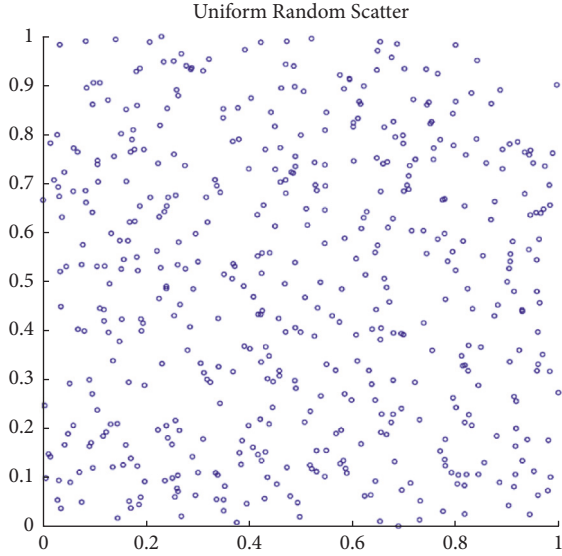
FIGURE 3: Population initialization with random number generator following the uniform sequence.
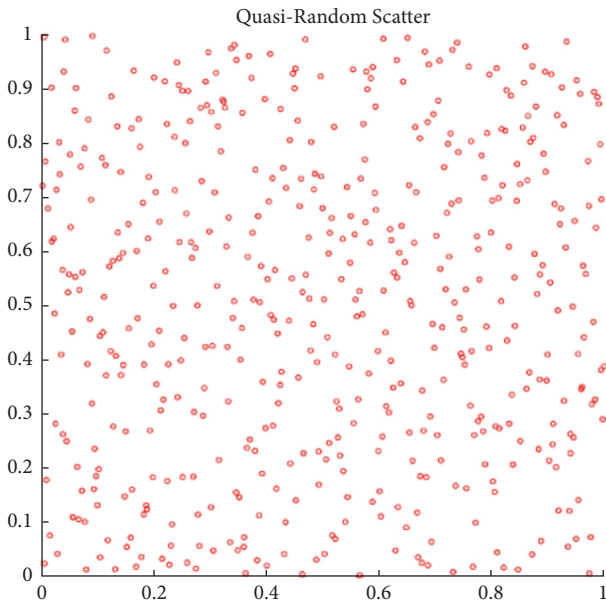
FIGURE 4: Population initialization with Sobol sequence following the random distribution.

more stable search. However, it reverses as the value of the fitness function entirely depends on the optimization problem:

$$fw = \left|\frac{x^*_{i,t \text{ fitness}}}{x_{i,t \text{ fitness}}}\right| - 0. \tag{15}$$

However, in unusual circumstances, when the weight factor belongs to absolute one as shown in (16), for instance, if the new solution is the global best solution or local best solution and the new solution is identical or operates similar fitness value:

$$fw = \left|\frac{x^*_{i,t \text{ fitness}}}{x_{i,t \text{ fitness}}}\right| - 1, \tag{16}$$

wf should be balanced enough to control the exploitation and exploration for the controlled convergence rate when leads to absolute 0 wf and absolute 1 wf. For this, the authors produce a chaotic effect by using sine maps to the weight factor wf between the interval of [0, 1].

3.4.5. Chaotic Sine Map. Chaotic maps produce uncontrolled groupings during the metaheuristic algorithm. The authors practiced the benefit of a chaotic sine pattern to update the weight factor [22]. The sine map is chaotic and used to produce a quarter effect between the interval of 0 and 1. When the weight factor becomes skewed towards 0, the sine wave covers the low balance and controls the low convergence rate. Similarly, when the weight factor becomes skewed towards 1, the sine wave covers the high balance and controls the high convergence rate. This phenomenon iteratively maintains the balance with optimal weight factor throughout the last epoch. A chaotic sine map can be defined as

$$S_{\text{map}} = \frac{m}{4}\sin(\pi x_i), \tag{17}$$

where $0 < m < 4$ is the controlling factor. The author chooses $m = 0.3$ with the most optimal sequence. In terms of the weight factor, the equation becomes

$$w_s = \frac{m}{4}\sin(\pi wf). \tag{18}$$

The enhanced variant of FDO utilized the following equation to update the fitness weight fw:

$$fw = \left|\frac{x^*_{i,t \text{ fitness}}}{x_{i,t \text{ fitness}}}\right| - w_s. \tag{19}$$

The flowchart of the enhanced FDO along the ELD application is presented in Figure 5.

4. Application Results

4.1. Dataset Overview. The performance of the enhanced variant of FDO is evaluated through 24 units taken from the 18-unit system and 20-unit system with each of the 6-unit case study chunks by optimizing the fitness function enlisted in (1). The parameter sets used in the experiment for each unit are listed below.

Total number of units used in the experiment = 24, total power demand = 400, 700, number of iterations = 100, 200, population size = number of bee scouts = 50, and the beta coefficient used for 24 units according to each chunk of 6 units in the exploring capacity with a power demand of 400 MW and 700 MW are presented as follows:

$$\text{Beta coefficient} = 1e^{-4} \times \begin{bmatrix} 1.4.17.15.17.60.13 \\ .15.13.65.19.16.17 \\ .26.15.24.22.20.19 \\ .19.26.22.16.15.20 \\ .17.24.19.71.30.25 \\ .30.69.32.25.32.85 \end{bmatrix}. \quad (20)$$

Detail of 24 units used to minimize the fuel cost, emission allocation, and transmission loss with 400 and 700 power load is presented in Table 2. In Table 2, Pmin and Pmax represent the lower and upper plant limits, respectively, whereas other parameters can be defined as $a =$ (.

The dataset used for the simulation consists of two chunks with 12 generating thermal units each. The first 12 generating units (1 to 12) are taken from Sys_18 U with all plant limitations and beta coefficients as represented in (20).

Similarly, the last 12 generating thermal units (13 to 24) are taken from Sys_20 U with all plant limitations and beta coefficients as described in (20). All thermal units employed for the empirical analysis are ramp-limits-free and do not endure in the prohibited zone for the smooth objective function. Comparison of simulation results on the ELD problem (FDO vs. enhanced FDO) with nonlinear optimization on 100 epochs with a power demand of 400 and 700 is presented in Tables 3 and 4. The total fuel error and the transmission cost are the minimum global fitness achieved by optimizing the ELD problem as minimize f (FC, EC), $\ni$, $\sum_{i=1}^{N} P_i = P_d + P_l$, $L_i \le P_i \le U_i$. Similarly, in Tables 4 and 5, a comparison of simulation results on the ELD problem (FDO vs. enhanced FDO) with nonlinear optimization on 200 epochs with a power demand of 400 and 700 is presented.

Certainly, for our problem statement, the accentuation is to distinguish, which epoch setting requires the most minimal fuel cost and transmission loss to discover arrangements of a specific worthy quality. Furthermore, the power demand is also analyzed in a roundabout way, to give in any event a complex reflection of the complexities of the various calculations considered in our relative examination.

The obtained results certainly take the fact that the population initialization and optimal fitness factor of FDO make some impact on the global best of FDO as compared to the enhanced variant of FDO in terms of optimal allocation emission. The reason behind developing each thermal generation chunk with 6 units is to investigate the impact of greater dispersion on the total fuel cost and minimum error. To visualize the obtained results emission allocation results, a convergence comparison of FDO with the enhanced variant of FDO on the first 6 thermal units with 100 epochs and different power demands are illustrated in Figure 6.

Moreover, convergence comparison (transmission loss) of FDO with the enhanced variant of FDO on the 24 thermal units with (100, 200 epochs) and different power demands are demonstrated in Figures 7 and 8. To validate the achieved results on the ELD problem (FDO vs. enhanced FDO), this research used an ANOVA test. The main reason behind performing the ANOVA test is to find the significant difference between the standard and enhanced FDO in terms of minimization. The additional reason to perform the ANOVA test is to determine which parameter delivers outcomes with critical contrasts, considering the target value accomplished by enhanced FDO from each run of the considerable number of tests performed. Graphical representation of one-way ANOVA test comparison (optimal allocation emission) of FDO with the enhanced variant of FDO on the 24 thermal units with (100, 200 epochs) and several power demands are illustrated in Figures 8 and 9, respectively.

4.2. Optimal Allocation Emission. It is earlier mentioned that each chunk of thermal units is tested on 100 independent runs with 100 and 200 epochs considering two different combinations of power demand. The authors observed a significant improvement in the optimal power allocation generated by the enhanced FDO for the first 6 thermal units as compared to the conventional FDO (referred to Table 4). All 5 units' results obtained by enhanced FDO outperformed FDO except the 6th unit with 162.2504561 optimal emission allocation on 100 epochs and 400 power demand.

As contrasted to the first chunk of thermal units, the performance of the enhanced algorithm was observed less when optimizing emission allocation. It can be seen from Table 4 that only thermal units 7 and 12 gained better emission rates, which lead to greater divergence of the whole population. However, thermal units 8, 9, 10, and 11 show equal empirical performance for both FDO and enhanced FDO with a 44.3 emission rate.

In the case of the third chunk, thermal units 13 to 18, the emission allocation rate is not significantly improved using enhanced FDO instead of the standard FDO. The enhanced version obtained 18 with unit 13, 58 with units 14, 15, 16, 17, and 58 with unit 18, on 100 epochs and 400 power demand, which shows a slight improvement. This slight impact of enhanced FDO reveals the impact of robust population initialization on the ELD emission allocation. Lastly, the fourth chunk of the thermal unit from Table 4 exhibits outstanding results of the Enhanced algorithm on the entire parameter setting except for the last thermal unit with a 19.62533477 emission rate.

From Table 5, when power demand raised 400 to 700, the enhanced algorithm also improves the swarm convergence, and hence, the optimal fitness factor works here. This phenomenon shows the inverse divergence of the global best computed with the enhanced fitness factor, which leads to the emission allocation of the thermal units 3 and 6 from 68.72201216, 162.2504561 to 83.36202838, and 347.0304788, respectively. Similarly, premature convergence is highly tackled by the enhanced algorithm when seeing a significant decrease in the average optimal allocation for the first six thermal units. Tables 5 and 6 show similar convergence behavior of enhanced FDO compared to the FDO using 200 epochs with 400 and 700 power demand. However, the optimal fitness factor produces less impact than the effect produced when testing on 100 epochs. This can be due to the dimension reduction that occurs in higher generations.
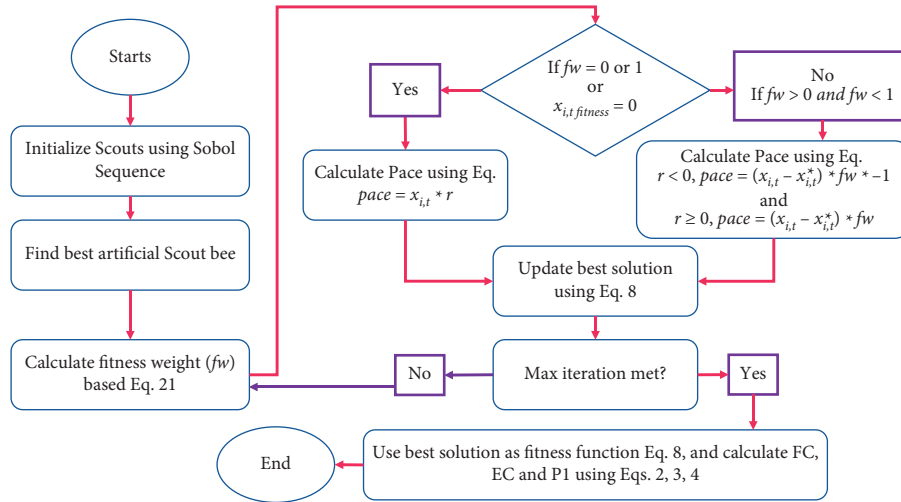
FIGURE 5: Flowchart for the enhanced FDO algorithm along with ELD application.

TABLE 3: Twenty-four units used with a chunk of 6 units in the exploring capacity with a power demand of 400 MW and 700 MW.

Units	Pmin	Pmax	a	B	C
1	7	15	0.602842	22.45526	85.74158
2	7	45	0.602842	22.45526	85.74158
3	13	25	0.214263	22.52789	108.9837
4	16	25	0.077837	26.75263	49.06263
5	16	25	0.077837	26.75263	49.06263
6	3	14.75	0.734763	80.39345	677.73
7	3	14.75	0.734763	80.39345	677.73
8	3	12.28	0.514474	13.19474	44.39
9	3	12.28	0.514474	13.19474	44.39
10	3	12.28	0.514474	13.19474	44.39
11	3	12.28	0.514474	13.19474	44.39
12	3	24	0.657079	56.70947	574.9603
13	150	600	0.00068	18.19	1000
14	50	200	0.00071	19.26	970
15	50	200	0.0065	19.8	600
16	50	200	0.005	19.1	700
17	50	160	0.00738	18.1	420
18	20	100	0.00612	19.26	360
19	25	125	0.0079	17.14	490
20	50	150	0.00813	18.92	660
21	50	200	0.00522	18.27	765
22	30	150	0.00573	18.92	770
23	100	300	0.0048	16.69	800
24	150	500	0.0031	16.76	970

TABLE 4: Comparison of simulation results on the ELD problem (FDO vs. enhanced FDO) with nonlinear optimization on 100 epochs and 400 power demand. The optimal values are exhibited in boldface.

Units	Power demand = 400	
	Optimal allocation emission (*lbs*)	
	FDO	Enhanced FDO
1	70.44063664	**69.91458228**
2	69.28036315	**68.72201216**
3	38.43849912	**37.87285707**
4	31.18554733	**30.67657178**
5	31.07224457	**30.56379958**
6	160.2587123	162.2504561
Total fuel cost ($)	2.05E + 05	**2.04E + 05**
Transmission loss	0.676	**2.79E-04**
7	111.6951678	**111.360789**
8	44.39	44.39
9	44.39	44.39
10	44.39	44.39
11	44.39	44.39
12	111.431507	**111.0794924**
Total fuel cost ($)	1.05E + 05	**1.04E + 05**
Transmission loss	0.6867	**2.81E-04**
13	18.32981341	**18.24689868**
14	59.0172677	**58.85809285**
15	58.93086524	**58.77200905**
16	58.7043227	**58.5458324**
17	58.92968203	**58.77222232**
18	146.7176828	146.8052033
Total fuel cost	1.25E + 06	**1.24E + 06**
Transmission loss	0.6296	**2.59E-04**
19	122.1815342	**122.1372312**
20	62.6886485	**62.53294149**
21	62.13776388	**61.98118821**
22	103.4629522	**103.3564131**
23	30.48017867	**30.36716223**
24	19.70886257	**19.62533477**
Total fuel cost ($)	1.31E + 06	**1.30E + 06**
Transmission loss	0.6599	**2.71E-04**

TABLE 5: Comparison of simulation results on the ELD problem (FDO vs. enhanced FDO) with nonlinear optimization on 100 epochs and 700 power demand. The optimal values are exhibited in boldface.

Units	Power demand = 700	
	Optimal allocation emission (*lbs*)	
	FDO	Enhanced FDO
1	85.7416	**85.74158**
2	85.7416	**85.74158**
3	108.9837	**83.36202838**
4	49.0626	49.06263
5	49.0626	49.06263
6	525.6109	**347.0304788**
Total fuel cost ($)	6.51E + 05	**6.46E + 05**
Transmission loss	2.2609	**9.27E-04**
7	259.65771	**258.266736**
8	44.39	44.39
9	44.39	44.39
10	44.39	44.39
11	44.39	44.39
12	265.4517495	**264.1743532**
Total fuel cost ($)	4.50E + 05	**4.45E + 05**
Transmission loss	2.6695	**0.0011**
13	33.35092403	**33.16801208**
14	104.2882238	**103.890753**
15	104.0208305	**103.6249588**
16	103.3245773	**102.9315033**
17	103.1925346	**102.8020805**
18	253.7447758	**253.5834798**
Total fuel cost ($)	3.73E + 06	**3.71E + 06**
Transmission loss	1.9219	**7.88E-04**
19	210.6754314	**210.3359493**
20	110.5273588	**110.1212242**
21	109.860002	**109.4539346**
22	181.2722755	**180.8655902**
23	54.11026706	**53.84868486**
24	35.56430084	**35.37544006**
Total fuel cost ($)	3.92E + 06	**3.90E + 06**
Transmission loss	2.0096	**8.23E-04**

4.3. Fuel Cost. Fuel cost minimization on the same hydroenergy and power demand is a big issue when several units are mimicking in parallel. This can be optimized by considering the current fuel cost as the global best for each of the individuals in the FDO. However, the minimal risk is premature convergence, which leads to double computing cases and wastage of time with the cost approximately equal to the standard fuel cost. Robust population initialization decreases the chance of premature convergence. Hence, the enhanced FDO used optimal fitness factors in combination with the Sobol operator to minimize the fuel cost.

It can also be observed in Tables 3–5 that the fuel cost difference between FDO and enhanced FDO is notable for all 24 thermal units on 100 and 200 epochs with a power demand of 400 and 700 sequentially. However, this significant minimal difference in cost can impact the whole unit generation cost. The trends for each chunk of the thermal unit from Tables 4–6 confirm the directly proportional relationship between the power demand and the fuel cost

directly. The minimization range is constant between them, which explicates the strong divergence and influential fitness factor. Increasing power demand will lead the fuel cost to increase with a constant proportion of difference produced by FDO and enhanced FDO.

4.4. Transmission Loss. Transmission error is essential for significant distance potential transmission, and it grows with an expansion in the measure of capacity to be dispatched. Therefore, the utilization of inexhaustible force from the sustainable plants close to the heap focuses diminishes the transmission losses. Appropriating the sustainable power sources all through the working time frames as opposed to utilizing them during their accessible period will assist with diminishing both expense and the transmission loss.

Compared to the optimal emission allocation and fuel cost, the authors have received the best optimal transmission loss results. The enhanced algorithm FDO reduced the loss

FIGURE 6: Convergence comparison (optimal allocation emission) of FDO with the enhanced variant of FDO on the first 6 thermal units with 100 epochs and different power demands.

FIGURE 7: Convergence comparison (transmission loss) of FDO with the enhanced variant of FDO on the 24 thermal units with 100 epochs and different power demands.

FIGURE 8: Convergence comparison (transmission loss) of FDO with the enhanced variant of FDO on the 24 thermal units with 200 epochs and different power demands.

FIGURE 9: One-way ANOVA test comparison (optimal allocation emission) of FDO with the enhanced variant of FDO on the 24 thermal units with 100 epochs and different power demands.

with a 60% rate on average. The authors can perceive that in Table 4, FDO minimizes the loss to 0.676, 0.6867, 0.6296, and 0.6296 for four chunks of the thermal unit with 100 epochs and 400 power demand as compared to the enhanced FDO, which reduces it to 2.79E-04, 2.81E-04, 2.59E-04, and 2.71E-

04 for four chunks of the thermal unit with 100 epochs and 400 power demand sequentially. Enhanced DFO significantly outperformed standard FDO for minimization transmission loss.

Likewise, the enhanced algorithm FDO decreased the loss by a 300% rate regularly. The study can comprehend that in Table 5, FDO minimizes the loss to 2.2609, 2.6695, 1.9219, and 2.0096 for four chunks of the thermal unit with 100 epochs and 700 power demand as contrasted to the intended

TABLE 6: Comparison of simulation results on the ELD problem (FDO vs. enhanced FDO) with nonlinear optimization on 200 epochs and 400 power demand. The optimal values are exhibited in boldface.

Units	Power demand = 400	
	Optimal allocation emission (*lbs*)	
	FDO	Enhanced FDO
1	69.91436113	**69.91435917**
2	68.72177747	**68.72177539**
3	37.87262247	**37.87262039**
4	30.67636117	**30.67635931**
5	30.56358919	**30.56358733**
6	162.251291	162.2512984
Total fuel cost ($)	2.04E + 05	2.04E + 05
Transmission loss	2.45E-06	**2.69E-12**
7	111.3606533	**111.360652**
8	44.39	44.39
9	44.39	44.39
10	44.39	44.39
11	44.39	44.39
12	111.0793492	**111.079348**
Total fuel cost ($)	1.04E + 05	1.04E + 05
Transmission loss	2.47E-06	**2.72E-12**
13	18.24686486	**18.24686456**
14	58.85802787	**58.85802729**
15	58.7719442	**58.77194362**
16	58.5457677	**58.54576712**
17	58.77215804	**58.77215747**
18	146.8052396	146.8052399
Total fuel cost ($)	1.24E + 06	1.24E + 06
Transmission loss	2.27E-06	**2.50E-12**
19	122.1372134	**122.1372132**
20	62.53287799	**62.53287743**
21	61.98112435	**61.98112378**
22	103.3563698	**103.3563694**
23	30.36711614	**30.36711573**
24	19.62530071	**19.62530041**
Total fuel cost ($)	1.30E + 06	1.30E + 06
Transmission loss	2.38E-06	**2.62E-12**

TABLE 7: Comparison of simulation results on the ELD problem (FDO vs. enhanced FDO) with nonlinear optimization on 200 epochs and 400 power demand. The optimal values are exhibited in boldface.

Units	Power demand = 700	
	Optimal allocation emission (*lbs*)	
	FDO	Enhanced FDO
1	85.74158	85.74158
2	85.74158	85.74158
3	83.36112205	**83.36111401**
4	49.06263	49.06263
5	49.06263	49.06263
6	347.0304661	**347.030466**
Total fuel cost ($)	6.46E + 05	6.46E + 05
Transmission loss	8.15E-06	**8.95E-12**
7	258.266172	**258.266167**
8	44.39	44.39
9	44.39	44.39
10	44.39	44.39
11	44.39	44.39
12	264.1738376	**264.173833**
Total fuel cost ($)	4.45E + 05	4.45E + 05
Transmission loss	9.57E-06	**1.05E-11**
13	33.16793762	**33.16793696**
14	103.8905912	**103.8905898**
15	103.6247976	**103.6247962**
16	102.9313433	**102.9313419**
17	102.8019216	**102.8019202**
18	253.5834156	**253.583415**
Total fuel cost ($)	3.71E + 06	3.71E + 06
Transmission loss	6.92E-06	**7.60E-12**
19	210.3358119	**210.3358107**
20	110.121059	**110.1210576**
21	109.4537694	**109.453768**
22	180.8654252	**180.8654237**
23	53.84857844	**53.8485775**
24	35.37536322	**35.37536254**
Total fuel cost ($)	3.90E + 06	3.90E + 06
Transmission loss	7.23E-06	**7.94E-12**

FDO, which overcome it to 9.27E-04, 0.0011, 7.88E-04, and 8.23E-04 for four chunks of the thermal unit with 100 epochs and 700 power demand sequentially.

Tables 6 and 7 explicitly formulate the same trends between standard and enhanced FDO with 200 significant differences in transmission error on 200 epochs and 400 and 700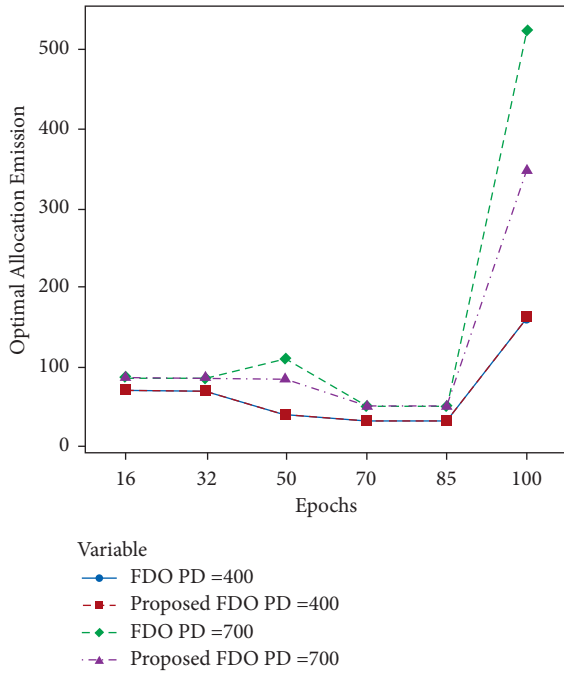 power demand. Figure 6 confirms the optimal convergence comparison in the case of optimal allocation emission (FDO with the enhanced variant of FDO on the first six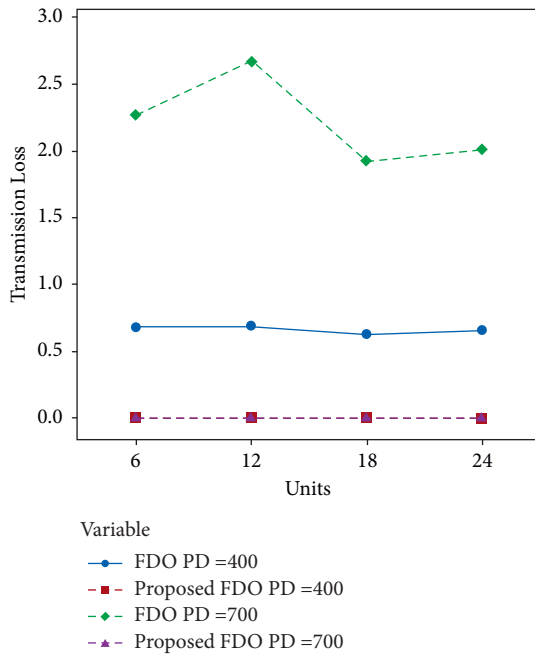 thermal units with 100 epochs and different power demands). Figures 7 and 8 dispense the clear-cut transmission loss difference. Units are presented on the *X*-axis, while transmission loss is enlisted on *Y*-axis.

To validate the obtained results, ANOVA statistical analysis for Figures 9 and 10 reinforces the best performance of the enhanced FDO algorithm and encourages the solution for other constraints as well. The box representation for enhanced FDO with 400 power demand is proved as an optimal solution with optimal chunk. Similarly, the interval plot representation for enhanced FDO with 700 power

demand is determined as an optimal solution with an optimal chunk.

5. Conclusion

This research work introduced a variant of the FDO algorithm motivated by scout bees in the hive exploring the process of seeking food from a pool of suitable options. The enhanced variant is utilized to solve the economic load dispatch problem. FDO and its modified version are motivated to upgrade the minimization capability during weight optimization of economic load dispatch. Each individual of the scout bee is represented as output power generated through each thermal unit. The study deals with three types of constraints in this work: power balance capacity, transmission loss, and optimal emission allocation. In the beginning, the exploration executed by enhanced FDO is dependent on a simplistic fitness factor that delivers a less optimal solution by sticking into local minima and transforms some of its decision variables through their constraint
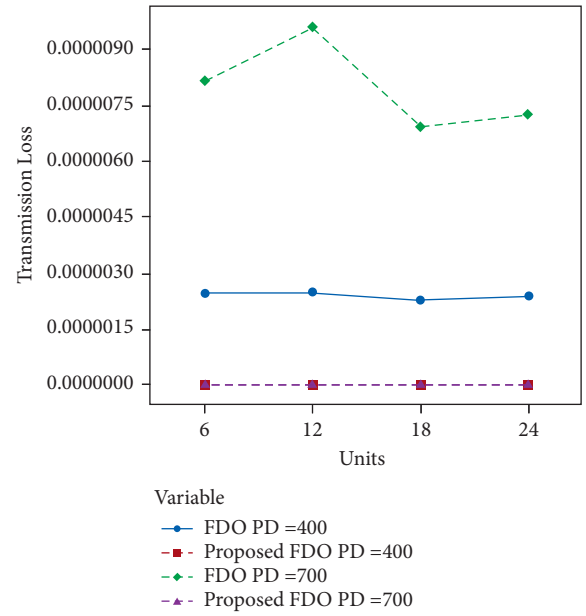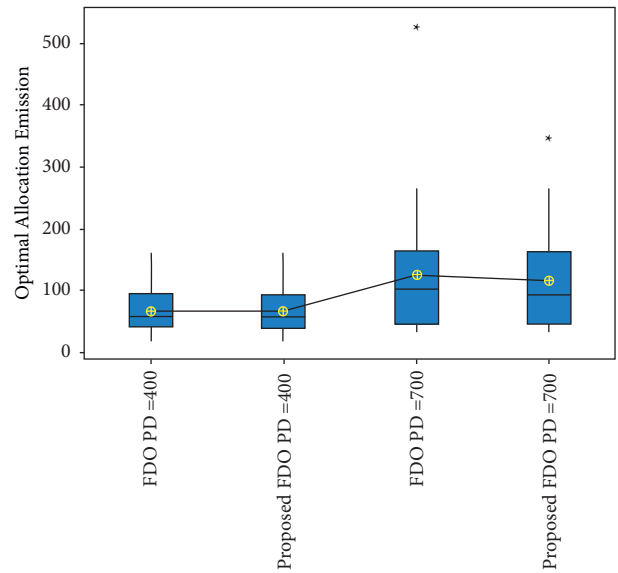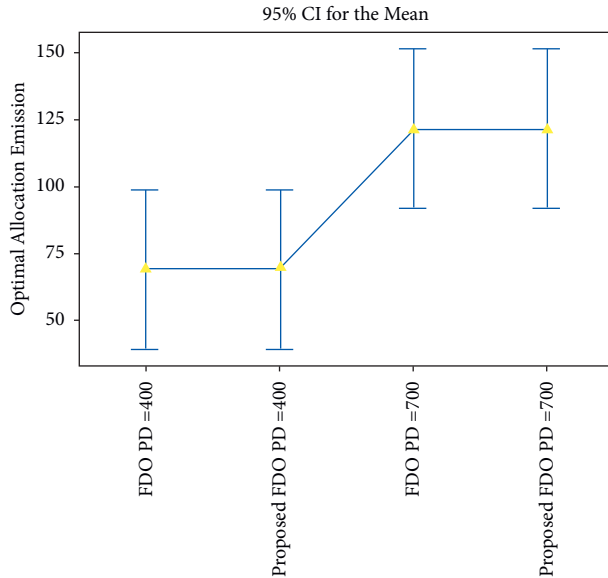
FIGURE 10: One-way ANOVA test comparison (optimal allocation emission) of FDO with the enhanced variant of FDO on the 24 thermal units with 200 epochs and different power demands.

violation. After applying the Sobol operator for population initialization and chaotic sine map for the optimal fitness, redistribution power operators are connected. The enhanced operator ensures the feasibility of a probable solution that the thermal unit will take as an input and barely estimate the balance power constraint. Furthermore, the enhanced population initialization approach consolidates a quasi-random Sabol sequence to create the initial solution in the multidimensional search space. A regular 24-unit system is applied with diverse power demands for experimental evaluation. The experiential results acquired utilizing the enhanced variant of FDO confirm the superior performance in terms of low transmission loss, low fuel cost, and low emission allocation compared to the standard FDO. As a part of our future work, the authors are inspired by the hybridization of FDO with other metaheuristic algorithms such as BA, DE, and PSO. The authors aimed at taking the best qualities from BA as local search capability, DE as optimal mutation factor, and PSO as inertia weight and incorporating them in FDO to achieve the best results. Furthermore, the authors are also interested in the fine-tuning of FDO parameters in combination with ELD constraint and their hyperparameter tuning. Additionally, the hybridized version of FDO will be evaluated to investigate the influence of objective evaluations on dimension reduction.

Abbreviations

ELD: Economic load dispatch
POZ: Prohibited operating zone
VPE: Valve-point effects
QBA: Quantum bat algorithm
CEED: Combined economic emission dispatch

ADE-MMS: Evolution algorithm integrating with multiple mutation strategies
WEV: EVs and wind farms
WE_DEED: Dynamic economic emission dispatching based on WEV system
IMA: Motion optimization algorithm
BA: Bat algorithm
ABC: Artificial bee colony
CSA: Chaotic-based self-adaptive
NSGA-RL: Learner nondominated sorting genetic algorithm
CCDE: Chaotic-crisscross differential evolution
DEA: Differential evolution algorithm
ADE-MMS: Self-adaptable differential evolution algorithm integrating with multiple mutation strategies
MCSA: Modified crow search algorithm
EEDP: Multi-objective economic and environmental dispatch problem
COA: Coyote optimization algorithm
CEED: Combined economic and emission dispatch
ECED: Emission-controlled economic dispatch
FDO: Fitness-dependent optimizer
FDOSOOPs: FDO with single-objective optimization problems
FDOMOOPs: Multi-objective optimization problems.

Data Availability

Data can be shared upon request from the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] T. Sale, "Alan turing at bletchley park in world war II," in *Alan Turing: Life and Legacy of a Great Thinker*, pp. 441–462, Springer, Berlin, Germany, 2004.

[2] G. G. Wang, S. Deb, and L. D. S. Coelho, "Earthworm optimisation algorithm: a bio-inspired metaheuristic algorithm for global optimisation problems," *International Journal of Bio-Inspired Computation*, vol. 12, no. 1, p. 1, 2018.

[3] A. Sheta, H. Faris, M. Braik, and S. Mirjalili, "Nature-Inspired Metaheuristics Search Algorithms for Solving the Economic Load Dispatch Problem of Power System: A Comparison Study," *Applied Nature-Inspired Computing: Algorithms and Case Studies*, pp. 199–230, Springer, Singapore, 2019.

[4] M. Ali, R. C. Deo, Y. Xiang et al., "Coupled online sequential extreme learning machine model with ant colony optimization algorithm for wheat yield prediction," *Scientific Reports*, vol. 12, no. 1, p. 5488, 2022.

[5] I. Ahmadianfar, S. Shirvani-Hosseini, A. Samadi-Koucheksaraee, and Z. M. Yaseen, "Surface water sodium (Na+) concentration prediction using hybrid weighted exponential regression model with gradient-based optimization," *Environmental Science and Pollution Research International*, pp. 1–26, 2022.

[6] M. Alrashidi, S. Rahman, and M. Pipattanasomporn, "Metaheuristic optimization algorithms to estimate statistical distribution parameters for characterizing wind speeds," *Renewable Energy*, vol. 149, pp. 664–681, 2020.

[7] L. A. Wulandhari, S. Komsiyah, and W. Wicaksono, "Bat algorithm implementation on economic dispatch optimization problem," *Procedia Computer Science*, vol. 135, pp. 275–282, 2018.

[8] A. A. Ewees, M. A. Gaheen, Z. M. Yaseen, and R. M. Ghoniem, "Grasshopper Optimization Algorithm with Crossover Operators for Feature Selection and Solving Engineering Problems," *IEEE Access*, vol. 10, 2022.

[9] I. Ahmadianfar, S. Shirvani-Hosseini, J. He, A. Samadi-Koucheksaraee, and Z. M. Yaseen, "An improved adaptive neuro fuzzy inference system model using conjoined metaheuristic algorithms for electrical conductivity prediction," *Scientific Reports*, vol. 12, no. 1, p. 4934, 2022.

[10] N. Mahmoudi, A. Majidi, M. Jamei et al., "Mutating fuzzy logic model with various rigorous meta-heuristic algorithms for soil moisture content estimation," *Agricultural Water Management*, vol. 261, Article ID 107342, 2022.

[11] I. Fister, X. S. Yang, J. Brest, and D. Fister, "A Brief Review of Nature-Inspired Algorithms for Optimization," *Elektrotehniski Vestnik/Electrotechnical Review*, Article ID 20635, 2013, https://arxiv.org/abs/1307.4186.

[12] J. Yuste, A. Duarte, and E. G. Pardo, "An efficient heuristic algorithm for software module clustering optimization," *Journal of Systems and Software*, vol. 190, 2022.

[13] V. Raman and N. S. Gill, "Review of different heuristic algorithms for solving Travelling Salesman Problem," *International Journal of Advanced Research in Computer Science*, vol. 8, no. 5, 2017.

[14] G. Dhiman and V. Kumar, "Seagull optimization algorithm: theory and its applications for large-scale industrial engineering problems," *Knowledge-Based Systems*, vol. 165, pp. 169–196, 2019.

[15] G. Dhiman and A. Kaur, "STOA: a bio-inspired based optimization algorithm for industrial engineering problems," *Engineering Applications of Artificial Intelligence*, vol. 82, pp. 148–174, 2019.

[16] R. Storn and K. Price, "Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces," *Journal of Global Optimization*, vol. 11, no. 4, pp. 341–359, 1997.

[17] J. M. Abdullah and T. Ahmed, "Fitness dependent optimizer: inspired by the bee swarming reproductive process," *IEEE Access*, vol. 7, Article ID 43486, 2019.

[18] W.-K. Hao, J.-S. Wang, X.-D. Li, M. Wang, and M. Zhang, "Arithmetic optimization algorithm based on elementary function disturbance for solving economic load dispatch problem in power system," *Applied Intelligence*, pp. 1–27, 2022.

[19] A. Ramadan, M. Ebeed, S. Kamel, M. I. Mosaad, and A. Abu-Siada, "Technoeconomic and environmental study of multi-objective integration of PV/Wind-Based DGs considering uncertainty of system," *Electronics*, vol. 10, no. 23, p. 3035, 2021.

[20] T. A. A. Victoire and A. E. Jeyakumar, "Hybrid PSO–SQP for economic dispatch with valve-point effect," *Electric Power Systems Research*, vol. 71, no. 1, pp. 51–59, 2004.

[21] A. A. E. Tawfiq, M. O. A. El-Raouf, M. I. Mosaad, A. F. A. Gawad, and M. A. E. Farahat, "Optimal reliability study of grid-connected PV systems using evolutionary computing techniques," *IEEE Access*, vol. 9, Article ID 42139, 2021.

[22] X. Yang, *Nature-Inspired Metaheuristic Algorithms*, Luniver Press, Russia, 2010.

[23] X. Yang and A. Hossein Gandomi, "Bat algorithm: a novel approach for global engineering optimization," *Engineering Computations*, vol. 29, no. 5, pp. 464–483, 2012.

[24] E. Gonçalves, A. R. Balbo, D. N. Silva, L. Nepomuceno, E. C. Baptista, and E. M. Soler, "Deterministic approach for solving multi-objective non-smooth Environmental and Economic dispatch problem," *International Journal of Electrical Power & Energy Systems*, vol. 104, pp. 880–897, 2019.

[25] D. Zou, S. Li, X. Kong, H. Ouyang, and Z. Li, "Solving the combined heat and power economic dispatch problems by an improved genetic algorithm and a new constraint handling strategy," *Applied Energy*, vol. 237, pp. 646–670, 2019.

[26] F. P. Mahdi, P. Vasant, M. Abdullah-Al-Wadud, V. Kallimani, and J. Watada, "Quantum-behaved bat algorithm for many-objective combined economic emission dispatch problem using cubic criterion function," *Neural Computing & Applications*, vol. 31, no. 10, pp. 5857–5869, 2018.

[27] Q. Zhang, D. Zou, N. Duan, and X. Shen, "An adaptive differential evolutionary algorithm incorporating multiple mutation strategies for the economic load dispatch problem," *Applied Soft Computing*, vol. 78, pp. 641–669, 2019.

[28] N. H. Awad, M. Z. Ali, R. Mallipeddi, and P. N. Suganthan, "An efficient Differential Evolution algorithm for stochastic OPF based active–reactive power dispatch problem considering renewable generators," *Applied Soft Computing*, vol. 76, pp. 445–458, 2019.

[29] M. J. Wang, J. S. Pan, T. K. Dao, and T. G. Ngo, "A load economic dispatch based on ion motion optimization algorithm," *Advances in Intelligent Information Hiding and Multimedia Signal Processing*, pp. 115–125, Springer, Singapore, 2019.

[30] T. U. Hassan, T. Alquthami, S. E. Butt, M. F. Tahir, and K. Mehmood, "Short-term optimal scheduling of hydro-thermal power plants using artificial bee colony algorithm," *Energy Reports*, vol. 6, pp. 984–992, 2020.

[31] F. Mohammadi and H. Abdi, "A modified crow search algorithm (MCSA) for solving economic load dispatch problem," *Applied Soft Computing*, vol. 71, pp. 51–65, 2018.

[32] T. C. Bora, V. C. Mariani, and L. D. S. Coelho, "Multi-objective optimization of the environmental-economic dispatch with reinforcement learning based on non-dominated sorting genetic algorithm," *Applied Thermal Engineering*, vol. 146, pp. 688–700, 2019.

[33] M. Kaur, J. S. Dhillon, and D. P. Kothari, "Crisscross differential evolution algorithm for constrained hydrothermal scheduling," *Applied Soft Computing*, vol. 93, Article ID 106393, 2020.

[34] S. T. Suganthi, D. Devaraj, K. Ramar, and S. H. Thilagar, "An Improved Differential Evolution algorithm for congestion management in the presence of wind turbine generators," *Renewable and Sustainable Energy Reviews*, vol. 81, pp. 635–642, 2018.

[35] B. Qiao and J. Liu, "Multi-objective dynamic economic emission dispatch based on electric vehicles and wind power integrated system using differential evolution algorithm," *Renewable Energy*, vol. 154, pp. 316–336, 2020.

[36] B. K. Panigrahi, P. N. Suganthan, and S. Das, "Swarm, Evolutionary, and Memetic Computing," in *Proceedings of the 5th International Conference, SEMCCO 2014*, pp. 18–20, Springer, Bhubaneswar, India, December 2014.

[37] U. Guvenc and E. Kaymaz, "Economic Dispatch Integrated Wind Power Using Coyote Optimization Algorithm," in *Proceedings of the 2019 7th International Istanbul Smart Grids and Cities Congress and Fair (ICSG)*, IEEE, Istanbul, Turkey, August 2019.

[38] I. N. Trivedi, P. Jangir, M. Bhoye, and N. Jangir, "An economic load dispatch and multiple environmental dispatch problem solution with microgrids using interior search algorithm," *Neural Computing & Applications*, vol. 30, no. 7, pp. 2173–2189, 2016.

[39] V. S. Aragón, S. C. Esquivel, and C. A. C. Coello, "An immune algorithm with power redistribution for solving economic dispatch problems," *Information Sciences*, vol. 295, pp. 609–632, 2015.

[40] K. M. Schultz, K. M. Passino, and T. D. Seeley, "The mechanism of flight guidance in honeybee swarms: subtle guides or streaker bees?" *Journal of Experimental Biology*, vol. 211, no. 20, pp. 3287–3295, 2008.

[41] S. Kirkpatrick, "Optimization by simulated annealing: quantitative studies," *Journal of Statistical Physics*, vol. 34, no. 5-6, pp. 975–986, 1984.

[42] S. Mirjalili, A. H. Gandomi, S. Z. Mirjalili, S. Saremi, H. Faris, and S. M. Mirjalili, "Salp Swarm Algorithm: a bio-inspired optimizer for engineering design problems," *Advances in Engineering Software*, vol. 114, pp. 163–191, 2017.

[43] I. M. Sobol', "On the distribution of points in a cube and the approximate evaluation of integrals," *USSR Computational Mathematics and Mathematical Physics*, vol. 7, no. 4, pp. 86–112, 1967.

Research Article

A Study for Development Suitability of Biomass Power Generation Technology Based on GHG Emission Reduction Benefits and Growth Potential

Deming Li [ID]

School of Economics and Management, Beijing Information Science and Technology University, Beijing, China

Correspondence should be addressed to Deming Li; lidm@bistu.edu.cn

Biomass energy can alleviate global warming and solve energy depletion, which is increasingly concerned by the world. Due to the different emission reduction benefits and growth potential of different regions and biomass power generation technologies, analyzing the suitability of these technologies combined with regional conditions can more accurately guide the long- and short-term development of biomass power industry. However, there is no comprehensive evaluation of the benefits and potential of several biomass power generation technologies in different regions. Therefore, this paper introduces development suitability indexes and constructs life-cycle environmental impact and time-value economic impact assessment models, and a growth potential dynamic assessment model based on Gaussian process and particle swarm optimization to illustrate the greenhouse gas emission reduction benefits and growth potential of biomass power generation technologies. The empirical research of China shows that biomass gasification and direct combustion power generation can bring the best environmental benefits, and biogas power generation can bring the best economic benefits. For regions with abundant biomass resources and bad air condition, it is more suitable to develop gasification and direct combustion power generation technology, while mixed-combustion and biogas power generation technology are more suitable in regions with high electricity consumption.

1. Introduction

With the rapid development of economy and the continuous growth of population, the demand for electric energy is increasing. Fossil energy dominated by coal, natural gas, and oil has always been the main energy source in all countries. However, the depletion of fossil fuels and environmental deterioration has become the shackles of social development [1]. The transition to green and clean renewable energy is essential for the sustainable development of society. According to the latest BP world energy statistics yearbook (2021), the global primary energy consumption and the carbon emissions from consumption in 2020 are both recorded the largest decline since 1945, with primary energy consumption falling by 4.5%, oil consumption falling by nearly 3/4 of the net reduction, and carbon emissions falling by more than 6%. The installed capacity of wind and solar energy increases rapidly, with electricity generation

increasing by 238,000 MW, 50% greater than any previous increase. The share of renewable energy in electricity generation increases from 10.3% to 11.7%, while the share of coal decreases by 1.3%. Renewable energy is maintaining strong growth.

Biomass has been a major source of energy since primitive times. Even in modern times, biomass has been the main fuel source for many developing countries [2]. As a kind of renewable energy, biomass has the characteristics of wide sources, abundant reserves, low emissions, and high application potential [3, 4], and is known as the fourth abundant energy after coal, natural gas, and oil [5]. Different from wind and solar energy, biomass stores solar energy in the form of chemical energy, and it uses agricultural residues as representative resources with low sulfur content, which can achieve carbon neutralization and effective emission reduction of GHG. Among the various energy conversion technologies, biomass power generation technology is

relatively mature, and biomass power is of great quality and high reliability, which can generate electricity continuously without time limit, effectively avoiding the intermittent problems caused by wind and solar power generation [6, 7]. If biomass power generation technology can be developed, it can not only alleviate the current contradiction between energy shortage, environmental pollution and economic development, but also create income for farmers, such as creating jobs and increasing income, which is conducive to sustainable social development [8].

Biomass energy has been applied in almost every department of modern industry. The development policies and R&D priorities for biomass energy vary from country to country due to differences in resource conditions and environmental requirements [9]. In one country, the development strategies of the biomass power industry in different regions should also be suitably adjusted due to the differences in the technical level, environmental quality, resource conditions, and industrial demand. The long-term development of the biomass power industry requires a stable supply of resources, mature and viable technologies, and outstanding environmental and economic benefits [10]. Therefore, it is of great significance to analyze the economy and growth potential of biomass power generation technology combined with regional conditions, which can help different countries and regions develop biomass power industry by using their own advantages and achieve the goal of zero carbon emission to the greatest extent.

Research on biomass power generation is extensive, mainly focusing on resource evaluation, modeling, and optimization of power generation technology or process, and has made progress. For example, Song et al. [10] analyzed the greenhouse gas (GHG) emission reduction benefits of straw direct combustion power generation projects using cost-effectiveness analysis and market value method. Xu et al. [11] evaluated the environmental impacts of five mature biomass power generation technologies in China from the perspective of life cycle. Irfan et al. [12] assessed the power generation potential of biomass resources such as sugarcane residues, straw, and animal manure in Pakistan and predicted the development trend of biomass resources. Although these studies can provide suggestions for the overall development of biomass industry, they cannot be applied to a specific region. The guidance provided has some deviation due to the differences between regions. There is no systematic and comprehensive evaluation of the benefits and potential of several specific biomass power generation technologies based on different regions now. There are many kinds of biomass power generation technologies, and the economic benefits, environmental benefits, and development potential of GHG emission reduction of different power generation technologies are different. Specific analysis of the above indexes of different regions and power generation technologies can more accurately guide the long- and short-term development of biomass power industry.

Based on the above analysis, this paper uses the life-cycle assessment (LCA) method and dynamic analysis method based on time value [9] to illustrate the environmental and economic benefits of different biomass power technologies in different regions, and uses Gaussian process (GP) [13] and particle swarm optimization (PSO) algorithm [14] to predict and analyze the regional development potential of different biomass power technologies. The purpose of this paper is to provide reference for the biomass resources development strategies in different regions through benefit and growth potential evaluation.

Since China is a developing country with typical biomass power industry development characteristics, this paper focuses on empirical research of important biomass power generation technologies in China to analyze the GHG emission reduction benefits and growth potential of biomass power generation in four cities.

Compared with the existing work, the contributions of this paper can be summarized as follows.

(1) It is the first time to analyze the GHG emission reduction benefits and growth potential of different biomass power generation technologies combined with the characteristics of one region.

(2) Integrate the environmental and economic benefits of biomass power generation technology in the whole life cycle, and construct the benefit analysis model of GHG emission reduction.

(3) A kernel function integrating long period kernel, short period kernel, rational kernel, and noise kernel is constructed, and GP is used to accurately fit the trend of environmental benefits of GHG emission reduction of biomass power generation technology.

(4) The average annual growth rate of electricity consumption, the number of rural population, the noncompliance rate of air quality, and the amount of biomass resources in various regions are constructed as the development suitability indexes, and the growth potential of biomass power generation technology is dynamically evaluated by PSO.

The remaining sections of this paper are organized as follows. Section 2 provides an overview and analysis of the existing research related to biomass resources. Section 3 describes the data used in the empirical research in this paper. Section 4 describes the economic and environmental benefits analysis methods, and the growth potential assessment methods of the GHG emission reduction of biomass power technology in detail. Section 5 describes the empirical analysis process and results of biomass power generation technology in China. Section 6 discusses the research results and gives some suggestions for the development of biomass. Section 7 summarizes the work of this paper.

2. Related Work

In recent years, the research on biomass resources mainly focuses on the assessment of resource benefits and potential and GIS-based regional distribution analysis, the quantitative and sustainability evaluation of the environmental and economic impacts of biomass power generation technologies, and the macro policy suggestions on various

factors affecting the development of biomass power industry [8].

2.1. Benefit and Potential Evaluation of Biomass Power Generation Resources.

Biomass resources come from a wide range of sources and are mainly classified as forest residues, agricultural residues, animal and poultry manure, and domestic waste [8]. Different biomass resources are converted into electricity, heat, solid, and liquid fuels through various energy conversion technologies, and the GHG emissions in this process are different. The typical is the agricultural residue represented by straw, whose CO_2 emissions during combustion are usually calculated as zero because they produce as much as CO_2 absorbed during growth. In addition, there are differences in the processing costs and power generation potential of different biomass resources.

Many mature methods have been used to analyze biomass resources in terms of resource efficiency and potential assessment. Song et al. [10] calculated the incremental cost of GHG emission reduction of straw direct-fired power generation projects based on the international general calculation method of GHG emission reduction, and evaluated the benefits of straw direct combustion power generation projects in China using the cost-effectiveness analysis and the market value method. Irfan et al. [12] assessed the power generation potential of biomass resources such as sugarcane residues, straw, and animal manure in Pakistan and predicted the development trend of biomass resources. Xu et al. [15] developed a comprehensive model to analyze the logistics cost of straw recycling using calculus methods and guided the site selection planning of biomass power plants and straw planting mode in China through the cost value. Wang et al. [16] used ARIMA to predict the potential of agricultural biomass resources in Heilongjiang province and analyzed the trend of resource distribution on a spatial and temporal scale in combination with GIS. Ji et al. [17] predicted the agricultural residues in China based on ANN algorithm considering the change of crop sown area. Che [18] simulated the spatial distribution of straw resources in China using GIS technique and predicted the future resource potential using grey prediction method. The above research focus more on the energy utilization of regional biomass resources in the process of efficiency and potential calculation and spatial and temporal distribution analysis, ignoring the impact of the resource spatial distribution density on the energy production of biomass resources.

2.2. Evaluation of Biomass Power Generation Technology.

The existing evaluation of biomass power generation technology mainly focuses on benefit and multi-index evaluation. Liang et al. [19] used cost-effectiveness analysis and market value approach to calculate the benefits of CO2 emission reduction, and compared with other renewable energy power generation technologies, concluded that attention should be paid to the development and research of biomass and other renewable energy power generation technologies. Xu et al. [11] evaluated the environmental impacts of five mature biomass power generation technologies in China from a whole life-cycle perspective. Malek et al. [20] summarized biomass scenarios in Southeast Asian and EU countries and used low, average, and high cost estimation methods to conduct economic analysis on biomass gasification power generation technology. Chen et al. [9] simulated and evaluated the environmental loads and economic benefits of biomass direct combustion, gasification, mixed-fired, and biogas power technologies in China using LCA and time-value-based dynamic analysis approach.

In addition, there are also a lot of researches that focus on the comparison of the development priorities of various biomass energy conversion technologies. For example, Zhou [21] used hierarchical analysis with ideal point method to integrate both biomass solid fuel and direct combustion power generation. Khishtandar et al. [22] combined the hesitant fuzzy language data and the preference of experts, adopted the multi-index method to deal with the priority of Iran's existing biomass technologies.

Most of the research on biomass power generation technologies focuses on modeling, optimization, and process evaluation of one technology or process in one region, and a few research focuses on analyzing the impact of specific multiple biomass power generation technologies on economic and environmental benefits. However, they did not explore the differences of regions, which are difficult to further guide the development of the biomass industry.

2.3. Macro Policy Analysis of Biomass Power Generation.

The third perspective of biomass power generation is macro policy recommendation and strategy research, which qualitatively analyzes industrial development constraints and indirectly reflects the suitability of industrial development in different regions. The development of biomass power industry is influenced by various factors, such as cultivated land area, resource potential, logistics network, market demand, environmental demand, government support, business operation, and technology level [22–24]. Zhao et al. [25] used the five forces competition model to assess the current and future development of China's biomass power industry, and provided suggestions on procurement strategies for the sustainable development of the industry according to five influencing factors. Zhu et al. [26] used the strategic analysis tool in the SWOT-PEST model to explore the development of China's biomass power industry. Song et al. [27] constructed a biomass power potential index system to evaluate the development level and potential of biomass power generation in various regions in China, and optimized the target quotas for each city's planning installation by combining the carbon emission intensity of provincial power grids. The above research focuses on the exploration of some influencing factors in industrial development and fails to present the potential of biomass power industry development accurately and comprehensively.

It is found that the above research cannot systematically and comprehensively evaluate the GHG emission reduction benefits and growth potential of biomass power generation

in different regions. They have one or more of the following problems.

(1) There is no comprehensive assessment of multi biomass power technologies, but different power technologies have different benefits and development potential.

(2) The analysis of biomass resources or power generation technology only for single region ignores the differences between regions, which is insufficient to guide the development of biomass power industry.

(3) The assessment of the power generation potential of biomass resources or technologies takes into account historical data on single factor only, ignoring other influencing factors.

Based on the above analysis, we construct evaluation indexes for the suitability of biomass power technology development in terms of resource potential, development demand, and development conditions; conduct empirical research of the biomass power industry in four cities of China based on data from various yearbooks, literature, and some simulations to illustrate the environmental and economic benefits; and assess the growth potential of different biomass power technologies. The analytical approach adopted in this paper can be extended to different regions of other countries, better provide reference for the development strategy of biomass resources, and further promote the development of global biomass industry.

3. Data Description

The data used in this paper are obtained from China's statistical yearbook from 2012 to 2021 [28], the literature [9, 29–34], and some simulation data. The data include the life-cycle list of biomass power technologies, the capital data for biomass and coal-fired power plants, and data on development suitability indexes for four cities (A, B, C, D) in China. The specific data and related descriptions are given below.

3.1. Life-Cycle List of Power Generation Technologies. The life-cycle list contains the 1 kWh life-cycle GHG emissions of four major biomass power technologies (biomass direct combustion power generation, gasification power generation, mixed-fired power generation, and biogas power generation) and conventional coal-fired power technologies in China. This paper mainly considers the emissions of CO_2 and CH_4. The data are obtained from the literature [9, 29–34] and some simulation data. The emissions of CO_2 and CH_4 in 2020 of five power generation technologies in four cities are shown in Table 1.

3.2. Capital Inflow-Outflow List of Biomass Power and Coal-Fired Power Plants. We obtain various capital input-output data from the literature [9] and simulated data for representative biomass and coal-fired power plants in China and process them. The results are shown in Table 2.

Table 1: Life-cycle list of different generation technologies.

City	Technology	CO_2(kg/kWh)	CH_4(kg/kWh)
A	Direct combustion	3.30×10^{-3}	2.30×10^{-6}
	Gasification	2.10×10^{-3}	2.10×10^{-6}
	Mixed-combustion	3.80×10^{-2}	9.00×10^{-5}
	Biogas	2.20×10^{-2}	2.40×10^{-6}
	Coal-fired	4.70×10^{-2}	1.10×10^{-4}
B	Direct combustion	2.21×10^{-2}	1.54×10^{-5}
	Gasification	1.39×10^{-2}	1.39×10^{-5}
	Mixed-combustion	2.57×10^{-1}	6.20×10^{-4}
	Biogas	1.51×10^{-1}	1.60×10^{-5}
	Coal-fired	3.16×10^{-1}	7.70×10^{-4}
C	Direct combustion	1.24×10^{-2}	8.60×10^{-6}
	Gasification	7.70×10^{-2}	7.80×10^{-5}
	Mixed-combustion	1.44×10^{-1}	3.50×10^{-4}
	Biogas	8.40×10^{-2}	8.90×10^{-6}
	Coal-fired	1.77×10^{-1}	4.30×10^{-4}
D	Direct combustion	6.11×10^{-2}	4.26×10^{-5}
	Gasification	3.88×10^{-2}	3.87×10^{-5}
	Mixed-combustion	6.77×10^{-1}	1.62×10^{-5}
	Biogas	4.17×10^{-1}	4.41×10^{-5}
	Coal-fired	8.70×10^{-1}	2.12×10^{-3}

The inflow capital during the construction and operation periods of the biomass power plant project we consider mainly includes by-product revenues such as electricity, and the outflow capital mainly includes civil construction, equipment procurement and installation, equipment commissioning and delivery, raw material and fuel procurement and transportation, labor management expenses, equipment depreciation, and other operating expenses. In data processing, the annual benchmark rate of return of the biomass power station is set to 10%, with an operating period of 20 years, an annual generating period of 5500 h, and a feed-in tariff of 0.8581 yuan/kWh, etc.

3.3. Development Suitability Index of Biomass Power Generation Technology. In this paper, we construct evaluation indexes for the suitability of biomass power technology development in terms of resource potential, development demand and development conditions, including the average annual growth rate of electricity consumption and the amount of biomass resources, the noncompliance rate of air quality, and the number of rural population. We select a representative city from North China, East China, Central China, and Southwest China, which are A, B, C, and D, respectively. Among them, B and C are large agricultural cities, A is an industrial city, and D is an animal husbandry city.

The electricity consumption value, cultivated land area, and forest area from 2010 to 2020 for four cities are obtained from China's statistical yearbook. The data from 2015 to 2020 are shown in Table 3. Among them, A and D have fewer rural laborers and smaller vegetation coverage and cultivated land area, which means that the biomass resources (such as straw) of the two cities are relatively less; B and C have more rural laborers and larger cultivated land area, especially C. Moreover, B and D have better air quality and lower

TABLE 2: Capital inflow-outflow list of different power generation technologies.

City	Technology	Initial investment (yuan)	Operation investment (yuan)	Operation profit (yuan)	Annual profit (yuan)	Initial investment (yuan)
A	Direct combustion	306,453,975	80,958,737	141,506,266	60,547,530	306,453,975
	Gasification	66,162,041	10,988,891	18,867,502	7,878,612	66,162,041
	Mixed-combustion	1,370,752,620	1,011,435,335	1,061,451,368	50,016,033	1,370,752,620
	Biogas	110,151,908	3,716,326	19,862,334	16,146,008	110,151,908
	Coal-fired	6,031,345,832	3,081,074,539	3,882,428,813	801,354,274	6,031,345,832
B	Direct combustion	245,163,180	64,766,989	113,205,013	48,438,024	245,163,180
	Gasification	52,710,791	8,791,111	15,094,002	6,302,891	52,710,791
	Mixed-combustion	109,660,173	809,148,264	849,161,092	40,012,828	109,660,173
	Biogas	88,121,530	2,973,061	15,889,864	12,916,803	88,121,530
	Coal-fired	4,825,076,669	2,464,859,630	3,105,943,049	641,083,419	4,825,076,669
C	Direct combustion	220,646,862	58,290,287	101,884,512	43,594,198	220,646,862
	Gasification	47,439,708	7,912,004	13,584,602	5,672,606	47,439,708
	Mixed-combustion	98,694,152	728,233,434	764,244,986	36,011,543	98,694,152
	Biogas	79,309,373	2,675,755	14,300,875	11,625,121	79,309,373
	Coal-fired	79,309,373	2,218,373,664	2,795,348,747	576,975,074	79,309,373
D	Direct combustion	147,097,908	38,860,197	67,923,008	29,062,811	147,097,908
	Gasification	31,626,478	5,274,666	9,056,401	3,781,735	31,626,478
	Mixed-combustion	65,796,098	485,488,962	509,496,657	24,007,695	65,796,098
	Biogas	52,872,922	1,783,837	9,533,920	7,750,083	52,872,922
	Coal-fired	2,895,046,001	1,478,915,776	1,863,565,831	384,650,055	2,895,046,001

TABLE 3: Biomass power generation technology development suitability indexes for four cities in China from 2015 to 2020.

Year	City	Average annual growth rate of electricity consumption (%)	Number of rural population	Air quality noncompliance rate (below secondary) (%)	Amount of biomass resources
2015	A	1.67	293	49.04	218.7
	B	6.75	2209	14.79	1618.2
	C	−1.36	5039	62.74	4756.9
	D	19.28	234	14.25	70.1
2016	A	1.67	293	49.18	211.8
	B	6.75	2,209	15.03	1,620.6
	C	−1.37	5,039	62.84	4,745
	D	10.93	234	14.48	70.1
2017	A	7.09	293	45.75	197.4
	B	8.76	2,154	12.88	1,771.5
	C	3.80	4,909	56.44	4,698.5
	D	21.44	233	14.25	54.6
2018	A	4.60	293	38.08	188.6
	B	9.38	2,098	17.81	1,785.8
	C	5.92	4,764	54.52	4,681.6
	D	18.37	233	1.10	81.3
2019	A	7.03	291	37.81	209.8
	B	10.43	2,044	10.41	1,868.8
	C	7.96	4,638	53.97	5,102.7
	D	18.97	237	1.92	91.3
2020	A	2.10	289	34.25	217.9
	B	7.49	1,987	11.78	1,967.1
	C	−1.58	4,511	51.51	5,549.4
	D	13.04	240	6.30	98.4

carbon emissions; the power demand of D is growing faster than that of the other four cities. Due to the differences in development suitability index values, there are also some differences in the development of biomass technology between different regions. The purpose of this paper is to explore the most suitable biomass technology in one region based on the above indexes.

4. Methodology

In this paper, we construct a life-cycle environmental impact evaluation model and a time-value economic impact evaluation model to analyze the environmental and economic benefits of GHG emission reduction of biomass power generation technologies, and construct a growth potential dynamic assessment model to predict and explore the benefits and development potential of these technologies.

4.1. Life-Cycle Environmental Impact Assessment Model. Inspired by the literature [9], this paper uses LCA to analyze the whole life-cycle environmental assessment of biomass power generation and compare it with coal-fired power generation from conventional coal-fired power plants to describe the current status of environmental emissions and emission reduction benefits of various biomass power generation technologies and coal-fired power generation technology. Due to the complexity of systems involving multiple power generation technologies, the life cycle of these technologies considered in this paper includes only raw material production, processing, transportation, and power generation operation stages.

The calculation of GHG emission reduction benefits of different biomass power technologies in different regions through the LCA approach first requires the construction of inventory analysis data, i.e., gas pollutant emissions for each power technology over the full life-cycle process, which is interpreted by the intensity of the contribution of each specific environmental exchange to the determined environmental impact type. The data are described in detail in Section 3, as shown in Table 2.

Next, the potential environmental impact values are calculated for each impact type, assuming potential environmental impact value is a and GHG is b, the calculation equation is as follows:

$$EP(a) = \sum EP_b(a) = \sum [Q_b \times EF_b(a)], \tag{1}$$

where $EP(a)$ represents the contribution of a, $EP_b(a)$ represents the contribution of b to a, Q_b represents the emission of b, and $EF_b(a)$ is the equivalence coefficient of b and a.

With this equation, we calculate the total environmental load of different power generation technologies, and in this paper we only consider the environmental impact category of global warming.

4.2. Time-Value Economic Impact Evaluation Model. In this paper, we construct a model for calculating the economic benefits of biomass technologies based on the economic

parameters of biomass power plant projects during the construction and operation periods. Inspired by the literature [10], we use dynamic analysis method based on time value to evaluate the economic benefits of different biomass power generation and coal-fired power generation technologies. The NPV will be used as a specific index to assess the economic benefits in this paper, which is calculated as follows:

$$NPV = \sum_{t=0}^{n} (I - O)_t (1 + i_c)^{-t}, \tag{2}$$

where I and O are capital inflow and capital outflow, respectively, $(I - O)_t$ represents the net financial flows for the year t, n represents the total number of years, and i_c represents the base earning rate.

4.3. Model for Dynamic Assessment of Growth Potential. The existing research have not established systematic potential evaluation methods for different biomass power generation technologies, but generally the potential of one biomass power generation technology in one region and its application prospects are analyzed in terms of biomass resource, cost-effectiveness, economic conditions, and development demand, such as literature [35, 36].

The growth potential assessment of biomass power generation in this paper consists of two parts: one part is to predict the GHG emission reduction benefits (economic and environmental benefits) of important biomass power generation technologies in four regions of China based on GP, and the other part is to calculate the growth potential of different power generation technologies based on PSO.

The prediction of GHG emission reduction benefits of biomass power generation technologies is based on the emission reduction benefits of power generation technologies in the regions in the past years to predict the emission reduction benefits for the next years. GP is determined by mathematical expectations and covariance functions, and some covariance functions with specific forms become kernel functions. In this paper, we choose the sum of four kernel functions $K = K_1 + K_2 + K_3 + K_4$ as the final kernel function of GP. K_1, K_2, K_3, and K_4 represent the long period kernel, short period kernel, rational kernel, and noise kernel functions, respectively. Since the GHG emission reduction benefits of biomass power generation technology may have short-term and long-term trends, K_1 and K_2 are used to fit this change, respectively; K_3 is smooth, which can effectively fit the irregular parts; and K_4 can model the residual noise in the data. Combining these four kernel functions, we can better capture the dynamic trend of emission reduction benefits. The equations of the four kernel functions are as follows:

$$K_1 = \exp\left(-\frac{1}{\varepsilon_1^2}\sin^2\frac{\pi(x_i - x_j)}{\lambda_1}\right),$$

$$K_2 = \exp\left(-\frac{1}{\varepsilon_2^2}\sin^2\frac{\pi\left(x_i - x_j\right)}{\lambda_2}\right),$$

$$K_3 = \exp\left(1 + \frac{\left(x_i - x_j\right)^2}{2\varepsilon_3^2}\right)^{-\alpha}, \qquad (3)$$

$$K_4 = \exp\left(-\frac{\left(x_i - x_j\right)^2}{2\varepsilon_4^2}\right),$$

where x_i and x_j represent the emission reduction benefits in year i and j, respectively; ε are developed to control how fast the kernel function changes with time; and λ is the cycle length.

After the kernel function configuration is completed, this paper solves the mathematical expectation, and kernel function of the Gaussian joint distribution is resolved by maximizing the log-likelihood function of the hyperparameters. Assuming that $m(t)$ is the mathematical expectation and $K(t)$ is the covariance, the equation is as follows:

$$\log P\left(x_{t+1}|X, m(t), K(t)\right) = -\frac{1}{2}\log|K(t)| - \frac{1}{2}(x_{t+1} - m(t))^T K(t)^{-1}\left(x_{t+1} - m(t)\right). \qquad (4)$$

In order to quantify the long-term joint impact of different indexes on the growth potential of biomass power generation technology, this paper uses PSO to achieve the goal [14]. PSO simulates the birds in the bird swarm through a massless particle. The current position of the particle is a candidate solution to the corresponding optimization problem, the flight process is the search process, and the flight speed can be dynamically adjusted according to the historical optimal position of the particle and the swarm. Each particle has two attributes of position and velocity. The optimal solution searched separately is called individual extremum, and the optimal individual extremum in the particle swarm is used as the current global optimal solution. In the process of iteration, the attributes of each particle are updated, and finally the optimal solution satisfying the termination condition is obtained. The flow of the algorithm is shown in Figure 1. Here, m_i represents a random particle, v_i and p_i are the velocity and position attributes of the particle, respectively, and p_{best} represents the combination of historical optimal development suitability indexes for a single particle, and g_{best} represents the combination of historical optimal development suitability indexes for a particle swarm. In this paper, the algorithm is used to mine the maximum GHG emission reduction benefits of biomass power technology and obtain the best combination of development suitability indexes, so the search objective of the particle swarm in this paper is to solve the maximum benefit. The algorithm first initializes N random particles m_i in the feature space. In the iteration S, $m_i(s) = (m_{i,1}(s), \ldots, m_{i,d}(s), \ldots, m_{i,D}(s))$, where D represents the number of development suitability indexes used in this paper. We use CfsSubsetEval to evaluate the current combination of developmental suitability indexes of the particles, and the output value $F(m_i(s))$ is used as the evaluation index of $m_i(s)$. Since the objective is to solve the benefit maximization problem, the larger the $F(m_i(s))$ of $m_i(s)$, the better the position it is than other particles. During the iteration, the particles always maintain p_{best} and g_{best}.

In this paper, PSO is used to find the best indexes combination that makes the biomass power generation technologies get the greatest environmental benefits. Then, we can get the factors that affect the development of one technology, and the growth potential in one region can be accessed based on the regional characteristics.

5. Case Study and Result Analysis

Taking China as an example, this paper analyzes the benefits of biomass direct combustion, gasification, mixed-fired and biogas power generation technologies in four cities (A, B, C, and D) from both environmental and economic perspectives, and introduces development suitability indexes to predict development trends and assess growth potential of these technologies in these cities. To calculate the emission reduction benefits, we introduce coal-fired power generation technology.

5.1. Environmental Benefit Analysis of Biomass Power Generation Technology. We select unit capacities of 30 MW, 4 MW, 300 MW, 2 MW, and 1,320 MW for biomass direct combustion power generation, gasification power generation, mixed-combustion power generation, biogas power generation, and coal-fired power generation, respectively, set the lifetime of all power generation projects to 20 years, and set the functional unit (Fu) as power generation 1 kWh.

In this paper, global warming is selected as an environmental impact category and assess the impact of GHG emissions from each power generation technology on global warming because the measured impact potential of this impact category is close to the true potential and the assessment is more accurate and representative [9]. Global warming is calculated in terms of CO2-equivalents. The characterization factor of CO2 is 1 and that of CH4 is 25.

According to Table 1 and equation (1), in 2020, the potential GHG impact values for each power generation technology are calculated as shown in Table 4. Taking the potential impact value of direct combustion power generation technology of city A as an example, the calculation process of these results is: the emissions of CO_2 and CH_4 are obtained from Table 1, which are, respectively, 3.30×10^{-3} and 2.30×10^{-6}, then the potential impact value is $3.30 \times 10^{-3} \times 1 + 2.30 \times 10^{-6} \times 25 = 3.35 \times 10^{-3}$.

```
Algorithm 1 PSO
1:  for each m_i do
2:      initialize v_i and p_i for m_i;
3:      evaluate m_i and set p_best = p_i;
4:  end for
5:  g_best = max(p_best);
6:  while not stop do
7:      for i = 1 to N do
8:          update v_i and p_i;
9:          evaluate m_i;
10:         if fit(p_i) > fit(p_best) then
11:             p_best = p_i;
12:         end if
13:         f fit(p_best) > fit(g_best)) then
14:             g_best = p_best;
15:         end if
16:     end for
17: end while
18: return g_best;
```

FIGURE 1: Flow of PSO.

TABLE 4: Characterization results of different power generation technology in 2020.

Technology	City A	City B	City C	City D
Direct combustion	3.35×10^{-3}	2.25×10^{-2}	1.26×10^{-2}	6.22×10^{-2}
Gasification	2.13×10^{-3}	1.42×10^{-2}	9.74×10^{-3}	3.94×10^{-2}
Mixed-combustion	4.07×10^{-2}	2.73×10^{-1}	1.53×10^{-1}	7.11×10^{-1}
Biogas	2.27×10^{-2}	1.52×10^{-1}	8.51×10^{-2}	4.19×10^{-1}
Coal-fired	4.50×10^{-2}	3.35×10^{-1}	1.88×10^{-1}	9.27×10^{-1}

According to Table 4, in 2020, the potential impact of gasification power generation technology on GHG is the least in each city, followed by direct combustion, biogas, and mixed-combustion power generation. Compared to the environmental load of conventional coal-fired power generation, their average environmental emission reduction benefits are 95.52%, 93.29%, 54.73%, and 19.79%, respectively.

In addition, this paper calculates the average environmental emission reduction benefits of each biomass power generation technology in A, B, C, and D (relative to coal-fired power generation technology) from 2010 to 2020, as shown in Figure 2. Overall, the environmental benefits of the four technologies maintain a relatively stable trend. The benefits of gasification and direct combustion power generation technologies are high, the benefits of biogas power generation are general, and the benefits of mixed-combustion power generation are relatively low.

5.2. Economic Benefit Analysis of Biomass Power Generation Technology.

The capital input-output data we used for biomass and coal-fired power plants are shown in Table 2, and the contents of which are described in detail in Section 3.

The economic efficiency index considered in this paper is NPV. According to Table 2 and equation (2), the economic efficiency indexes for each generation technology are shown in Table 5.

Taking the economic benefit calculation process of direct combustion power generation technology of city A as an example: $I = 141,506,266 + 60,547,530 = 202,053,796$;
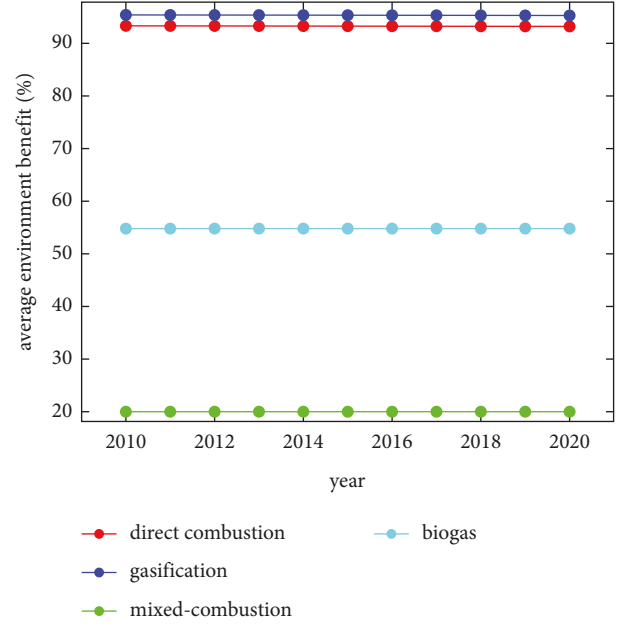

FIGURE 2: Average environmental benefits of four biomass technologies in China from 2010 to 2020.

$O = 306,453,975 + 80,958,737 = 441,412,712$; then the value of NPV is $\sum_{t=0}^{20}(202,053,796 - 441,412,712)_t (1 + 0.1)^{-t} = 6,963,824$.

Based on the results of the economic evaluation of different biomass power technologies in different cities, on the whole, the technology for biomass biogas power has the highest NPV, followed by direct combustion power, coal-fired power technology, and lower NPV for mixed-combustion power and gasification power. In the four cities, direct combustion and biogas power generation have greater economic benefits compared to coal-fired power generation.

5.3. Development Trend Prediction of Biomass Power Generation Technology.

In this paper, we analyze the development trend of different biomass power generation technologies in each city from an environmental perspective, and use the environmental benefits of GHG emission reduction of the important biomass power generation technologies in four cities of China A, B, C, and D from 2010 to 2020 to predict the environmental benefits of each technology in each city from 2021 to 2030. To ensure the accuracy of the results, we first divide the data from 2010 to 2020 into training set and test set for experiments to evaluate the prediction performance of GP. Among them, the data from 2010 to 2017 are used as the training set and the rest data are used as the test set. The experimental results are shown in Table 6. It can be seen that the average prediction accuracy of GP is 86%, which can effectively fit the trend of environmental benefits of various biomass power generation technologies.

Next, we use the trained GP model to predict the environmental benefits of four biomass power generation technologies in four cities from 2021 to 2030, as shown in Figure 3. The environmental benefits of these technologies vary greatly, especially biogas power generation technology.

TABLE 5: Economic evaluation results of different power generation technologies.

City	Technology	Scale (MW)	NPV (yuan·MW^{-1})
A	Direct combustion	30	6,963,824
	Gasification	4	228,697
	Mixed-combustion	300	1,417,921
	Biogas	2	13,653,211
	Coal-fired	1,320	5,168,552
B	Direct combustion	30	5,571,061
	Gasification	4	182,958
	Mixed-combustion	300	1,134,339
	Biogas	2	10,922,570
	Coal-fired	1,320	4,134,838
C	Direct combustion	30	5,013,954
	Gasification	4	164,665
	Mixed-combustion	300	1,020,902
	Biogas	2	9,830,314
	Coal-fired	1,320	3,721,360
D	Direct combustion	30	3,342,633
	Gasification	4	109,771
	Mixed-combustion	300	680,602
	Biogas	2	6,553,537
	Coal-fired	1,320	2,480,906

For direct combustion power generation, A shows an obvious downward trend, while B, C, and D remain unchanged; for gasification power generation, A decreases slightly and C increases significantly; the environmental benefits of hybrid and biogas power generation technology in A and D show a slight downward trend; C shows an upward trend of biogas power generation year by year, and B fluctuates slightly but remains basically unchanged. From 2021 to 2030, biomass power generation technology with the best environmental benefits in the four cities is still gasification and direct combustion, followed by biogas and hybrid power generation.

5.4. Assessment of Growth Potential of Biomass Power Generation Technologies. This paper introduces the average annual growth rate of electricity consumption and the amount of biomass resources biomass resources, the air quality noncompliance rate, and the number of rural population as indexes affecting the development of biomass power generation technology in one city, where the average annual growth rate of electricity consumption and biomass resources reflect the resource potential of a city, the air quality noncompliance rate reflects the need for environment improvement in the city, and the number of rural population reflects the agricultural biomass cultivation, collection, and human conditions in the transportation process.

We construct a PSO model to search for the best impact indexes of each biomass power technology based on the data of environmental benefits and development suitability indexes, so as to analyze the most suitable biomass power technology to be developed in each city. We construct the model on WEKA 3.9.5 and iterate 20 times to obtain the results as shown in Table 7.

TABLE 6: Environmental benefits prediction accuracy of GP.

\	Train	Test
Accuracy (%)	87.89	84.73

According to Table 7, the air quality noncompliance rate and the amount of biomass resources are the main factors affecting the biomass direct combustion and gasification power generation technologies, and the annual growth rate of electricity consumption is the main factor affecting the mixed-combustion and biogas power generation technologies. Combining Tables 3 and 7, city A and city C have higher air quality noncompliance rates and city C has abundant biomass resources, so biomass direct combustion power generation and gasification power generation technologies have greater development potential in these two cities, which confirms our previous conclusion that these two technologies have the best environmental benefits in terms of GHG emission reduction. The average annual growth rate of electricity consumption in city B and D is much higher than that of other cities, which indicates that the demand for electricity in these two cities is high, so the mixed-combustion and biogas power generation technologies have greater development potential in city B and C. In addition, city B also has abundant biomass resources, so it is also suitable for developing biomass direct combustion and gasification power generation technologies.

6. Discussion and Policy Recommendations

Biomass power generation has become a great need for sustainable development and clean energy. However, due to the differences of technical level and environmental quality in various regions, the biomass power generation technology suitable for development is also different. Only by analyzing the economic benefits, environmental benefits and growth potential of biomass power generation technology in combination with regional conditions, can we promote the development of biomass industry in various regions and achieve the goal of zero emission to the greatest extent. For China dominated by agriculture, biomass gasification power generation and direct combustion power generation technology can bring better environmental benefits, and biogas power generation technology can bring better economic benefits. Specifically, cities with serious air pollution in China, development priority should be given to biomass power generation technologies, such as direct combustion power generation and gasification power generation, if there are a large number of local rural areas and abundant biomass resources such as straw and forestry waste, it is also more suitable for the development of these two biomass industries. For areas with high electricity consumption, development priority should be given to the promotion of mixed-combustion and biogas power generation technology.

Based on our research findings, the following policy recommendations are given for the China government.

(1) Establish suitability indicators for the development of biomass industry: the government should divide regions according to the average annual growth rate
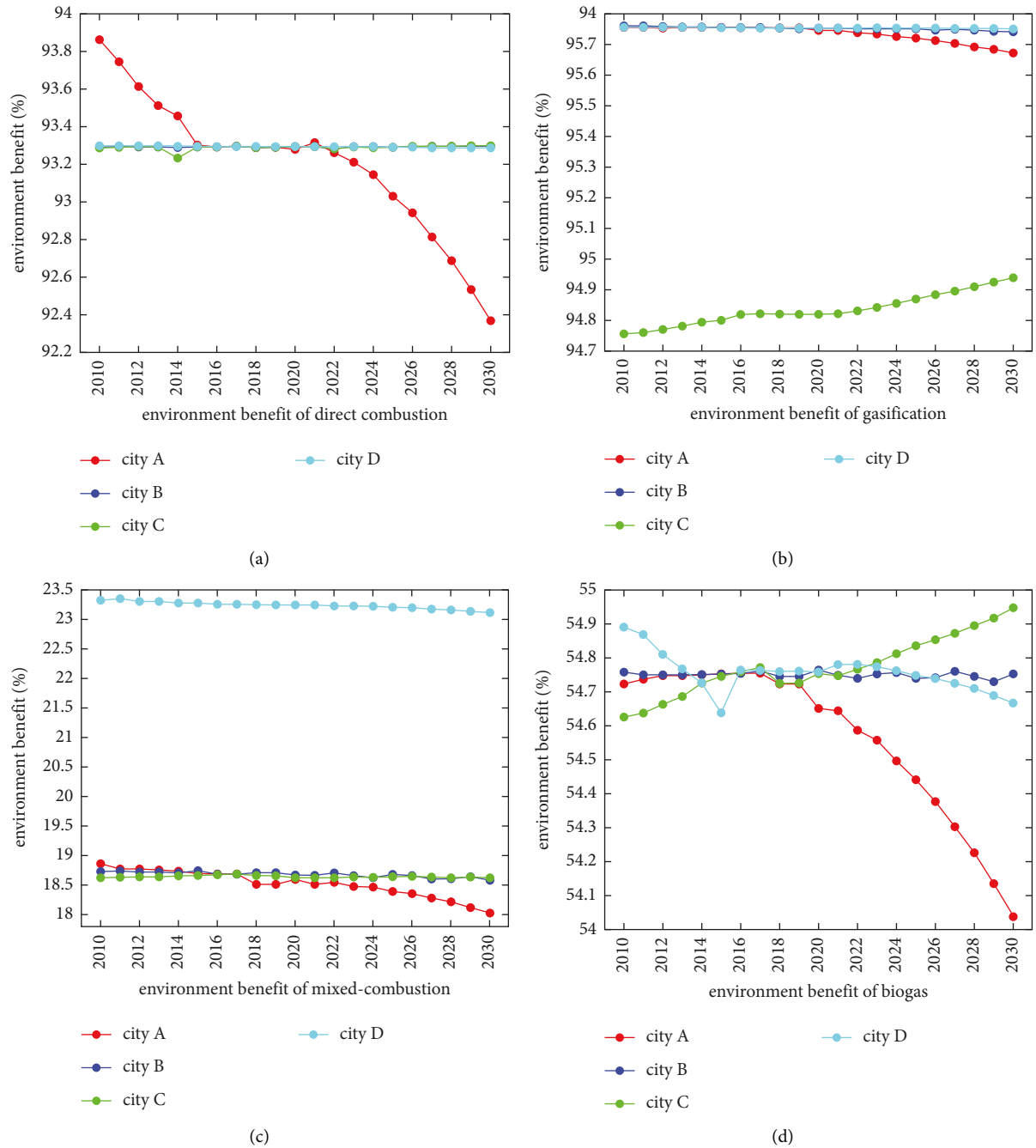
(a)

(b)

(c)

(d)

FIGURE 3: Environmental benefits of GHG emission reductions for four important biomass power technologies in four cities of China from 2021 to 2030. (a) Environment benefit of direct combustion. (b) Environment benefit of gasification. (c) Environment benefit of mixed-combustion. (d) Environment benefit of biogas.

TABLE 7: Economic evaluation results of different power generation technologies.

Technology	Indexes
Direct combustion	Air quality noncompliance rate, amount of biomass resources
Gasification	Air quality noncompliance rate, amount of biomass resources
Mixed-combustion	Annual growth rate of electricity consumption
Biogas	Annual growth rate of electricity consumption

of electricity consumption, biomass resources, air quality noncompliance rate, rural population and other indicators, and make development strategies according to local conditions.

(2) Further stimulate people's environmental awareness: the government should educate the public about the benefits of biomass for sustainable development, especially for rural residents, as they are located in regions with sufficient biomass resources, they

should be educated to make the best of straw, livestock manure, and other resources to avoid resource waste.

(3) Explore new markets and technologies: the government should promote the export of technologies and products of local biomass energy enterprises. For example, for agricultural countries such as Southeast Asia, direct combustion power generation and gasification power generation technology can be promoted. In addition, the government should constantly absorb modern technology and enhance the economic and environmental benefits of biomass technology.

7. Conclusion

In this paper, the environmental load and economic benefits of the biomass power generation process are simulated using the LCA and the dynamic analysis method based on time value, and the environmental and economic benefits of different biomass power generation technologies are illustrated by comparing them with coal-fired power generation technologies. In addition, the GP and PSO algorithm are used to predict and analyze the development potential of different biomass power generation technologies in different regions. According to the empirical analysis of China, biomass direct combustion and gasification power generation technology are suitable for development in regions with concentrated biomass resources and serious air pollution; mixed-combustion and biogas power generation technology are suitable for use in regions with high electricity demand. China's geographical situation and resource characteristics determine that it should vigorously develop biomass gasification and direct combustion power generation technology, and develop mixed-combustion and biogas power generation technology according to local conditions. Although the research of this paper focuses on the agricultural countries represented by China, it is also beneficial to other typical geographical patterns and can effectively guide the development of the biomass industry in different regions of other countries.

Data Availability

All data could be accessed by request.

Conflicts of Interest

The author declares no conflicts of interest.

References

[1] D. Gielen, F. Boshell, D. Saygin, M. D. Bazilian, N. Wagner, and R. Gorini, "The role of renewable energy in the global energy transformation," *Energy Strategy Reviews*, vol. 24, pp. 38–50, 2019.

[2] A. Kumar, N. Kumar, P. Baredar, and A. Shukla, "A review on biomass energy resources, potential, conversion and policy in India," *Renewable and Sustainable Energy Reviews*, vol. 45, pp. 530–539, 2015.

[3] A. A. A. Abuelnuor, M. A. Wahid, S. E. Hosseini et al., "Characteristics of biomass in flameless combustion: a review," *Renewable and Sustainable Energy Reviews*, vol. 33, pp. 363–370, 2014.

[4] Y. B. Deng, C. Liu, and S. B. Wu, "A review on fast pyrolysis of biomass for bio-oil," *Adv. New Renew. Energy*, vol. 2, pp. 334–341, 2014.

[5] M. Gan, Z. Ji, X. Fan, Y. Zhao, X. Chen, and Y. Fan, "Insight into the high proportion application of biomass fuel in iron ore sintering through CO-containing flue gas recirculation," *Journal of Cleaner Production*, vol. 232, pp. 1335–1347, 2019.

[6] J. Ren and B. K. Sovacool, "Prioritizing low-carbon energy sources to enhance China's energy security," *Energy Conversion and Management*, vol. 92, pp. 129–136, 2015.

[7] J. He, Y. Liu, and B. Lin, "Should China support the development of biomass power generation?" *Energy*, vol. 163, pp. 416–425, 2018.

[8] K. Li, *Research on Regional Suitability Evaluation and Energy, Environmental and Economic Benefits of Biomass Power Generation Industry (Master's Thesis*, Jilin University, Changchun, China, 2019.

[9] S. Chen, H. Feng, J. Zheng et al., "Life cycle assessment and economic analysis of biomass energy technology in China: a brief review," *Processes*, vol. 8, no. 9, p. 1112, 2020.

[10] S. Yan-Pei, J. Xiu-Mei, Y.-J. Zhuang et al., "Greenhouse gas mitigation benefits of middle-scale biomass power generation project," *Guangzhou Chemical Industry*, vol. 44, no. 12, pp. 12–14, 2016.

[11] C. Xu, J. Hong, J. Chen, X. Han, C. Lin, and X. Li, "Is biomass energy really clean? An environmental life-cycle perspective on biomass-based electricity generation in China," *Journal of Cleaner Production*, vol. 133, pp. 767–776, 2016.

[12] M. Irfan, Z. Y. Zhao, M. K. Panjwani et al., "Assessing the energy dynamics of Pakistan: prospects of biomass energy," *Energy Reports*, vol. 6, pp. 80–93, 2020.

[13] R. M. Dudley, "Sample functions of the Gaussian process," *Selected Works of RM Dudley*, Springer, New York, NY, USA, pp. 187–224, 2010.

[14] Y. Shi, "Particle swarm optimization," *IEEE connections*, vol. 2, no. 1, pp. 8–13, 2004.

[15] X. L. Xu and Y. J. Chen, "A comprehensive model to analyze straw recycling logistics costs for sustainable development: evidence from biomass power generation," *Progress & Sustainable Energy*, vol. 39, no. 4, Article ID e13394, 2020.

[16] W. Wang, W. Ouyang, F. Hao, and G. Liu, "Temporal-spatial variation analysis of agricultural biomass and its policy implication as an alternative energy in northeastern China," *Energy Policy*, vol. 109, pp. 337–349, 2017.

[17] L. Q. Ji, "An assessment of agricultural residue resources for liquid biofuel production in China," *Renewable and Sustainable Energy Reviews*, vol. 44, pp. 561–575, 2015.

[18] Li Che, "Study on crop residue estimate, spatial distribution and use potential," Doctoral dissertation, Dalian University of Technology, Dalian, China, 2014.

[19] Y.-J. Liang and J.-C. Fan, "The benefit of renewable energy power generation technologies to the reduction of the greenhouse gas emission," *Renewable Energy Review Journal*, no. 1, pp. 23–25, 2004.

[20] A. A. Malek, M. Hasanuzzaman, and N. Abd Rahim, "Prospects, progress, challenges and policies for clean power generation from biomass resources," *Clean Technologies and Environmental Policy*, vol. 22, no. 6, pp. 1229–1253, 2020.

[21] Y. Zhou, "Evaluation of biomass solid fuel based on forest logging residues," Master's thesis, Fujian Agricultural and Forestry University, Fuzhou, China, 2016.

[22] S. Khishtandar, M. Zandieh, and B. Dorri, "A multi criteria decision making framework for sustainability assessment of bioenergy production technologies with hesitant fuzzy linguistic term sets: the case of Iran," *Renewable and Sustainable Energy Reviews*, vol. 77, pp. 1130–1145, 2017.

[23] R. Golecha and J. Gan, "Effects of corn stover year-to-year supply variability and market structure on biomass utilization and cost," *Renewable and Sustainable Energy Reviews*, vol. 57, pp. 34–44, 2016.

[24] J. L. Thompson and W. E. Tyner, "Corn stover for bioenergy production: cost estimates and farmer supply response," *Biomass and Bioenergy*, vol. 62, pp. 166–173, 2014.

[25] Z. Y. Zhao, J. Zuo, P. H. Wu, H. Yan, and G. Zillante, "Competitiveness assessment of the biomass power generation industry in China: a five forces model study," *Renewable Energy*, vol. 89, pp. 144–153, 2016.

[26] L. Zhu, E. Hiltunen, E. Antila, F. Huang, and L. Song, "Investigation of China's bio-energy industry development modes based on a SWOT-PEST model," *International Journal of Sustainable Energy*, vol. 34, no. 8, pp. 552–559, 2015.

[27] K. Song, J. Zhou, P. Zhang, and S. Kan, "Assessment of biomass power potential on provincial scale and analysis on plan target quota," *Forum on Science and Technology in China*, no. 1, pp. 124–129, 2016.

[28] National Bureau of Statistics, *China Statistical Yearbook*, China Statistics Press, Beijing, China, 2012-2021.

[29] W. Wang, D. Q. Zhao, H. L. Yang, J. Y. Cai, and P. Chen, "Life cycle analysis on biomass gasification power generation system and inquiry to assessment method," *Acta Energ*, vol. 26, pp. 752–759, 2016.

[30] Y. J. Jia, Z. Yu, and C. Z. Wu, "4 MW Life cycle assessment of biomass gasification combined cycle power generation system," *Acta Energ*, vol. 25, pp. 56–62, 2004.

[31] J. W. Liu, "Study on the initial conditions and life cycle assessment of biomass energy utilization system," Master's thesis, Beijing University of Chemical Technology, Beijing, China, 2010.

[32] Y. Gong, "Technical economy and environmental benefit analysis of biogas power generation project," Master's thesis, North China Electric Power University, Beijing, China, 2018.

[33] F. C. Pan, "Life cycle assessment of the utilization of kitchen waste to produce biogas for power generation," Master's thesis, Guangxi University, Nanning, China, 2018.

[34] X. H. Di, Z. R. Nie, and T. Y. Zuo, "Life cycle amission inventories for the fuels consumed by thermal power in China," *China Environ*, vol. 25, pp. 632–635, 2005.

[35] E. M. Yan, C. D. Cai, and W. H. Wang, "Introduction to biomass gasification power generation technology and application potential," *Yunnan Electric Power*, vol. 5, pp. 16–21, 2012.

[36] T. Y. Qi, X. L. Zhang, X. M. Ou, Z. Liu, and S. Y. Chang, "The regional cost of biomass combustion power generation in China and development potential analysis," *Renewable Energy Resources*, vol. 29, no. 2, pp. 115–118, 2011.

Research Article

ECANP: A Topic Influence Evaluation Model for Hot Topics

Yiru Chang ⓘ,[1] **Zhiyuan Zhang** ⓘ,[1] **and Guixun Luo** ⓘ[2]

[1]*School of Electronic and Information Engineering, Key Laboratory of Communication and Information Systems,*
 Beijing Municipal Commission of Education, Beijing Jiaotong University, Beijing 100044, China
[2]*School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China*

Correspondence should be addressed to Zhiyuan Zhang; zhangzhiyuan@bjtu.edu.cn

Social network is an important product of industrial society. In recent years, the research related to hot topics has focused on topic detection, topic trend prediction, and topic tracking. However, the important role of topic influence evaluation in hot topic research has not received enough attention, which leads the problem of inaccurate influence calculation. In order to solve the above problems, this paper proposes a novel model to evaluate the real-time relative influence of topics in social network. The proposed model can quantify the influence of topics, and some influential factors which determine topic hotness will be analyzed and identified. In this model, five impact indicators are defined, namely user engagement, topic coverage, topic activity, topic persistence, and topic novelty to consider the topic characteristics more finely. Moreover, the proposed model not only consider traditional simple factors of like, forward and comments, but also pay attention to the relative influence and time attenuation characteristics of the topics. Further, the experimental results show that our method could quickly aggregate the influence factors of hot topics and accurately provide the influence indicator of topics.

1. Introduction

In the society of smart industry, the Internet has increasingly become a way of information dissemination that cannot be ignored. On the one hand, as the most timely and widest media for information access, the Internet has become the main channel for the central government, relevant government departments, and authoritative platforms to release news and access information [1]. On the other hand, due to its good interaction, various forms of expression, and outstanding appeal, it has not only attracted major websites and organizations to use it as a publicity channel, but also gained more and more users. By December 2021, the number of Internet users in China had reached 1.032 billion, and the Internet penetration rate had reached 73.0% [2], an increase of 1.4 percentage points over half a year ago. It can be seen that the Internet plays an extremely essential role in information dissemination and daily life.

The emergence of social media, especially the application of mobile communication technology, enables people to break the restrictions of time and space simultaneously, and garner, share, and exchange information from the Internet whenever and wherever possible, which led to an increasing number of Internet users and the rapid expansion of information on various platforms. In particular, Sina Weibo has become one of the most mainstream social platforms in China with its weak interactivity and sharing. Users can follow their friends and interact with interested bloggers. When people are interested in an event on the Internet, it will be liked, forwarded, or commented, and spread quickly, resulting in relevant events. At the same time, the same or similar events discussed by everyone form a topic, which has a certain social influence. A hot topic will be formed when a topic is highly concerned or participated by Internet users just like the microblog hot search list. The collective public opinions, emotional attitudes, and values in hot topics form public opinion and have a great impact on public life. Therefore, how to find influential topics from a large number of articles and opinions on microblog is meaningful for the government to grasp the current thoughts and concerns of the people. It is necessary to effectively evaluate the influence of the topic and study the influence of the topic on the society.

However, the openness and inclusiveness of the Internet make it difficult for users to obtain information. Some blogs contain important information, while others contain trivial and meaningless information. People are eager for information that can describe social dynamics. How can we timely obtain the centralized and organized social hot spots and understand the focus information in the vast network? This problem attracts researchers to put forward many effective solutions in the field of topic detection, such as using topic detection method to identify emerging topics in network information flow, and find hot topics through hotness evaluation. Topic hotness evaluation, which can be regarded as a filtering process of hot topics, is used for hot topic discovery, topic recommendation, and topic trend prediction. The hotness of the topic is used to measure the influence of the topic. Traditional influence evaluation methods only consider the frequency of news reports for hotness evaluation. In their opinion, the more news reports contained in one topic, the higher the attention, and the greater the hotness. Or they think that the hotness of the topic is mainly reflected in the number of comments and clicks. Although the influence evaluation method has been improved in some follow-on evaluation algorithms, such as TF-PDF [3–5] hotness evaluation algorithm, which only considers the influence of media on topic attention; literature [6] takes into account the users' attention to topics, but these influencing factors are not comprehensive enough and are not applicable to all fields. Therefore, we are more concerned about how to fully extract the effective information in blog posts and analyze which influencing factors could be quantified as the evaluation factors of hot topics.

Owing to the strong interactivity, fast propagation speed, and easy use of social networks, the network platform will produce a large amount of data every day, forming a lot of topics, which resulting in a rapid update of topic hotness, and the new hot topics will cover the old hot topics. When an event on the website may trigger massive relevant information, forwarding, and comments in a short time, the event will immediately become a hot topic. In addition, because the hotness of the topic varies in different periods of time, and new topics will continue to emerge in the same period of time, especially the topics related to judicial cases and political reports in microblog may last for a long time, but there are not many articles related to the topic produced during this period, so the hotness of the topic will be relatively small. However, if a topic produces extensive relevant articles in a short time, the topic will be more popular. Hot topics have a lifecycle, and they will go through the process from "generation" to "extinction". Therefore, we can analyze that interactive behavior, topic volume, and time are all significant factors affecting topic hotness.

In order to more accurately evaluate the popularity of topics, this paper proposes an evaluation model of real-time relative influence of topics in social networks (named ECANP), according to the initials of the five indicators---E (Engagement), C (Coverage), A (Activity), N (Novelty), and P (Persistence). Our model will analyze the law of increase and decline of topic hotness from five aspects: user engagement, topic coverage, topic activity, topic novelty, and

topic persistence, so as to better distinguish hot topics from cold topics.

The main contributions of this work are summarized as follows.

(i) We emphasize the relative influence and time attenuation of the topic. At the same time, multiple topics may appear at the same time, and the hotness ranking of topics depends on the relative influence between topics. In addition, topics have a lifecycle. With the passage of time, new topics gradually emerge and old topics are slowly replaced, so the time attenuation characteristics of topics are closely considered in the paper.

(ii) A new model for topic influence evaluation, ECANP, has been proposed, which comprehensively evaluates the relative popularity of topics by integrating user engagement, topic coverage, topic activity, topic novelty, and topic persistence.

(iii) Extensive experiments have been conducted on real datasets to prove the effectiveness of ECANP.

The remainder of this paper is organized as follows. Section 2 describes the related works. Section 3 presents the influence evaluation model in detail. Section 4 constructs experiments to verify the model and analyze the effectiveness of the model. Finally, conclusion is provided in Section 5.

2. Related Work

As an important channel of information dissemination and sharing, social networks bear the overload of information. Compared with the early era of lack of information, the biggest difference of social networks is that users can create information while browsing information. Moreover, the online social networking provided by the Internet almost imitates real life. Even if there is no contact between people, the purpose of information communication, sharing and dissemination can be achieved, and this way is more free and flexible. Due to the great participation of users, many subjective blog posts have been formed on the network, and continue to develop into topics, resulting in public opinion, which has a great impact on the public's point of view and life. In order to monitor public opinion in time, find hot topics and extract valuable public opinions from a large number of unorganized and complex data, numerous research scholars devote themselves to the research in this field.

Since 1996, the Defense Advanced Research Projects Agency (DARPA) proposed the concept of Topic Detection and Trace (TDT) for automatically judging the theme of news data flow without human intervention, which has attracted extensive attention, including well-known universities such as IBM Watson Research Center, BBN Company, and Carnegie Mellon University. Companies and research institutions have participated in the evaluation of subsequent TDT. Although domestic research on TDT started relatively late, since National Taiwan University participated in the evaluation of TDT topic detection task in

1999, Peking University, Chinese Academy of Sciences, and Fudan University began to study TDT-related technologies, and put forward many researches using topic detection and tracking to solve the tasks of topic clustering and hot topic discovery, which have achieved prominent results.

The emergence of TDT promotes the discovery and tracking of new events in news reports [7, 8]. The existing topic detection methods mainly focus on the methods based on machine learning [9–13]. For example, literature [14–16] proposes a topic detection method based on incremental clustering to automatically discover and track online news. In reference [17], Ma et al. tracked the online news topics based on the improved vector space model (VSM) model, extracted as characteristics of feature vectors through the lexical chain based on HowNet, and finally used the initial weight and structural weight of the features to construct the semantic vector space model.

Topic detection can undoubtedly help people quickly find out what topics people are discussing from a large number of online articles, but it cannot focus on hot topics, especially when there are many topics and time is limited, it is impossible to choose which part of the more important topics to participate in. Therefore, topic influence assessment is particularly important in hot topic detection tasks.

Hot topic detection is generally divided into two tasks: topic detection and hotness evaluation. Topic discovery depends on topic detection technology, and the generation of hot topics depends on the hotness evaluation method. As for how to measure the hotness of topic, different researchers have successively analyzed the factors affecting the evaluation of topic hotness from different aspects based on the concept of "hot topic", and all captured the characteristics of topic hotness to varying degrees.

Chen et al. [18] constructed a topic hotness evaluation model based on four factors: media attention, topic competition, topic intensity, and topic cohesion; Deng et al. [19] believed that opinions of comments represented the attitudes of different reviewers towards the topic, and comments with more opinions were hotter. Thus, they not only considered the number of reviews, comments, and publication time, but also took the comment opinion into account to evaluate the hotness of the blog topic; Li et al. [20] divided the factors affecting hotness into internal characteristics and external characteristics. Internal characteristics refer to number of clicks, reply, participating users and topic post, and external characteristics refer to duration time of topic, post source, number of released post, and topic quality. However, these methods do not take into account the characteristics of topic lifecycle.

Zhong [21] sorted out the characteristics of hot topics by analyzing the meaning of topics and hot topics, and proposed a method to extract hot terms that can represent hot topics from text documents based on the two key attributes of hot terms: persistence (the frequency with which a term appears in a set of documents) and topicality (the variation in the frequency of usage of a term over time). Through the distribution and change of hot terms in time, the clustering of hot topics in a given time period can be identified by weighting and applying $TF * PDF$ and aging theory,

respectively. Wang et al. [22] extracted the time attribute, report attribute, user attention, and other characteristic parameters of news reports, and established evaluation model of hot topics to evaluate the popularity of clustering topics. Although these methods consider the attenuation characteristics of topic hotness, they do not fully mine the attention of users.

Liu and Hu [22] introduced aging theory while concerned about the inconsistency between media focus and user focus. Assuming that the value of energy attenuation remains unchanged, they use the energy function to express the hotness. However, the lifecycle of different topics is different, and the attenuation degree with time is also different. Therefore, this assumption is unreasonable.

Based on the statistical idea and time characteristics, this paper comprehensively considers the attenuation characteristics of user attention and topics from five aspects, and proposes a new evaluation model of real-time relative influence of topics in social networks by integrating user engagement, topic coverage, topic activity, topic persistence, and topic novelty. The model not only analyzes the user's attention to the topic from the user behavior, but also considers that the influence of the topic is relative. Finally, it also concentrates on the time attenuation characteristics of the topic and considers the topic characteristics in a more fine-grained manner.

3. The Proposed Model

In this section, we first introduce the three important concepts involved in this paper. Then, we elaborate on the factors that determine the evaluation model. At last, we propose an influence evaluation model, called ECANP, and present the detailed components of ECANP.

3.1. Problem Definition. For ease of understanding, this section gives explanations of important concepts and lists the description of important symbols in Table 1.

Definition 1. TopicIn Topic Detection and Tracking (TDT) research, a topic is defined as a composition of core events or activity and events or activities directly related to it. An event is something that occurs at a specific time and place, involves some people or things, and may be accompanied by some inevitable results. Specifically, the topic is not a large field (such as the "national two sessions") or a certain type of event (such as the "Winter Olympics Games"), but a very specific "event", such as the "Fengxian event", and then the reports on the "Fengxian event" are considered to be related to the topic of "Fengxian event". In general, we can simply treat a topic as a collection of several reports related to an event [23].

Definition 2. Topic HotnessDifferent topics have different effects on the public. When the event forms a topic, some of which will receive more attention and discussion from people for a period of time, while others only get the

attention and participation of a few people. According to the different degree of attention and participation received by the topic, it is expressed as the different influence of the topic, which is quantified as the topic hotness. Topic that has a high hotness is called hot topic. Hotness is a relative concept. Ranking according to the hotness of the topic could get the ranking list of hot topics, so as to distinguish hot topics from the rest of the topics.

Definition 3. Influence EvaluationHow to quantify the impact of a topic is the focus of evaluation. Whether a topic is popular or not depends on the feedback from users. On the basis of the user's feedback on the topic, such as likes, forwards, comments, original microblogs, and other forms of engagement, quantifying the influence generated by the topic as hotness is the influence evaluation.

3.2. Key Study. The general process of topic hotness evaluation is shown in Figure 1. The general process consists of four layers: input layer, data layer, model layer, and output layer. The input layer clusters the data required by the model through topic detection and discovery technology, and each topic contains many blog posts belonging to the topic. The data layer is responsible for extracting the evaluation factors of calculation hotness for each topic. The model layer uses the proposed hotness evaluation model to calculate the hotness of each topic. Finally, the hot topic list is generated in the output layer according to the hotness value obtained from the topic.

Targeting the topic influence evaluation problem, we propose a new solution, called ECANP. As shown in Figure 2, the model is divided into five components: i.e., user engagement, topic coverage, topic activity, topic persistence, and topic novelty. We first discuss how the five indicators affect the topic hotness in Section 3.2.1, then present the specific evaluation model, and introduce the calculation details of each component in Section 3.2.2.

3.2.1. Evaluation Factors Analysis of the Topic Hotness. With the development and application of topic detection technology in research and industry, the massive amount of information on the Internet is integrated into orderly and classified modules, which is convenient for users to view other articles related to the topic of an article, that is, they belong to the same topic, so as to form their own opinions. But in reality, we find that even if information is aggregated into topics, the number of topics is still huge and various. If the topics can be sorted according to certain strategies, users will access the information they demand more efficiently and conveniently, thereby meeting the needs of users.

One strategy for ranking topics is to quantify the influence of topics at the topic level, that is, hotness, and rank them according to the hotness of topics. The hotter the topic, the higher the influence, thus the higher the ranking, and vice versa. Different from ordinary topics, hot topics usually have close user attention, wide coverage, high release frequency, and other characteristics. Therefore, we should first

TABLE 1: Symbols and description used in this paper.

| Symbols	Description
α	Weight of likes to user engagement
β	Weight of forwarding number to user engagement
γ	Weight of comment numbers to user engagement
L_j	Total likes of topic j
R_j	Total forwards number of topic j
C_j	Total comment number of topic j
Q_j	The number of articles related to topic j
Q	Total number of articles on all topics
T_j	Duration, that is, the time from the appearance to the end of the topic j
n_u	The number of time units in which the topic was reported like day or month
n	The number of time units segmented from the earliest release time of all topics to the acquisition time
$\nabla t(j)$	The number of time units between the current influence evaluation time and the topic release time
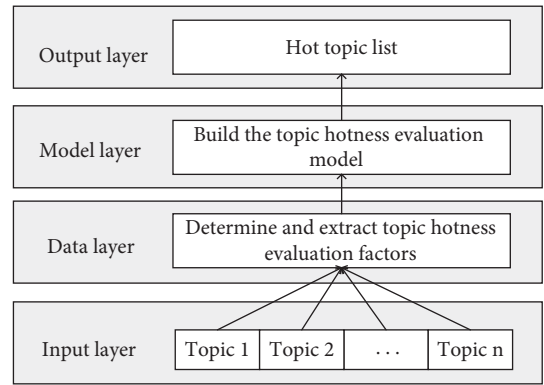
FIGURE 1: The general process of topic hotness evaluation.

determine some factors related to topic hotness, and establish a topic influence evaluation model by analyzing the different effects of various factors on topic influence. As follows, this paper defines five factors related to topic hotness evaluation at the topic level.

(1) User Engagement. In online social networks, the formation of hot topics is affected by many factors, but extensive user engagement is the foundation of forming a hot topic. Because only when a certain number of people browse, participate, and pay attention to, the topic will have a certain social influence, thereby attracting more people to be involved. In real life, through face-to-face contact and communication, a large number of people gather and participate in a certain place for a period of time, resulting in an influential activity. The network world is a reflection of the real world. Depending on the dissemination and sharing of the Internet, people can communicate freely across time and space. This is how social networks come into being. Different from the real social interaction, people's participation in online social networking sites is expressed as explicit participation and implicit participation. Publishing original articles, forwarding articles, and liking or commenting on articles are explicit participation behaviors. Browsing and
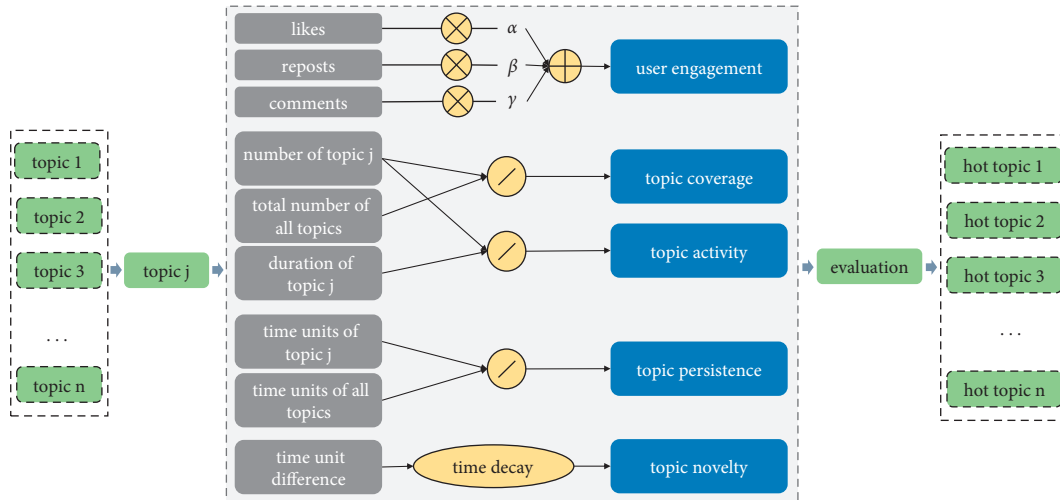
FIGURE 2: The overall framework of ECANP for topic hotness evaluation.

searching related content are the implicit participation behaviors. From the perspective of topic, the explicit feedback and implicit feedback they receive exactly reflect user's attention and participation in the topic. The like behavior shows that the liked content can attract user's attention, cause their resonance, and reflect their favor and appreciation of the content. When users are interested in a blog post, have some opinions, and want to get more information related to content, they tend to give comments, which reflects the user's awareness and interest in the content. When users are approve of the content they are interested in, they will be prompted to forward the content, so that the content can obtain more exposure and share with more people, reflecting users' recognition of the content.

Due to users' explicit feedback is easy to obtain, this paper measures users' engagement through their likes, forwards, and comments on topic articles. The more likes, forwards, and comments, the more attention user pays to the topic, and then topic will be widely spread owing to user's forwarding behavior. Due to user's comment behavior, more users will be appealed to participate in the discussion of the topic, thus expand the dissemination and influence of topic.

(2) Topic Coverage. A topic is a collection of seed events and related events. Hot topics usually have a broad user base. They will publish many original articles related to the topic, which makes the number of articles contained in the topic continue to increase. The number of articles contained in a topic reflects the hotness of topic to a certain extent. Compared with the number of articles related to other topics in the same period, the number of articles related to the topic reflects relative influence of the topic. Generally speaking, the topic which has more relevant articles is hotter. The more the number of articles related to a topic accounts for total number of articles on all topics at the same time, indicating that the topic has greater influence than other topics. For instance, if there are 1000 related articles on topic A from its emergence to its demise, but there are 10000 related articles

on topic B, we believe that topic A is more popular than topic B in terms of topic coverage.

It can be seen that the number of articles related to topic is an important factor to measure influence of topic. Therefore, the coverage of a specific topic in all topics is of great significance to quantify the hotness of topic.

(3) Topic Activity. Articles on social networks (such as microblog) can be regarded as a text stream on the timeline. When a blog post is published, with the attention and participation of users, a certain number of relevant articles will be generated in succession over a period of time to form a topic. If this time period is short, a topic generates a large number of relevant articles, while another topic only generates a small number of articles, the topic with generous relevant articles in a short time will get more attention and higher hotness. Or in another case, if a topic produces many articles, but it is distributed over a long period of time, and the average number to the time unit is small, the hotness value will be smaller than that of the topic that produces many relevant articles in the short term. That is, the more relevant articles on topic are published per unit of time, the more its influence can be reflected. Therefore, as a hot topic, we should not only consider the proportion of the number of articles related to topic, but also take the activity per unit time into account.

Generally speaking, in the whole time period, the more frequently a topic is discussed, the more active the topic is, the more relevant articles it has compared with other topics, and the greater its influence. Therefore, the number of articles produced by the topic per unit time is also one of the major factors affecting topic hotness, thus we can acquire the activity of the topic.

(4) Topic Novelty. Hot topics are those topics that are frequently discussed and concerned by the public for a period of time and within a certain range. The hotness of topic will accumulate with the increasing attention of users and media, otherwise it will gradually decay over time, which is

consistent with the life cycle of the topic modeled by Liu and Hu [24] based on aging theory. Affected by the "life cycle", hot topics will go through a process from "generation" to "extinction". Accordingly, their hotness will change with the change of life cycle and eventually decline naturally. In addition, people prefer new topics and current events to old ones. In particular, with the migration of time, new topics gradually emerge and attract users' new attention, naturally, fewer and fewer people pay attention to the old topics, and the number of articles related to them also decreases. However, more and more attention is paid to the new topics, resulting in the old topics being gradually replaced by the new topics, and gradually fade out of people's memory, and people turn to pay more attention to the development of new topics.

Therefore, the earlier the topic first appears from the current time, the smaller its impact on users and the smaller its hotness value. On the contrary, the closer the topic appears to the current moment, the more active it is, and the more it can draw the attention of users, the higher its contribution to hotness value, and the more likely it is to become a hot topic. Hence, the novelty indicator of topic is obtained by using the attenuation function with number of time unit intervals between current time and first release time of topic.

(5) Topic Persistence. Traditional topic hotness evaluation model tends to consider the impact of media attention (i.e., the number of relevant reports) and user attention (i.e., user clicks and participation) on the hotness. Recently, more researchers have considered the characteristics of topic life cycle, but still neglected another property of topic itself, that is topic persistence. A topic always develops with time, and it is not easy for an event to become a topic overnight, which requires magnanimous users' long-term discussion and participation. The longer a topic is discussed and concerned by users, the more it can arouse users' interest, and the more likely it is to attract more users to participate. Nevertheless,

some topics are not necessarily discussed every day and have nodal property. Such as the "jiangge incident" that has lasted until now in 2016, which will appear again on the social platform and become a hot topic whenever there is new progress in the case. Consequently, the total number of time units in which topic is continuously discussed in specified time period will be acted as an important indicator to measure topic hotness in this paper.

User engagement is used to calculate the user influence related to topic. Topic coverage is used to calculate the propagation coverage of related topics. Topic activity is used to calculate the activity of related topics. Topic novelty is used to calculate the contribution value of the novelty of related topics to hotness. And topic persistence is used to calculate the time when relevant topics are continuously active on the social platform. Based on the above analysis, the newer the topic, the more public participation and discussion, the wider the coverage, the higher the activity, and the longer the duration, and the more likely topic is to have a high hotness value and become a hot topic. We will propose topic hotness evaluation model in Section 3.2.2, and verify the model through experiments in Section 4.

3.2.2. Topic Hotness Evaluation Model: ECANP. Based on the factors related to topic influence evaluation analyzed in the previous section, it can be concluded that user engagement, topic coverage, topic activity, topic persistence, and topic novelty can be used as evaluation factors to quantify a topic influence. Following the above indicators, multiple topics will be generated successively in a period of time, and the relative influence of each topic is quantified as the hotness evaluation. Because topic coverage, topic activity and topic persistence belong to topic attributes, besides, topic attributes, and user participation will gradually weaken with the passage of time, so the operation of influence evaluation integrating five indicators can be formulated as follows:

$$
\text{Hot}_j = \left(\begin{array}{c} \text{Engagement}(j) + \text{Coverage}(j) * \text{Activity}(j) \\ * \text{Persistence}(j) \end{array} \right) * \text{Novelty}(j), \tag{1}
$$

where Engagement (j) denotes the user engagement in topic j. We first employ the entropy weight method to determine the weight of each indicator affecting user participation, and then apply the sum aggregator to aggregate the impact of the three indicators for expressing the influence of users' participation behavior on the topic. Utilizing entropy weight method to determine the indicator weight can be divided into two steps: data standardization processing and entropy weight determination of the indicator.

(1) Data Standardization Processing. We consider that there are m topics in a period of time, and the user engagement factor of each topic has n measurement indicators. Let

$R \in R^{m \times n}$ denote the judgment matrix, and build R before standardizing the data:

$$
R = \left(r_{ji} \right)_{m \times n}, \quad (j = 1, 2, \ldots, m; i = 1, 2, \ldots, n). \tag{2}
$$

Then, for the sake of eliminating the adverse effects caused by singular sample data, we employ maximum and minimum normalization to standardize the judgment matrix R to acquire the standard data limited in the range of $[0, 1]$:

$$
b_{ji} = \frac{r_{ji} - r_{\min}}{r_{\max} - r_{\min}}, \tag{3}
$$

where

$$r_{\min} = \min(r_{1i}, i, \ldots, r_{mi}), \tag{4}$$

$$r_{\max} = \max(r_{1j}, r_{2j}, \ldots, r_{mj}). \tag{5}$$

In equation (2), m represents the number of topics, n denotes the number of indicators to measure user engagement, and r_{ji} means value of the i-th indicator of the j-th topic;

In equation (3), $r_{\max}$ and $r_{\min}$ represent the maximum and minimum values of the number of likes, forwards, and comments of the relevant articles on different topics under the same measurement indicator, respectively. And b_{ji} stands for maximum and minimum normalization value of r_{ji}.

(2) Entropy Weight Determination of the Indicator. We define the weight of all likes, forwards, and comments of each topic as follows:

$$W = (w_i)_{1 \times n}, \tag{6}$$

where

$$w_i = \frac{1 - H_i}{n - \sum_{i=1}^{n} H_i}. \tag{7}$$

According to the definition of entropy, the weight of all measurement indicators can be determined:

$$H_i = -\frac{1}{\ln m} \sum_{j=1}^{m} f_{ji} \ln f_{ji}, \quad (0 < H_i < 1), \tag{8}$$

$$f_{ji} = \frac{1 + b_{ji}}{\sum_{j=1}^{m} (1 + b_{ji})}, \quad (j = 1, 2, \ldots, m; i = 1, 2, \ldots, n), \tag{9}$$

where f_{ji} denotes the proportion of the j-th indicator in the i-th topic.

Through analysis, we can know that if the value of f_{ji} in equation (8) is 0, ln 0 will inevitably occur. For solving this problem, our paper adopts the following formula amend f_{ji}'.

$$f_{ji}' = \frac{b_{ji}}{\sum_{j=1}^{m} (b_{ji})}, \quad (j = 1, 2, \ldots, m; i = 1, 2, \ldots, n). \tag{10}$$

In line with above contents, the weight of each measurement indicator under each topic can be calculated. Then, the final user engagement can be obtained through the sum aggregator.

Based on the analysis of topic influence factors in Section 3.2.1, we finally determine that user engagement is affected by three indicators: likes, forwards, and comments. Therefore, this paper calculates the weight of these three indicators and aggregates them through equation (11):

$$\text{Engagement}(j) = \alpha \times L_j + \beta \times R_j + \gamma \times C_j. \tag{11}$$

Then, coverage (j) denotes the topic coverage in topic j, which is calculated based on the proportion of the number of articles related to topic j in all topic articles, that is,

$$\text{Coverage}(j) = \exp \frac{Q_j}{Q}. \tag{12}$$

Next, activity (j) denotes the topic activity in topic j. By calculating the number of articles on topic j in unit time, we can obtain the activity of the topic j, that is,

$$\text{Activity}(j) = \ln \frac{Q_j}{T_j}. \tag{13}$$

Similarly, novelty (j) denotes the topic novelty in topic j. According to the difference of time units between current time and the time when the topic j was first published, the novelty indicator of topic j is obtained:

$$\text{Novelty}(j) = (\nabla t(j) + 1)^{-k}, \tag{14}$$

$$\nabla t(j) = t_t - t_p, \tag{15}$$

where t_t is current time (for example, if the collected experimental data is from April 1 to April 30, 2020, the current time is April 30), t_p is the first release time of topic j, and $\nabla t(j)$ is the difference of time units between the current time and the first release time of the topic, time in days. k is the attenuation factor, which controls the attenuation rate of topic j over time. The larger $\nabla t(j)$, the smaller novelty (j), and the less contribution of this indicator to hotness.

Finally, persistence (j) denotes the topic persistence in topic j. According to the duration of the topic j in the life cycle and the proportion of the number of units in the whole topic monitoring time, the persistence of the topic is obtained:

$$\text{Persistence}(j) = \frac{n_u}{n}, \tag{16}$$

the total duration n_u of topic j is obtained by equation (17), that is,

$$n_u = n_e - n_b, \tag{17}$$

where n_u is the number of time units in which topic j is reported and discussed, time in days. n_e indicates when the topic dies, and n_b represents when the topic arises.

Further, the influence evaluation of topic j in equation (1) can be obtained by aggregating the above five indicators. Therefore, the hotness value Hot_j of topic j is described as follows:

$$\text{Hot}_j = \left((\alpha \times L_j + \beta \times R_j + \gamma \times C_j) + \exp \frac{Q_j}{Q} \times \ln \frac{Q_j}{T_j} \times \frac{n_u}{n} \right)$$

$$\times (\nabla t(j) + 1)^{-k}, \tag{18}$$

where k is the attenuation coefficient.

3.2.3. Interpretability Discussion of the Model.
Entropy weight method is an objective method to determine the weight, which has certain accuracy compared with subjective methods such as analytic hierarchy process. Moreover, the

weight value determined by this method could be modified, which determines its high adaptability. The formula for calculating entropy value in entropy weight method was put forward by information scientist Shannon. When the data is more dispersed and the entropy is smaller, it can be considered that the data contains more information, so the weight is larger. According to Section 3.2.1, a topic will get users' likes, forwards, and comments. These behaviors represent users' different degrees of preference for the topic and reflect users' engagement. Therefore, in order to identify the contribution of user behavior factors to users' engagement, entropy weight method is used to calculate their weight. After that, user engagement can be obtained by weighted summation of behavior factors.

In addition to considering users, the attributes of the topic itself, including text attributes and time attributes, should not be ignored. The text attributes of topic include topic coverage and topic activity, and the time attributes comprise topic novelty and topic persistence. In a series of blog posts, the more blog posts related to the topic, the wider their coverage, which is expressed as the proportion of the number of blog posts related to the topic in the total blog posts, i.e., Q_j/Q. Over a period of time, the more blog posts related to the topic, the higher their frequency, and the easier it is to catch the user's eye, which is expressed as the number of blog posts related to the topic in unit time, i.e., Q_j/T_j. Generally, $(Q_j/Q) \in (0, 1]$.

The topic of sustainability that people pay attention to can gain higher hotness. The longer the topic lasts, it means that people are more interested in the topic; the topic continues to ferment and has a greater impact on more users. We use the proportion of the number of time units in which the topic is reported to the total time units to express topic persistence, i.e., n_u/n.

Topics are always updated iteratively. "From emergence to extinction" is a process that every topic will experience. Over time, the contribution of each attribute of the topic to its hotness is gradually weakening, as described by the aging theory [22]. The attenuation degree of topic determines the novelty of topic, which depends on two factors, including the time span of topic and the attenuation factor. For the attenuation function, we choose the inverse proportional function of time factor, because its value range is larger, the attenuation degree of new topic and old topic is clearly distinguished, and the attenuation is stable.

4. Experiments and Analysis

We testify our proposed model by conducting extensive experiments on real-world corpus, aiming to answer the following key questions.

(i) Q1: How can the topic influence evaluation model effectively explain the topic hotness?

(ii) Q2: How does ECANP perform compared with state-of-the-art influence evaluation models?

(iii) Q3: Can five indicators (i.e., user engagement, topic coverage, topic activity, topic persistence, and topic novelty) reasonably explain the impact on ECANP?

TABLE 2: The correspondence between topic name mark and real topic name.

Topic name mark	Real topic name
Topic 1	The murderer of the passion fruit girl case was commuted to death
Topic 2	Murder with a knife in Kaiyuan, Liaoning
Topic 3	Peking University Wu Xieyu's mother killing case opens
Topic 4	The case of Lao Rongzhi in Nanchang, Jiangxi Province was opened
Topic 5	Henan "9-year-old Lao Lai case"
Topic 6	Lai Xiaomin was sentenced to death in the first instance

4.1. Datasets. To demonstrate the performance of our method ECANP and compare it with the baseline methods, a corpus of microblog articles is adopted. Six topic data are used for experiments to verify the universal applicability of the model ECANP in this paper. The topic names are shown in Table 2. The data in the datasets comes from the judicial cases provided by Yifang, and the time range of topic is from December 14, 2020 to January 11, 2021, with an overall data volume of more than 50000 pieces. This paper evaluates the influence of topic by utilizing the six topics that users have participated in for a long time provided by Yifang and the analysis of the hotness of each topic. Before using the data, we check the dataset through conventional data preprocessing method to remove the unusable or invalid data. The specific data statistics are shown in Table 3.

4.2. Experimental Settings

4.2.1. Evaluation Metrics. Since there is no unified evaluation indicator for topic influence evaluation, in order to prove the effectiveness of the model, this paper consulted a large number of relevant literature. Inspired by [19], we finally determined to carry out experimental verification from three aspects, namely the effectiveness of model (Effectiveness verification, abbreviated as EFVC), the comparison of ability to distinguish hot and cold topics with the baseline evaluation methods (Ability to distinguish hot and cold topics, abbreviated as ADHCT), and the impact of each evaluation indicator on the model (Control variable analysis, abbreviated as CVA), corresponding to the above three questions.

4.2.2. Baselines. To support the effectiveness of ECANP model, we compare it with the following five baseline models, in which the first four models are verified with the same and only one dataset, and the fifth is verified with the six datasets used in this paper due to the particularity of its method. In addition, some comparison models do not have a name, for ease of display, we give a name according to the naming method in this paper. The experimental settings of proposed model and baselines are introduced in the next subsection.

TABLE 3: Statistics of the datasets.

Topic	Forwards	Comments	Blogs	Start time	End time	Duration (days)
Topic 1	55707	43621	8288	2020/12/28 0 : 00	2020/12/31 23 : 59	4
Topic 2	6785	11576	3520	2020/12/27 0 : 00	2020/12/31 23 : 59	5
Topic 3	16311	32170	6088	2020/12/23 0 : 00	2020/12/30 23 : 59	8
Topic 4	26780	70562	15371	2020/12/21 0 : 00	2020/12/30 23 : 59	10
Topic 5	17966	31265	11894	2020/12/14 0 : 00	2020/12/20 23 : 59	7
Topic 6	15971	20406	6578	2021/1/5 0 : 00	2021/1/10 23 : 59	6

(i) BHEM-TOA [20]: This is a blog hotness evaluation model based on comment opinion analysis, which realize blog hotness evaluation through the number of reviews, comments, publication time, and the opinioned comments.

(ii) FSTCC [1]: Such a model is proposed to calculate the hotness value of online news topics about the emergency events, which considers reporting frequency of topics, the number of report sources, time property, click rates of users, and the number of comments.

(iii) HFTC [24]: This is a method to evaluate topic hotness by exploiting the frequency of topic tags. Specifically, it takes several keywords with the highest probability of occurrence in topic as the tag set, calculates topic hotness by using topic tag frequency without relying on any information other than the text itself, and finally determines the latent topic with the highest hotness value as hot topic in the unified network.

(iv) IEFE [21]: This is a topic hotness evaluation model, where considers the internal and external factors impacting the hotness. The characteristics of hot topics are analyzed by internal factors such as number of clicks, comments, and user participation, as well as external factors such as topic duration, topic quality, and topic concentration.

(v) DMCBF [19]: This model is based on decay, media attention, topic competition, and topic cohesion. The hotness value of each day is calculated through the energy function, and the accumulation of hotness value of each day is regarded as the accumulated hotness of the topic after d days.

4.2.3. Experimental Settings. For the three experiments in the next section, we adopt different settings and data processing methods.

For the first experiment, we implemented our ECANP model in six experimental datasets, presenting the hotness results of each topic, and the results of some influencing factors.

For the second experiment, the settings of comparison models are divided into two categories. To realize comparison between our proposed method and the first four models, we selected Topic 1 in datasets for experiment. Firstly, we clustered multiple subtopics and topic names under Topic 1 through topic detection method and topic name detection, and then ran ECANP and four baseline models respectively to calculate the hotness of each detected subtopic. For comparability, the hotness of all models is normalized in the range of 0–100. The normalization formula is equation (19). Due to the particularity of the fifth comparison model, we regard each dataset as a separate topic, divide each topic according to the number of days, and get the relevant articles of the topics in each day. ECANP and DMCBF are used to calculate the hotness of six topics respectively. Finally, we select the hotness value of the five hottest topics and the five coldest topics obtained by the first four models, as well as the hotness value of all topics of the last model, contrast ECANP model with them respectively, and compare the performance according to the judgment formula of hot and cold topic discrimination ability:

$$\text{Hot}_j\_\text{norm} = \frac{\ln \text{Hot}_j}{\ln \max \left(\text{Hot}_1, \text{Hot}_2, \ldots, \text{Hot}_j \right)}. \quad (19)$$

For the third experiment, to clearly see the change of topic hotness and its indicators over time, we divided the life cycle of topic into 8 time periods, obtained the articles related to the topic in each time period, and took the last time point of the time period as the coordinate label.

The settings of each baseline model are as follows.

BHEM-TOA: we removed the part of text comments in this baseline method, because the author evaluates the hotness of Blog websites, which is composed of the hotness of multiple topics. Multiple topics under a website do not distinguish comments, so it is set as a constant in this paper, and its value is 0.

FSTCC: different hotness evaluation methods use different indicators for different scenarios. Owing to there is no report sources in our datasets used in this baseline method, such information is ignored. In addition, the time interval used by the author is the difference between the current time and the topic publishing time, but the experimental results show that the publishing time of the topic in the dataset is too long from the current time, and the time attenuation is quite large, resulting in the hotness value of 0. Therefore, this paper uses the duration of the topic instead of the current time in the paper. The time interval is set to 1 day.

HFTC: the author applies the proposed model to cross social networks. To compare with our method, only the single platform topic hotness evaluation in author's paper is used.

TABLE 4: Results of algorithm model.

Topic name	Hotness value	Ranking	Top	Ranking list of hot topic	Hotness value
Topic 1	100	1	Top 1	Topic 1	100
Topic 2	83.35	6	Top 2	Topic 4	99.64
Topic 3	92.85	4	Top 3	Topic 5	92.99
Topic 4	99.64	2	Top 4	Topic 3	92.85
Topic 5	92.99	3	Top 5	Topic 6	90.11
Topic 6	90.11	5	Top 6	Topic 2	83.35

IEFE: in the implementation of this method, since there are no number of clicks in our datasets, this item and the publishing sources are ignored, and the number of microblog articles is used to replace the number of user participation.

DMCBF: the time used in the model is in days.

In equation (18), the attenuation coefficient is set to 0.1.

4.3. Experimental Results and Analysis

4.3.1. Result 1: Validity Verification (Answer the Mentioned Q1). Based on the statistical experimental data in Section 4.1, we use the topic hotness evaluation model proposed in this paper to evaluate the hotness of all topics in the dataset, calculate the hotness value of each topic, and rank them according to their hotness value. As shown in Table 4, it presents the model results and topic ranking.

Due to the long real name of the topic, for convenient representation and viewing, the topic name is listed in the form of "topic + number" in the charts in this paper. The actual correspondence between topic name mark and real topic name is shown in Table 2.

In Table 4, the left part shows the hotness value of each topic and its corresponding ranking results, and the right part lists the topic ranking results from high to low according to the hotness value. From the above results, we can conclude that among the six groups of experimental data, Topic 4 has the highest hotness, while Topic 2 has the lowest hotness. Since the hotness of topic is relative over a period of time, the influence of topic with the highest hotness is quantified as 100. By sorting the relative influence of topics, the topic ranking table is obtained.

In this part of the experiment, to validate the effectiveness of our topic influence evaluation model, we evaluated and analyzed the hotness values of six topics. ECANP first extracts the number of forwards, comments, related blog posts, topic duration, hotness evaluation time period, and current time of each topic, then utilizes these information to calculate the hotness of each topic through each indicator calculation method and hotness evaluation formula proposed in Section 3.2.2, and finally presents the histogram of the number of forwards, comments, and blog posts of each topic in Figure 3; meanwhile, the hotness value of each topic is also shown in Figure 3 as a line chart. It can be seen that the change trend of topic hotness value is consistent with the trend of forwards, comments, and blog posts. For example, on the whole, the number of forwards, comments, and blog posts of Topic 2 is less than that of Topic
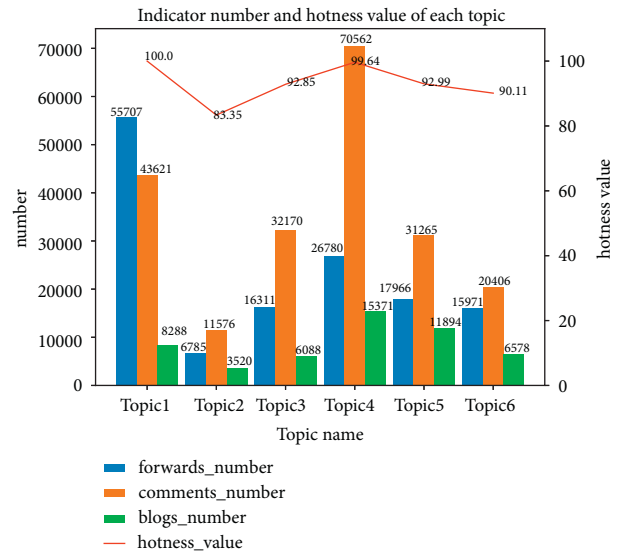

FIGURE 3: Comparison of the number of some indicators and hotness value of different topics.

1, accordingly, the hotness value of Topic 2 is also less than that of Topic 1, and so on; from the perspective of single factor, the change trend of forwarding number from Topic 1 to Topic 6 is "down-up-up-down-down", correspondingly, the change trend of topic hotness is the same, and other single factor analysis of topics is followed by analogy. In particular, the number of comments of Topic 4 is greater than that of Topic 1, but the hotness value of Topic 4 is less than that of Topic 1, which is due to Topic 1 is more novel than Topic 4. Combined with entropy weight method and hotness evaluation formula, it is concluded that the hotness value of Topic 1 is slightly higher than that of Topic 4.

The ranking of hot topics given by ECANP model is the same as expected, which shows that the topic hotness evaluation model proposed by this paper is reasonable and effective.

4.3.2. Result 2: Performance Comparison (Answer the Mentioned Q2). In this section, we contrast the performance of our ECANP with five baselines, compare their ability to distinguish hot topics from cold topics. Note that, for the baseline BHEM-TOAm FSTCC, HFTC and IEFE, we conduct experiments on the same topic, regard a topic as an event, use the clustering algorithm to obtain the subtopics under topic, and calculate and distinguish the hotness of subtopic. The experimental flow is shown in Figure 4, and
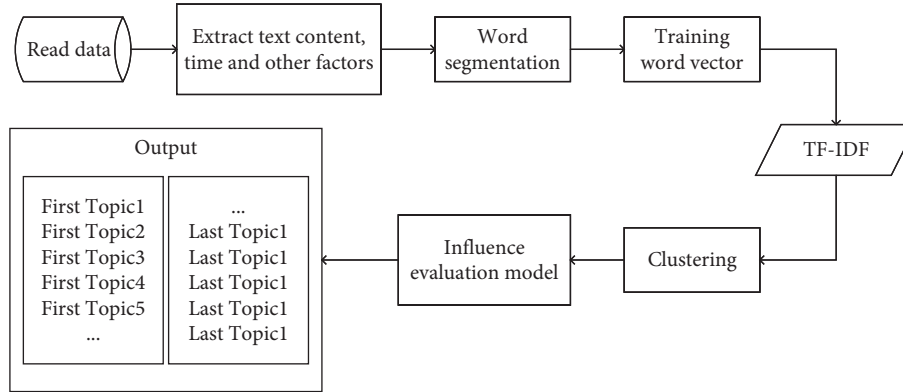
Figure 4: Data processing flow in comparison test.

Table 5: Performance comparison of our model and baselines (part one).

Topic label	Topic description	ECANP (ours) Hotness	Topic description	BHEM-TOA Hotness
HT1	①	100.0	①	26.4781
HT2	②	99.9504	②	6.1957
HT3	④	86.5919	③	3.1946
HT4	③	77.0561	④	1.9596
HT5	⑤	29.0486	⑤	1.1666
CT1	⑨	0.0004	⑨	0.0009
CT2	⑩	0.0004	⑩	0.0009
CT3	⑪	0.0004	⑪	0.0009
CT4	⑫	0.0004	⑫	0.0009
CT5	⑬	0.0004	⑬	0.0009
Dist (HT, CT)		0.9999908339562471		0.9999070183251324

Table 6: Performance comparison of our model and baselines (part two).

Topic label	Topic description	FSTCC Hotness	Topic description	HFTC Hotness	Topic description	IEFE Hotness
HT1	①	28.694	①	99.8995	①	99.9984
HT2	③	25.8758	②	15.355	②	97.0389
HT3	②	13.0456	③	15.0499	④	57.737
HT4	⑧	3.7081	⑦	7.8471	③	51.934
HT5	⑲	3.4714	⑳	7.2633	⑤	18.9157
CT1	⑭	0.0905	⑨	0.1	⑫	0.021
CT2	⑩	0.0905	⑭	0.1	⑮	0.021
CT3	⑪	0.0905	⑩	0.1	⑯	0.021
CT4	⑬	0.0905	⑪	0.1	⑰	0.021
CT5	⑫	0.0905	⑬	0.1	⑱	0.021
Dist (HT, CT)		0.991984356		0.99724503648		0.99947001200

then reports the performance of these baselines on the topic. For the DMCBF model, we use the six topics used in this paper to verify, then calculate the hotness of each topic, and report their performance.

Tables 5–8 indicate the experimental results of our method and five baselines, respectively. The table shows the results using equation (20) to normalize, where in Tables 5 and 6, α of the ECANP, FSTCC, and HFTC model is set to 0.001, the BHEM-TOA model is set to 0.1, and the IEFE model is set to 0.0001. In Tables 7 and 8, α of the ECANP is

Table 7: The result of our model and DMCBF.

Topic name	ECANP	DMCBF
Topic 1	19.338433668350692	88.19006162822104
Topic 2	3.582932515952112	27.3980491541227
Topic 3	9.41422561422112	79.2579948124188
Topic 4	18.66498030322313	74.4511671473541
Topic 5	9.554358308112763	99.90728389941798
Topic 6	7.128933737362833	31.796818192414495

TABLE 8: Performance comparison of our model and baselines.

	ECANP		DMCBF	
	Ranking	Hotness	Ranking	Hotness
HT1	Topic 1	19.338433668350692	Topic 5	99.90728389941798
HT2	Topic 4	18.66498030322313	Topic 1	88.19006162822104
HT3	Topic 5	9.554358308112763	Topic 3	79.2579948124188
CT1	Topic 3	9.41422561422112	Topic 4	74.4511671473541
CT2	Topic 6	7.128933737362833	Topic 6	31.796818192414495
CT3	Topic 2	3.582932515952112	Topic 2	27.3980491541227
	Dist (HT, CT)	0.276654078917006	Dist (HT, CT)	0.2079302077825828

0.00001, and α of the DMCBF is 1. As shown in Tables 5, 6, and 8, for the topic hotness evaluated by different methods, we use equation (21) to calculate the hotness distance [25] between topics.

Table 9 shows the subtopic names generated in the comparative experiment between ECANP and BHEM-TOA, FSTCC, HFTC, and IEFE. It can be seen from the results that the performance of our method in distinguishing hot topics from cold topics is better than the baselines, exceeding the performance of the optimal baseline model by 0.008%. This is because we make good use of indicators of five dimensions and effectively combine user characteristics and topic attributes. The results show that our method can better identify hot topics and cold topics, making hot topics more popular and cold topics less popular.

$$\text{norm\_Hot}_j = \frac{e^{\text{Hot}_j \times \alpha} - e^{-\left[\text{Hot}_j \times \alpha\right]}}{e^{\text{Hot}_j \times \alpha} + e^{-\left[\text{Hot}_j \times \alpha\right]}}, \tag{20}$$

$$\text{Dist}(\text{HT, CT}) = \frac{\sum_{i=1}^{5}\left(H^{ht(i)} - H^{ct(i)}\right)^2}{\sum_{i=1}^{5}\left(H^{ht(i)}\right)^2 + \sum_{i=1}^{5}\left(H^{ct(i)}\right)^2}, \tag{21}$$

where H^{ht} is the hotness of hot topic ht and H^{ct} is the hotness of cold topic ct.

4.3.3. Result 3: Case Study of ECANP (Answer Q3).

Obviously, the user attention and topic attributes in different time periods are not invariable, which enable the hotness of each topic varying with time. In order to clearly show that the change of topic hotness over time is affected by relevant influencing indicators, Figure 5 displays each relevant indicator value and topic hotness of each topic in different time periods of its life cycle. The figure (a–f) reflects the change trend of each indicator and hotness value of six topics in the form of broken line diagram, respectively. Note that the life cycle of each topic is different, and some have long life cycles and some have short life cycles. If the results are displayed at a unified time interval, the graph will be very unsightly and affect the intuition. Therefore, in order to facilitate viewing the results, each topic adopts the same number of time nodes, that is, the time interval of each topic is different, which does not affect the experimental results.

As shown in Figure 5, for each topic, the values of each indicator and topic hotness change in each period. Taking (c) in Figure 5 as an example, the hotness value of Topic 3 at the

TABLE 9: The correspondence between topic label and real subtopic name.

Subtopic name	Topic label
The murderer of the passion fruit girl case was commuted to death	①
Yang Guangyi, the defendant in the passion fruit girl case, was sentenced to death	②
Passion fruit girl's mother responded to the murderer's death sentence	③
Niu Bo, Chaohua	④
Sulfuric acid boy	⑤
Record the process of China's rule of law society bit by bit	⑥
Short comment on court newspaper	⑦
'LOEWE'	⑧
Xiao He said something	⑨
Yan Mu 1	⑩
Yan Mu 2	⑪
Nothing 1	⑫
Yan Mu 3	⑬
What can you do	⑭
Nothing 2	⑮
Nothing 3	⑯
Nothing 4	⑰
Nothing 5	⑱
Guangxi Yang Guangyi's rape case was retried and sentenced to death	⑲
Nanjing Denghuang 728 extravagant lighting show	⑳

first time point is 0, while the hotness value rises in the next time period, which means that the topic is still in the embryonic stage at the first time point, and then it obtains extensive user attention and participation, and reaches the hottest at the second time point. With the passage of time, the user attention and the value of topic attributes decrease, resulting in the decrease of topic hotness. At the sixth time node, the trend of user engagement, topic persistence, topic coverage, and topic activity have increased, and the trend of topic hotness is also rising. This shows that proactive user participation, lasting and active topic discussion, and extensive topic coverage will bring about high hotness of a topic and produce great influence. Although the novelty (purple line) and persistence (red line) of Topic 3 gradually increase as time goes on, the change range is very small, so it has little impact on the hotness of the topic. (Note, in the figure, since the values of these two indicators are close, the two lines almost coincide.) On the contrary, if the topic
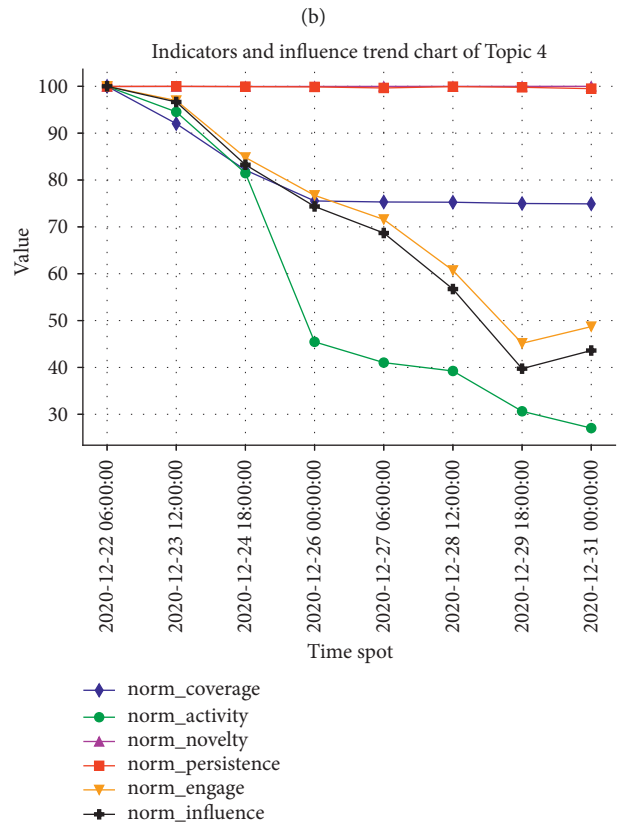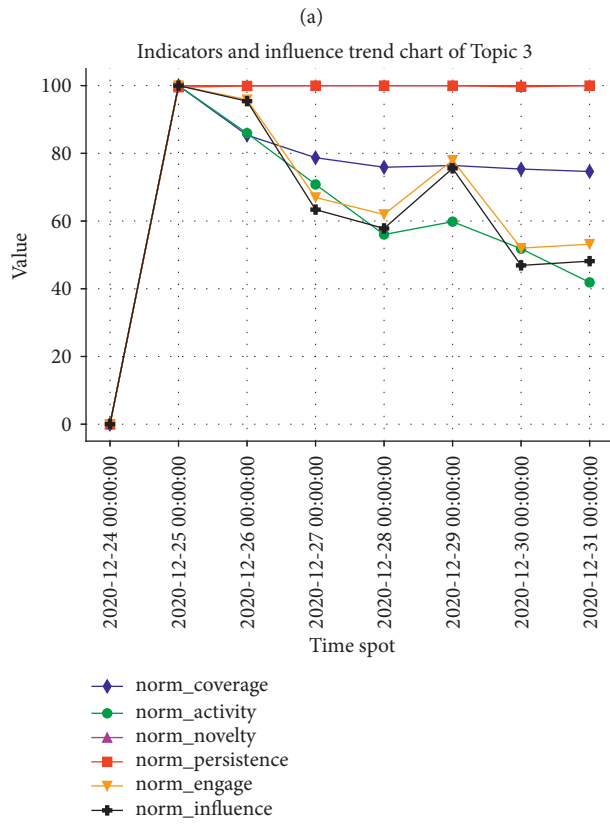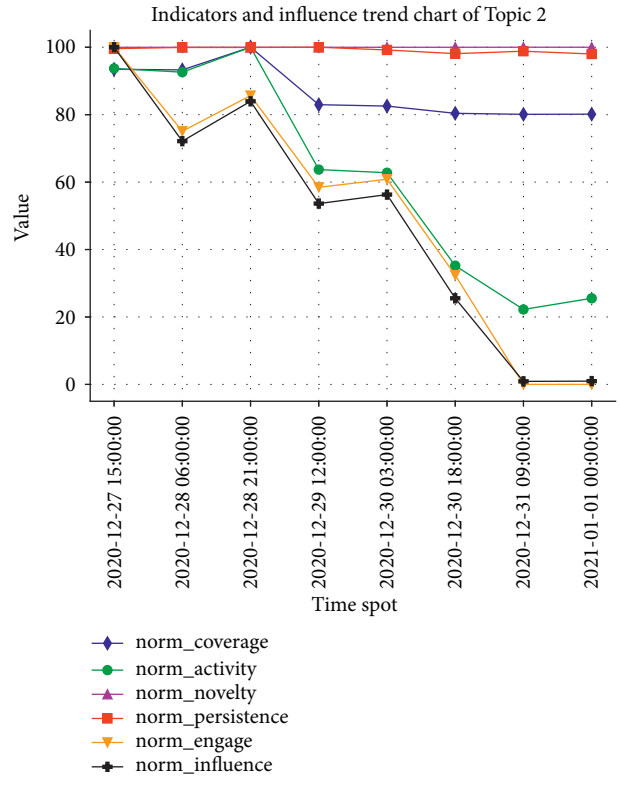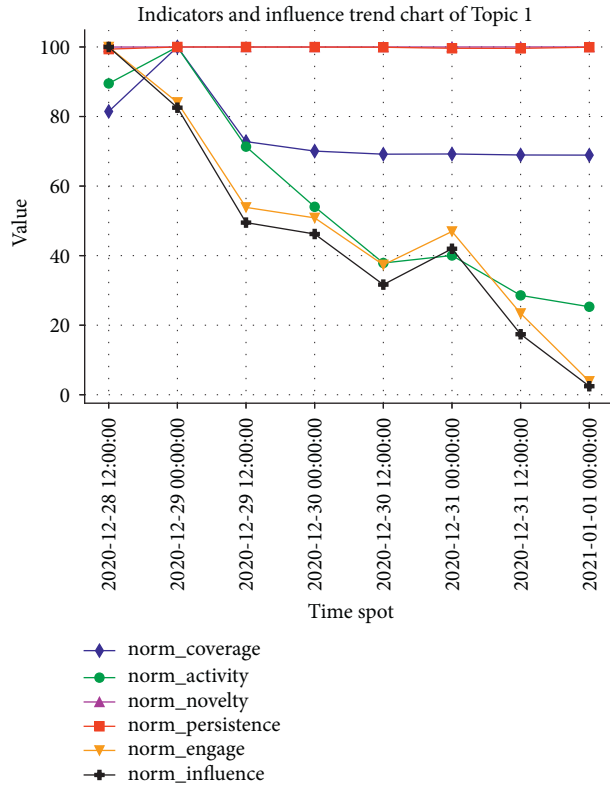
(a)
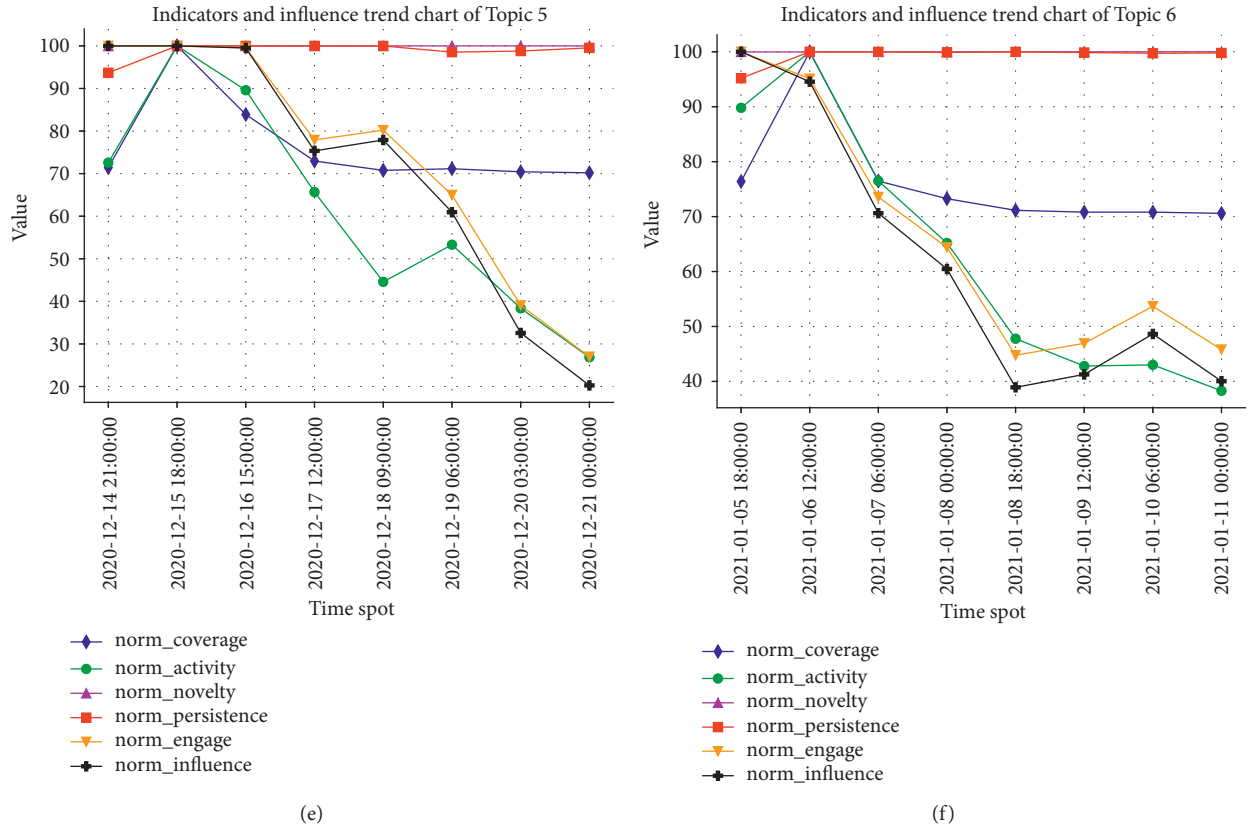
(b)

(c)

(d)

FIGURE 5: Continued.

Figure 5: The trend chart of topic indicators and hotness over time.

novelty and topic persistence change greatly in different time periods, it will have a great impact on the topic.

The hotness value of topic is jointly determined by various factors such as the forwarding, comments, and publishing time of the topic-related articles, including user characteristics and topic attributes. Due to the large base of users' forwarding, comments, and other behaviors, and the topic is generated almost at the same time period, the user engagement has a great impact on the final hotness of the topic.

ECANP relatively comprehensively analyzes the influencing factors of topic hotness, and makes full use of the factors such as the number of comments, the number of forwarding, the number of articles, and the release time to evaluate the topic hotness from five indicators. This method can effectively quantify the influence of topics and reflect the impact of user participation, topic coverage, topic activity, topic persistence, and topic novelty on the hotness evaluation results. It is more reasonable and practical for guiding topic ranking.

5. Conclusion

In this paper, we analyzed the relative influence and time attenuation characteristics of topics and hot topics, as well as each indicator affecting the topic hotness according to the reality that the influence of a topic is relative, not absolute. The topic hotness is measured from five indicators: user

engagement, topic coverage, topic activity, topic persistence, and topic novelty, which involves the number of comments, number of microblog articles, and time attenuation characteristics of blogs in the topic, and establishes a topic hotness evaluation model that could quickly aggregate hot topics and evaluate the influence of hot topic. In the experimental stage, we propose to verify the topic hotness evaluation performance from three aspects. Through effectiveness analysis, baseline method performance comparison, and indicator impact analysis, we realize the effective verification and analysis of the topic influence evaluation method proposed in this paper. The results show that ECANP model can effectively evaluate the influence of topics in a period of time, and give a reasonable topic ranking according to its hotness value. Our model involves the calculation of five indicators, but experiments show that the model is low complexity, time-consuming, and easy to understand. The computational complexity will not increase exponentially with the increase of the number of blogs, but linearly with the increase of topics.

Nevertheless, the results of this study have to be seen in light of some limitations. The first is that the research results of this paper calculate the relative influence of topics. Since this paper proposes that it is more meaningful to study whether one topic is more hot than another topic for topic hotness ranking, we are studying the relative influence of topics, which requires the participation of multiple topics, considering the impact of different topics, and finally giving

the topic popularity ranking, rather than just calculating the absolute influence of a specified topic. The second limitation relates to the evaluation method of topic influence specifically designed for single domain and single platform in this paper, which has not yet involved cross-domain social platforms.

In view of the above limitations, we will solve them in future work. On the one hand, we will continue to explore the nature of topic and the characteristics of network public opinion. Through the analysis of the multidimensional characteristic attributes of topic, we will find more valuable information. Combined with the high-quality topic propagation influence structure, we will design a more effective topic hotness evaluation model, test the evaluation performance of different topic characteristic models, and realize the evaluation method using the law of topic propagation influence. On the other hand, because of its heterogeneity, multisource and high capacity, cross-domain social platforms have aroused widespread interest and posed many challenges, we will further study the use of multiplatform information features to beyond the evaluation of single platform topic influence.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] L. Chen, J. Du, K. Shimohara, and M. Song, "The hot topics evaluation and key issues in online news about emergency events," in *Proceedings of the 10th World Congress on Intelligent Control and Automation*, pp. 407–412, Beijing, China, January 2012.

[2] L. Chen and M. Song, "Study on hot topics identification and key issues in on-line news about emergency events," in *Proceedings of the 2011 International Conference on Advanced Intelligence and Awareness Internet*, pp. 250–254, IET, Shenzhen, October 2011.

[3] K.-L. Nguyen, B.-J. Shin, and S. J. Yoo, "Hot topic detection and technology trend tracking for patents utilizing term frequency and proportional document frequency and semantic information," in *Proceedings of the 2016 International Conference on Big Data and Smart Computing*, pp. 223–230, IEEE, Hong Kong, China, January 2016.

[4] K. K. Bun and M. Ishizuka, "Topic extraction from news archive using TF ∗ PDF algorithm," in *Proceedings of the Third International Conference on Web Information Systems Engineering, 2002. WISE 2002*, pp. 73–82, IEEE, Singapore, December 2002.

[5] K.-L. Nguyen, "Hot topic detection and technology trend tracking for patents utilizing term frequency and proportional document frequency and semantic information," in *Proceedings of the 2016 International Conference on Big Data and Smart Computing (BigComp)*, pp. 223–230, Hong Kong, China, January 2016.

[6] X. Wang, "Hot topic detection in news blog," in *Proceedings of the International Conference on Advances in Materials Science and Information Technologies in Industry*, pp. 1114–1118, Xian, China, 2014.

[7] J. Li and X. Ma, "Research on hot news discovery model based on user interest and topic discovery," *Cluster Computing-The Journal of Neteorks Software Tools and Applications*, vol. 22, no. 4, pp. 8483–8491, 2019.

[8] Y. Hong, Y. Zhang, T. Liu, and S. Li, "Topic detection and tracking review," *Journal of Chinese Information Processing*, vol. 21, no. 6, pp. 71–87, 2007.

[9] C. Wang, M. Zhang, S. Ma, and L. Ru, "Automatic online news issue construction in Web environment," in *Proceedings of the 17th international conference on World Wide Web*, pp. 457–466, New York, United States, April 2008.

[10] Y. Jin, S. H. Myaeng, and Y. Jung, "Use of place information for improved event tracking," *Information Processing & Management*, vol. 43, no. 2, pp. 365–378, 2007.

[11] D. A. Smith, "Detecting and browsing events in unstructured text," in *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 73–80, ACM, New York, USA, August 2002.

[12] A. Wang and J. Zhang, "Topic discovery method based on topic model combined with hierarchical clustering," in *Proceedings of the 2020 IEEE 5th Information Technology and Mechatronics Engineering Conference*, pp. 814–818, IEEE, Chongqing, China, June 2020.

[13] N. Li and D. D. Wu, "Using text mining and sentiment analysis for online forums hotspot detection and forecast," *Decision Support Systems*, vol. 48, no. 2, pp. 354–368, 2010.

[14] J. Liu, Q. Wang, Y. Liu, and Y. Li, "A short text topic discovery method for social network," in *Proceedings of the 33rd Chinese Control Conference*, pp. 512–516, IEEE, Nanjing, China, July 2014.

[15] G. Xu, Z. Yu, C. Wang, and A. Wang, "Research on topic discovery technology for Web news," *Neural Computing & Applications*, vol. 32, no. 1, pp. 73–83, 2020.

[16] J. Allan, V. Lavrenko, and R. Papka, "On-line new event detection and tracking," in *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 37–45, ACM, New York, United States, 2017.

[17] J. Ma, F. Wu, and C. Li, "The topic tracking based on semantic similarity of sememe's lexical chain," in *Proceedings of the 2nd International Conference on Software Engineering, Knowledge Engineering and Information Engineering*, pp. 118–121, Singapore, March 2014.

[18] K.-Y. Chen, L. Luesukprasert, and S.-C. T. Chou, "Hot topic extraction based on timeline analysis and multidimensional sentence modeling," *IEEE Transactions on Knowledge and Data Engineering*, vol. 19, no. 8, pp. 1016–1025, 2007.

[19] C. Deng, H. Deng, and Y. Liu, "Online hot topic discovery and hotness evaluation," in *Proceedings of the 3rd International Conference on Computer Science and Application Engineering*, pp. 1–8, ACM, Sanya, China, October 2019.

[20] J. Li, X. Zhang, Y. Weng, and C. Hu, "Blog hotness evaluation model based on text opinion analysis," in *Proceedings of the 2009 Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing*, pp. 235–240, IEEE, Chengdu, China, December 2009.

[21] M. Zhong, "Hot topic discovery in online community using topic labels and hot features," *Tehnicki Vjesnik-Technical Gazette*, vol. 26, no. 4, pp. 1068–1075, 2019.

[22] C. Wang, M. Zhang, L. Ru, and S. Ma, "Automatic online news topic ranking using media focus and user attention based on aging theory," in *Proceedings of the 17th ACM conference on Information and knowledge management*, pp. 1033–1042, ACM, New York, USA, October 2008.

[23] H. Chen, H. Yin, X. Li, M. Wang, W. Chen, and T. Chen, "People opinion topic model: opinion based user clustering in social networks," in *Proceedings of the the 26th International Conference on World Wide Web*, pp. 1353–1359, CC, Perth, Australia, April 2017.

[24] C. Liu and R. Hu, "Hot topic discovery across social networks based on improved LDA model," *KSII Transactions on Internet and Information Systems*, vol. 15, no. 11, pp. 3935–3949, 2021.

[25] Y. Luo, *Internet Hot Topic Discovery Model Research Based on User Browse Behavior*, pp. 1–49, Beijing University of Posts and Telecommunications, Beijing, China, 2008.

Research Article

Nonlinear Dynamic Analysis of Bistable Piezoelectric Energy Harvester with a New-Type Dynamic Amplifier

Dawei Man [1,2] **Gaozheng Xu,**[1] **Huaiming Xu,**[1] **Deheng Xu,**[1] **and Liping Tang**[1,2]

[1]*School of Civil Engineering, Anhui Jianzhu University, Hefei 230601, China*
[2]*BIM Engineering Center of Anhui Province, Hefei 230601, China*

Correspondence should be addressed to Dawei Man; mandawei@ahjzu.edu.cn

A distributed parametric mathematical model of a new-type dynamic magnifier for a bistable cantilever piezoelectric energy harvester is proposed by using the generalized Hamilton principle. The new-type dynamic magnifier consists of a two-spring-mass system, one is placed between the fixed end of the piezoelectric beam and the L-shaped frame, and the other is placed between the L-shaped frame and the base structure. We used the harmonic balance method to obtain the analytical expressions for the steady-state displacement, steady-state output voltage, and power amplitude of the system. The effect of the distance between the magnets, the spring stiffness ratio and mass ratio of the two dynamic magnifiers, and the load resistance on the performance of the harvester is investigated. Analytical results show that compared with the bistable piezoelectric energy harvester with a typical spring-mass dynamic magnifier, the proposed new-type energy harvester system with a two-spring-mass dynamic magnifier can provide higher output power over a broader frequency band, and increasing the mass ratio of the magnifier tip mass to the tip magnet can significantly increase the output power of the BPEH + TDM system. Properly choosing the stiffness ratio of the two dynamic amplifiers can obviously improve the harvested power of the piezoelectric energy harvester at a low excitation level.

1. Introduction

In recent years, the rapid development of wireless sensor networks in building structure health and environmental monitoring has put forward higher requirements for the sustainability of its power supply. Piezoelectric energy harvesting technology is one of the most commonly used energy harvesting technologies, which collects vibration energy from the surrounding environment and converts it into useable energy [1, 2]. In the early stages, different types of linear resonant piezoelectric energy harvesters were designed to generate electrical energy from ambient vibrations. The electro-mechanical coupling equation of a linear cantilever piezoelectric energy harvester was derived and experimentally validated by Erturk and Inman [3, 4]. The ambient vibration excitation frequency usually has the characteristics of time-varying and broadband, so if the ambient vibration frequency does not match the harvester's resonant frequency, the efficiency of the linear piezoelectric

energy harvester is not high [5–9]. This makes it difficult to meet the requirements of the practical application for this linear piezoelectric energy harvester [10].

The nonlinear techniques enable piezoelectric energy harvesters to achieve energy harvesting in a wider frequency band. Due to the increase of the working frequency bandwidth, the nonlinear piezoelectric energy harvester is less sensitive to the change of the external excitation frequency than the linear piezoelectric energy harvester and is more suitable for harvesting energy from the ambient vibration in practical applications [11–14]. The nonlinearity of piezoelectric energy harvesters induced by magnetic forces is usually classified into three main categories, namely, monostable [15, 16], bistable [17, 18], and tristable [19, 20]. Bistable piezoelectric energy harvesters (2 stable and 1 unstable equilibrium positions) have been extensively investigated and their broadband advantages over linear energy harvesters have been verified in simulations and experiments [21, 22]. Stanton et al. [23] established an

analytical model of a bistable piezoelectric energy harvester consisting of a permanent magnet and a piezoelectric cantilever beam and investigated the dynamic characteristics of the system using numerical simulations and experimental methods. Stanton et al. [24] studied the voltage output of a bistable cantilever piezoelectric energy harvester system under different excitation intensity and analyzed the influence of magnet spacing on the system response. He and Daqaq [25] investigated the influence mechanism of asymmetric potential well characteristics on bistable piezoelectric energy harvester under white noise excitation. Kim et al. [26] proposed an electro-mechanical coupling equation for a hysteresis reversible magneto-elastic piezoelectric energy harvester, and the analytical solutions of the system response are obtained by the multiscale method and the high-dimensional harmonic balance method, respectively. The operating bandwidth and output power of the bistable piezoelectric energy harvester have been substantially increased after entering the interwell motion. However, it requires higher excitation strength. If the excitation strength is low, the bistable energy harvester may exhibit intrawell motion which greatly reduces the output performance of the system [27].

To improve the output performance of the bistable energy harvester under low-level excitation, researchers try to make it easier to oscillate with large amplitude interwell motion. Sebald et al. [28, 29] found that external intervention and increasing the excitation amplitude can help the bistable energy harvester jump from intrawell motion to large amplitude interwell motion. However, the excitation level of the vibration in the real environment is low, and it is difficult to enter the large-scale interwell movement [30]. Ma et al. [31] proposed an asymmetric tristable energy harvester, which has a shallower and wider potential well, so that it can extract vibration energy in a wider frequency range, even at a relatively low excitation level, but the interwell output power amplitude is low in this case. Wang et al. [32] propose a configuration that includes an elastic amplifier to amplify the base excitation and provide enough kinetic energy to overcome the tristable potential well barriers, thus leading to large amplitude bistable intermotion. They only consider to amplify the vibration displacement of the base but do not consider how to further amplify the vibration amplitude of the cantilever beam. In order to further improve the performance of the energy capture device under weak excitation, a new-type bistable piezoelectric cantilever energy harvester (BPEH) with two dynamic magnifiers (TDMs) is proposed in this paper. It can amplify the amplitude of the low-level base excitation and the vibration amplitude of the fixed end of the piezoelectric cantilever beam at the same time, so as to dramatically improve the output power and effective bandwidth of the piezoelectric energy harvester. Considering the size effect of the tip magnet, the distributed parameter electro-mechanical coupling equation of the bistable piezoelectric energy harvester with two dynamic magnifiers (BPEH + TDM) is established based on the generalized Hamilton principle, and the analytical solution of the energy capture system is derived by using the harmonic balance method. The effects of the distance between

the magnets, the mass of the dynamic magnifiers, the load resistance, and the stiffness ratio of the two dynamic magnifiers on the performance of the energy capture system were studied. The results show that compared with the typical bistable piezoelectric energy harvester with a dynamic magnifier, the piezoelectric proposed energy harvester system with a two-spring-mass dynamic magnifier can collect higher output power over a broader frequency band. By reasonably selecting the design parameters of the amplifier, the harvested power can be significantly increased and the effective bandwidth of the harvester can be improved. The mathematical model of the BPEH + TDM is described in Section 2. The harmonic balance method is used for analytical expressions for the steady-state displacement, steady-state output voltage, and power amplitude of the BPEH + TDM in Section 3. The effects of parameter variations of the BPEH + TDM on its dynamic characteristics are numerically investigated in Section 4.

2. Mathematical Model of the BPEH + TDM

The BPEH + TDM configuration considered in this paper is schematically shown in Figure 1. The BPEH comprises a piezoelectric cantilever beam and two magnets (denoted as A and B). The piezoelectric cantilever beam is composed of a substrate layer, covered with a pair of piezoelectric layers (PZTs) on both of its surfaces, and poled oppositely in the thickness direction. The two piezoelectric layers are electrically connected in series with a load resistance (R), representing the equivalent resistance of a low power electronic device. Magnet A (called the tip magnet) is attached to the tip of the cantilever beam and the external magnet B is fixed at the right wall of the L-shaped frame. The TDM comprises two dynamic magnifiers (denoted as DM1 and DM2), the DM1 is basically a spring (k_f)-mass (M_f) system placed between the fixed end of the piezoelectric beam and the bottom of the L-shaped frame, and the DM2 composed an L-shaped frame and a spring k_b, and the L-shaped frame is mechanically connected in series with the spring k_b. M_f and M_m represent the mass of DM1 and DM2, respectively. The horizontal gap between the tip magnet and magnet B is d. Here, l and b are the length and width of the piezoelectric cantilever beam, respectively; h_s and t_p denote the thickness of the substrate layer and the PZTs, respectively; e is the eccentricity of the tip magnet.

$v_m(t)$ and $v_b(t)$ represent the vibration displacement of the DM2 and the base, respectively. s is the coordinate along the neutral axis of the beam, and $v(s, t)$ represents the displacement of the beam at s position relative to its fixed end. The constitutive equations of the piezoelectric cantilever beam are assumed as follows:

$$\left.\begin{aligned} T_1^s &= Y_s S_1^s \\ T_1^p &= Y_p\left(S_1^p - d_{31}E_3\right) \\ D_3 &= d_{31}T_1 + \varepsilon_{33}^T E_3 \end{aligned}\right\}. \tag{1}$$

Here, Y is Young's modulus, subscript/superscript p and s represent the piezoelectric layers and substrate layer, and S_1
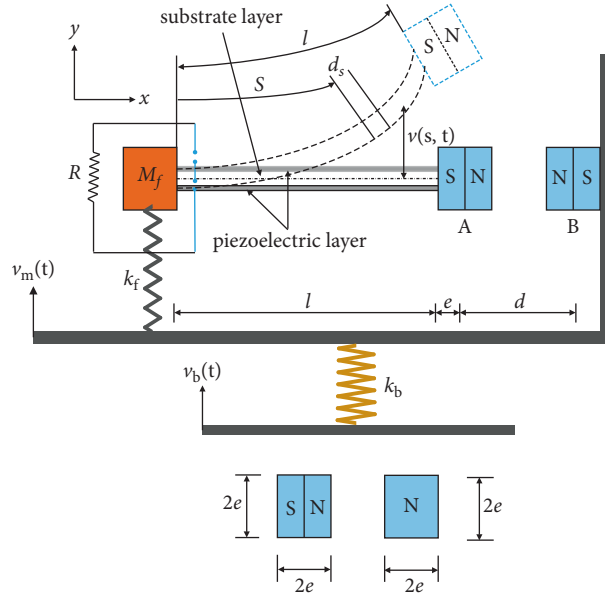
FIGURE 1: Schematic of the considered BPEH + TDM.

and T_1 are the strain and the stress of the beam, respectively. D_3 is the electric displacement and d_{31} and ε_{33}^T are the piezoelectric constant and dielectric constant, respectively. $E_3 = -V(t)/(2t_P)$ is the electric field, in which $V(t)$ represents the voltage. The strain generated in the piezoelectric beam can be assumed as $S_1^s = S_1^P = -yv''$.

The generalized Hamilton's principle of the BPEH + TDM system is as follows:

$$\int_{t_1}^{t_2} \left[\delta \left(T_k + W_e - U_e - U_m - U_d \right) + \delta W \right] \mathrm{d}t = 0. \tag{2}$$

Here, T_k, W_e, U_e, U_m, U_d, and W are the kinetic energy, the electrical energy, the strain energy, the magnetic potential energy, the elastic potential of the dynamic magnifiers, and the external work, respectively. T_k and W_e are as follows:

$$T_k = \frac{1}{2} \int_0^l m \left(\dot{v} + \dot{v}_m(t) \right)^2 \mathrm{d}s + \frac{1}{2} M_t \left(\dot{v}(l,t) + e\dot{v}'(l,t) + \dot{v}_m(t) \right)^2 + \frac{1}{2} J\dot{v}'(l,t)^2$$

$$+ \frac{1}{2} M_f \left(\dot{v}(0,t) + \dot{v}_m(t) \right)^2 + \frac{1}{2} M_m \dot{v}_m(t)^2, \tag{3}$$

$$W_e = \frac{1}{2} Y_p b d_{31} \left(h + \frac{t_p}{2} \right) V(t) \int_0^l v'' \mathrm{d}s + bl\varepsilon_{33}^S \frac{V(t)^2}{4t_p}. \tag{4}$$

Here, $m = 2\rho_p t_p b + \rho_s h_s b$, in which ρ_p and ρ_s are the density of the piezoelectric layers and substrate layer, respectively. M_t is the tip magnet mass and J represents the rotary inertia of the tip magnet, ε_{33}^s is the permittivity.

U_e is expressed as follows:

$$U_e = \frac{1}{2} \int_0^l \left[YIv''^2 - Y_p b d_{31} \left(h + \frac{t_p}{2} \right) V(t)v'' \right] \mathrm{d}s. \tag{5}$$

Here, $h = (h_s/2)$, $YI = (2/3) [Y_s bh^3 + Y_p b (3h^2 t_p + 3ht_p^2 + t_p^3)]$.

U_d is expressed as follows:

$$U_d = \frac{1}{2} k_f v(0,t)^2 + \frac{1}{2} k_b v_m^2. \tag{6}$$

Here, k_f and k_b represent the stiffness of DM1 and DM2, respectively. Considering the eccentricity of the tip magnet, U_m can be given by the following equation:

$$U_{\mathrm{m}} = \mu_0 M_A V_A M_B V_B \left\{ -\left(v(l,t) + \frac{ev'(l,t)}{\sqrt{1 + v'(l,t)^2}} \right)^2 + 2\left[d + e\left(1 - \frac{1}{\sqrt{1 + v'(l,t)^2}} \right) \right]^2 \right.$$

$$-3\left[d + e\left(1 - \frac{1}{\sqrt{1 + v'(l,t)^2}} \right) \right]\left(v(l,t) + \frac{ev'(l,t)}{\sqrt{1 + v'(l,t)^2}} \right)v'(l,t) \right\} \tag{7}$$

$$4\pi\sqrt{1 + v'(l,t)^2} \left\{ \left[d + e\left(1 - \frac{1}{\sqrt{1 + v'(l,t)^2}} \right) \right]^2 + \left(v(l,t) + \frac{ev'(l,t)}{\sqrt{1 + v'(l,t)^2}} \right)^2 \right\}^{5/2}.$$

Here, $\mu_0 = 4\pi \times 10^{-7} H \cdot m^{-1}$ is the magnetic permeability constant. M_A (M_B) and V_A (V_B) are the magnetization intensity and volume of the magnet A (B), respectively.

Using the Galerkin approach, $v(s,t)$ can be written as follows:

$$v(s,t) = \phi_r(s)\eta_r(t). \tag{8}$$

Here, $\varphi_r(s)$ and $\eta_r(t)$ represent the R-order mode shape function and the generalized mode coordinates of the beam, respectively.

The modal shape function satisfies the following orthogonal relations:

$$\int_0^l \phi_s(s)m\phi_r(s)\mathrm{d}s + \phi_s(l)M_t\phi_r(l) + \phi_s(l)M_t e\phi_r'(l) + \phi_s(0)M_f\phi_r(0)$$

$$+ \phi_s'(l)\left(J + M_t e^2\right)\phi_r'(l) + \phi_s'(l)M_t e\phi_r(l) = \delta_{rs}, \tag{9}$$

$$\int_0^l \frac{\mathrm{d}^2\phi_s(s)}{\mathrm{d}s^2} YI \frac{\mathrm{d}^2\phi_r(s)}{\mathrm{d}s^2}\mathrm{d}s + \phi_s(0)k_f\phi_r(0) = \omega_r^2\delta_{rs}. \tag{10}$$

Here, δ_{rs} represents the Kronecker delta. $\omega_r = \lambda_r^2\sqrt{YI/(ml^4)}$ represents the resonance frequency of the r-th mode, in which λ_r is the eigenvalue. The calculation process of the λ_r is described in the literature [33, 34].

Substituting equation (8) into (7), the Taylor's expansion of U_m at $\eta(t) = 0$ can be expressed as follows:

$$U_m = k_0 - \frac{1}{2}k_1\eta_1^2 + \frac{1}{4}k_2\eta_1^4 + o\left(\eta_1^5\right). \tag{11}$$

Here, $k_0 = 2\kappa/d^3$,

$$k_1 = \frac{\kappa\left(10q_1 + 2d^2\phi_1'(l)^2 + 2q_2\right)}{d^5},$$

$$k_2 = \frac{\kappa\left[8d^2q_3 + 35q_1^2 + 10\left(d^2\phi_1'(l)^2 + q_2\right)q_1 + \left(3d^2\phi_1'(l)^4 + 2q_2\phi_1'(l)^2 + 4q_4\right)d^2\right]}{d^7},$$

$$q_1 = de\phi_1'(l)^2 + e^2\phi_1'(l)^2 + 2e\phi_1(l)\phi_1'(l) + \phi_1(l)^2,$$

$$q_2 = \left(e\phi_1'(l) + \phi_1(l)\right)^2 - 2\ de\ \phi_1'(l)^2 + 3\ d\left(e\phi_1'(l) + \phi_1(l)\right)\phi_1'(l),$$

$$q_3 = 2.5\left(0.75\ de\ \phi_1'(l)^4 + 0.75e^2\phi_1'(l)^4 + e\phi_1(l)\phi_1'(l)^3\right),$$

$$q_4 = \left(e\phi_1'(l) + \phi_1(l)\right)e\phi_1'(l)^3 - 1.5\ de\ \phi_1'(l)^4 + 0.5e^2\phi_1'(l)^4$$

$$- 3\left[-0.5\ de\ \phi_1'(l)^3 + 0.5e\phi_1'(l)^2\left(e\phi_1'(l) + \phi_1(l)\right)\right]\phi_1'(l),$$

$$\kappa = \frac{\mu M_A V_A M_B V_B}{4\pi},$$ (12)

$$\delta W = \delta v_m \ddot{v}_b\left(M_m + M_t + ml + M_f\right) + \delta\eta(t)\ddot{v}_b\left(M_t\phi_1(l) + m\int_0^l \phi_1(s)\mathrm{d}s + M_t e\phi_1'(l) + M_f\phi_1(0)\right).$$

The external virtual work can be defined as follows.

Substituting equation (8) into (2) and considering only the 1st order mode, Lagrange's equation for the BPEH + TDM system is given by the following equation:

$$\begin{cases} \dfrac{\mathrm{d}}{\mathrm{d}t}\left(\dfrac{\partial L}{\partial v_m}\right) - \dfrac{\partial L}{\partial v_m} + \dfrac{\partial W}{\partial v_m} = 0, \\[3mm] \dfrac{\mathrm{d}}{\mathrm{d}t}\left(\dfrac{\partial L}{\partial \dot{\eta}}\right) - \dfrac{\partial L}{\partial \eta} + \dfrac{\partial W}{\partial \eta} = F(t), \\[3mm] \dfrac{\mathrm{d}}{\mathrm{d}t}\left(\dfrac{\partial L}{\partial \dot{V}}\right) - \dfrac{\partial L}{\partial V} + \dfrac{\partial W}{\partial V} = Q(t). \end{cases}$$ (13)

Here, $F_1(t) = -2\xi_1\omega_1\dot{\eta}_1(t)$ is the generalized dissipative force, ξ_1 is the damping ratio, and $Q(t) = V(t)/R$ represents the generalized output charge.

The electro-mechanical coupling equations of the BPEH + TDM system can be obtained by using the following equation:

$$\begin{cases} M_0\ddot{\eta}_1(t) + M_1\ddot{v}_m(t) + k_b v_m = -M_1\ddot{v}_b(t), \\[3mm] \ddot{\eta}_1(t) + 2\xi_1\omega_1\dot{\eta}_1(t) + \omega_1^2\eta_1(t) - k_1\eta_1(t) + k_2\eta_1(t)^3 - \theta_1 V(t) + M_0\ddot{v}_m(t) = -M_0\ddot{v}_b(t), \\[3mm] C_p\dot{V}(t) + \dfrac{V(t)}{R} + \theta_1\dot{\eta}_1(t) = 0. \end{cases}$$ (14)

Here, $M_0 = m\int_0^l \varphi_1(s)\mathrm{d}s + M_t\varphi_1(l) + M_t e\varphi_1'(l) + M_f\varphi_1(0)$, $M_1 = ml + M_t + M_f + M_m$, $\omega_1^2 = YI\int_0^l \varphi_1''(s)^2\mathrm{d}s + k_f\varphi_1(0)^2$, $\theta_1 = Y_p bd_{31}(h + (t_p/2))\int_0^l \varphi_1''(s)\mathrm{d}s$, $C_p = bl\varepsilon_{33}^S/2t_p$. Here, $\omega_1^2 = YI\int_0^l \varphi_1''\mathrm{d}s$, $g_0 = mg\int_0^l \varphi_1(s)\mathrm{d}s + M_t g\varphi_1(l)$, $\Gamma_1 = m\int_0^l \varphi_1(s)\mathrm{d}s + M_t(\varphi_1(l) + e\varphi_1'(l))$, $\theta_1 = Y_p bd_{31}(h + (t_p/2))\int_0^l \varphi_1(s)\mathrm{d}s$, and $C_p = bl\varepsilon_{33}^s/2t_p$.

The excitation acceleration is assumed to be $\ddot{v}_b(t) = \ddot{v}_b\cos(\omega_e t)$, where $\ddot{v}_b$ denotes the excitation amplitude, ω_e denotes the circular frequency, and C_p denotes the capacitance. Introducing the dimensionless parameters $x = \eta_1/l$, $V_m = v_m/l$, $V_b = v_b/l$, $\ddot{V} = (VC_p/l\theta_1)$, $\tau = \omega_1 t$, equation (14) can be rewritten as the following equation in the dimensionless form:

$$\begin{cases} \dfrac{M_1 - M_0^2}{K_b}x^{(4)} + \dfrac{2M_1\xi_1}{K_b}x^{(3)} + \dfrac{M_1(1 - K_1) + K_b}{K_b}\ddot{x} + 2\xi_1\dot{x} + (1 - K_1)x + K_2 x^3 \\[3mm] + \dfrac{M_1 K_2}{K_b}\left(6x\dot{x}^2 + 3x^2\ddot{x}\right) - \dfrac{M_1\Theta}{K_b}\ddot{V} - \Theta\ddot{V} = F\cos(\omega\tau), \\[3mm] \ddot{V} + \alpha\ddot{V} + \dot{x} = 0. \end{cases}$$ (15)

Here, $K_b = k_b/\omega_1^2$, $K_1 = k_1/\omega_1^2$, $K_2 = k_2 l^2/\omega_1^2$, $\Theta = \theta_1^2/C_p\omega_1^2$, $\alpha = 1/C_pR_L\omega_1$, $F = -M_0\dot{v}_b/\omega_1^2 l$.

3. Harmonic Balance Analysis

The solution of equation (15) is assumed to be

$$
\begin{cases}
x = A(\tau) + B(\tau)\sin(\omega\tau) + C(\tau)\cos(\omega\tau), \\
\overline{V} = D(\tau)\sin(\omega\tau) + E(\tau)\cos(\omega\tau).
\end{cases}
\tag{16}
$$

Here, A, B, C, D, and E are undetermined coefficients, so the displacement amplitude can be expressed as $a = \sqrt{B^2 + C^2}$ and the output voltage amplitude can be expressed as $u = \sqrt{D^2 + E^2}$.

Substituting equation (16) into (15), let the constant terms on both sides of the equation and the coefficients of $\sin(\omega\tau)$ and $\cos(\omega\tau)$ consistent and ignoring the high-order harmonic term and partial zero term, we can obtain the following equations:

$$
Z_1\ddot{A} + 2\xi_1\dot{A} + (1 - K_1)A + K_2A^3 + \frac{3}{2}K_2A\left(B^2 + C^2\right) = 0,
\tag{17}
$$

$$
Z_1\left(\ddot{B} - 2\omega\dot{C}\right) + 2\xi_1\dot{B} + Z_2C + Z_3B + Z_4D = 0,
\tag{18}
$$

$$
Z_1\left(\ddot{C} - 2\omega\dot{B}\right) + 2\xi_1\dot{C} + Z_3C - Z_2B + Z_4E - F = 0,
\tag{19}
$$

$$
\dot{D} - \omega E + \alpha\, D + \dot{B} - \omega C = 0,
\tag{20}
$$

$$
\dot{E} + \omega\, D + \alpha E + \dot{C} + \omega B = 0.
\tag{21}
$$

Here,

$$
\begin{aligned}
Z_1 &= \frac{K_1M_1 + K_b}{K_b}, \\[4pt]
Z_2 &= \frac{2\xi_1M_1}{K_b}\omega^3 - 2\xi_1\omega, \\[4pt]
Z_3 &= \frac{M_1 - M_0^2}{K_b}\omega^4 - \frac{(1 - K_1)M_1 + K_b}{K_b}\omega^2 + 1 - K_1 + K_2\left(3A^2 + \frac{3}{4}a^2\right) - \frac{K_2M_1}{K_b}\omega^2\left(3A^2 + \frac{3}{4}a^2\right), \\[4pt]
Z_4 &= \frac{\Theta M_1}{K_b}\omega^2 - \Theta.
\end{aligned}
\tag{22}
$$

As the undetermined coefficients A, B, C, D, and E in equations (17)–(21) change slowly, it can be considered that

$$
\begin{cases}
\dfrac{dA}{d\tau} = \dfrac{dB}{d\tau} = \dfrac{dC}{d\tau} = \dfrac{d\,D}{d\tau} = \dfrac{dE}{d\tau} = 0, \\[8pt]
\dfrac{d^2A}{d\tau^2} = \dfrac{d^2B}{d\tau^2} = \dfrac{d^2C}{d\tau^2} = \dfrac{d^2D}{d\tau^2} = \dfrac{d^2E}{d\tau^2} = 0, \\[8pt]
\dfrac{d^3A}{d\tau^3} = \dfrac{d^3B}{d\tau^3} = \dfrac{d^3C}{d\tau^3} = 0, \\[8pt]
\dfrac{d^4A}{d\tau^4} = \dfrac{d^4B}{d\tau^4} = \dfrac{d^4C}{d\tau^4} = 0.
\end{cases}
\tag{23}
$$

Using equations (20) and (21), we obtain the following equations:

$$
D = \frac{\omega}{\omega^2 + \alpha^2}(\alpha C - \omega B),
\tag{24}
$$

$$
E = -\frac{\omega}{\omega^2 + \alpha^2}(\omega C + \alpha B).
\tag{25}
$$

Then, substituting formulas equations (20) and (21) into equations (18) and (19), respectively, we obtain the following equations:

$$
B = -\frac{F\left(Z_2 + Z_4\left(\alpha\omega/\omega^2 + \alpha^2\right)\right)}{\left(Z_3 - \left(Z_4\omega^2/\left(\omega^2 + \alpha^2\right)\right)\right)^2 + \left(Z_2 + \left(Z_4\alpha\omega/\left(\omega^2 + \alpha^2\right)\right)\right)^2},
\tag{26}
$$

$$
C = -\frac{F\left(\left(Z_4\omega^2/\left(\omega^2 + \alpha^2\right)\right) - Z_3\right)}{\left(Z_3 - \left(Z_4\omega^2/\left(\omega^2 + \alpha^2\right)\right)\right)^2 + \left(Z_2 + \left(Z_4\alpha\omega/\left(\omega^2 + \alpha^2\right)\right)\right)^2}.
\tag{27}
$$

Therefore, the displacement amplitude and the voltage amplitude can be expressed as follows:

$$
a^2\left[\left(Z_2 + Z_4\frac{\alpha\omega}{\omega^2 + \alpha^2}\right)^2 + \left(Z_3 - Z_4\frac{\omega^2}{\omega^2 + \alpha^2}\right)^2\right] = F^2.
\tag{28}
$$

Here, the steady-state displacement response amplitude a can be obtained by equation (23), and the steady-state output voltage amplitude and output power amplitude can then be expressed in the following forms:
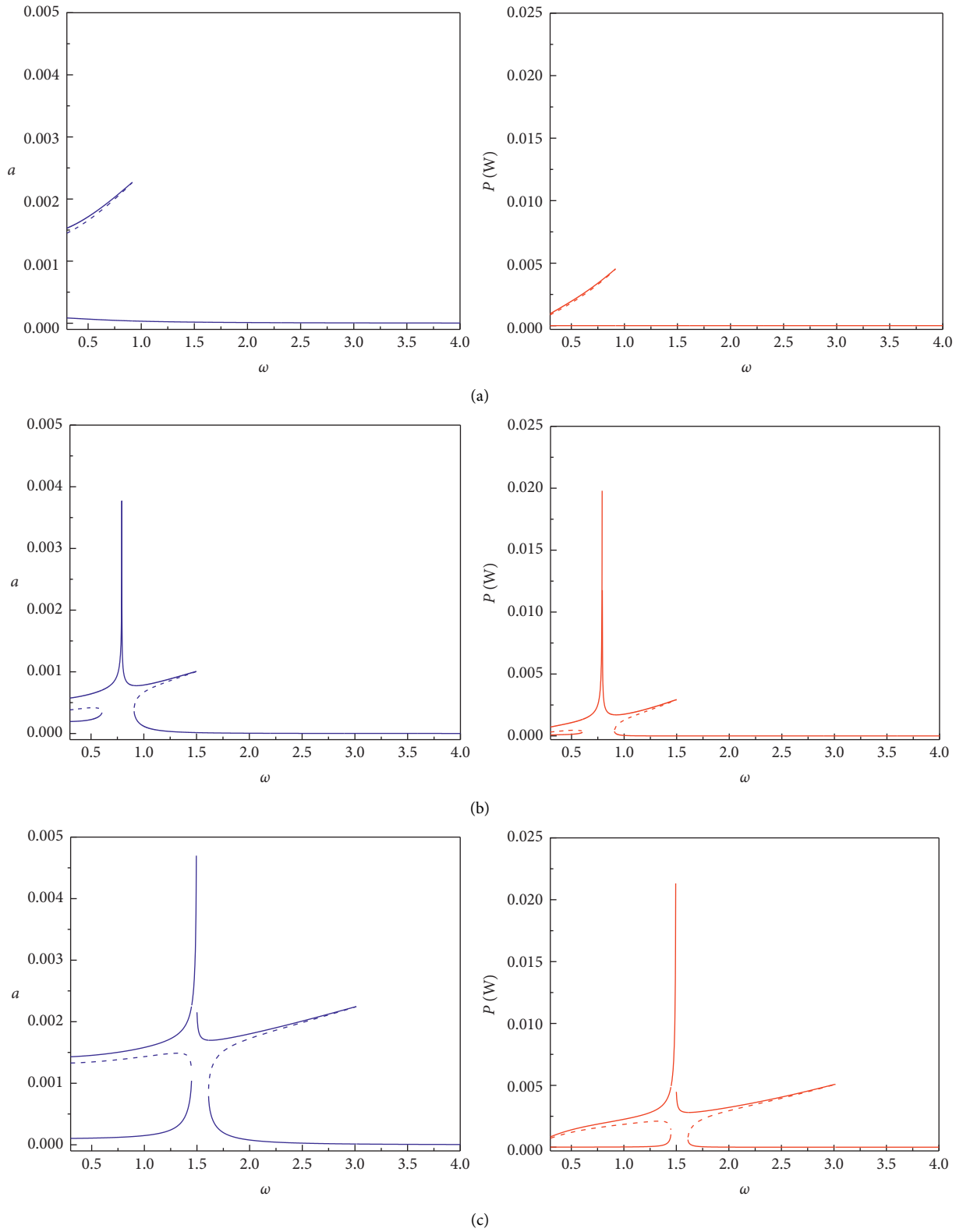
FIGURE 2: Displacement amplitude (left column) and output power amplitude (right column) versus excitation frequency for: (a) BPEH + DM1. (b) BPEH + DM2. (c) BPEH + TDM when $d = 16$ mm.
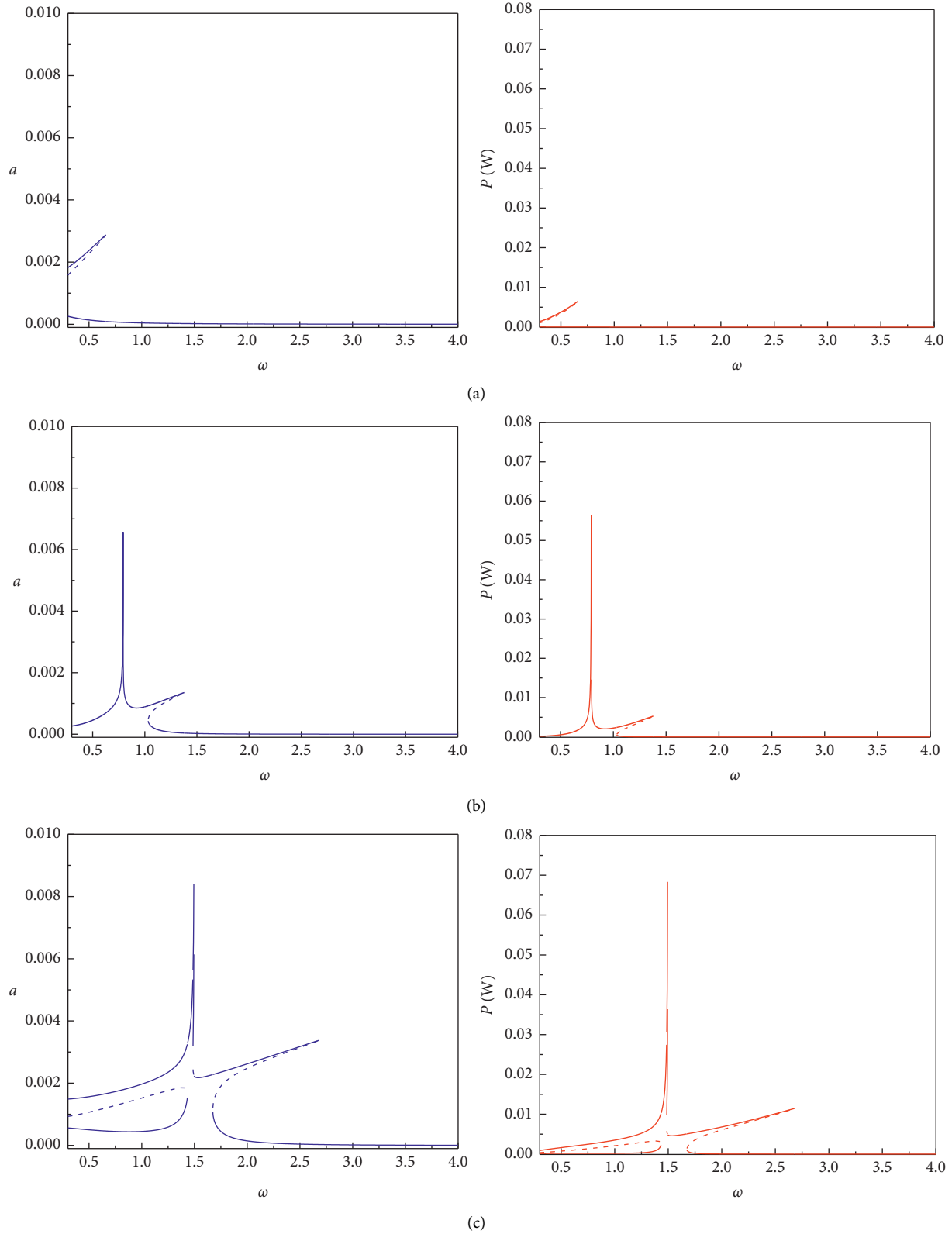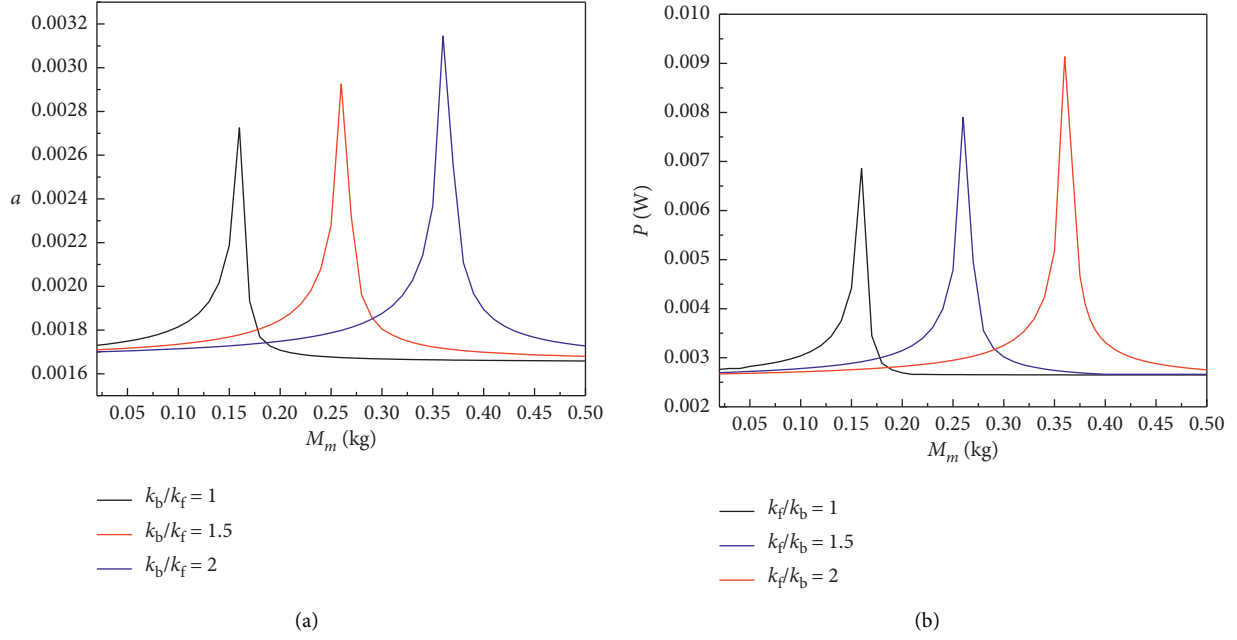
(a)

(b)

(c)

Figure 3: Displacement amplitude (left column) and output power amplitude (right column) versus excitation frequency for: (a) BPEH + DM1. (b) BPEH + DM2. (c) BPEH + TDM when $d = 20$ mm.

FIGURE 4: (a) Displacement amplitude and (b) output power amplitude versus the mass of the base amplifier M_m for excited frequency $\omega = 1.4$ with different value of the stiffness ratio k_f/k_b.
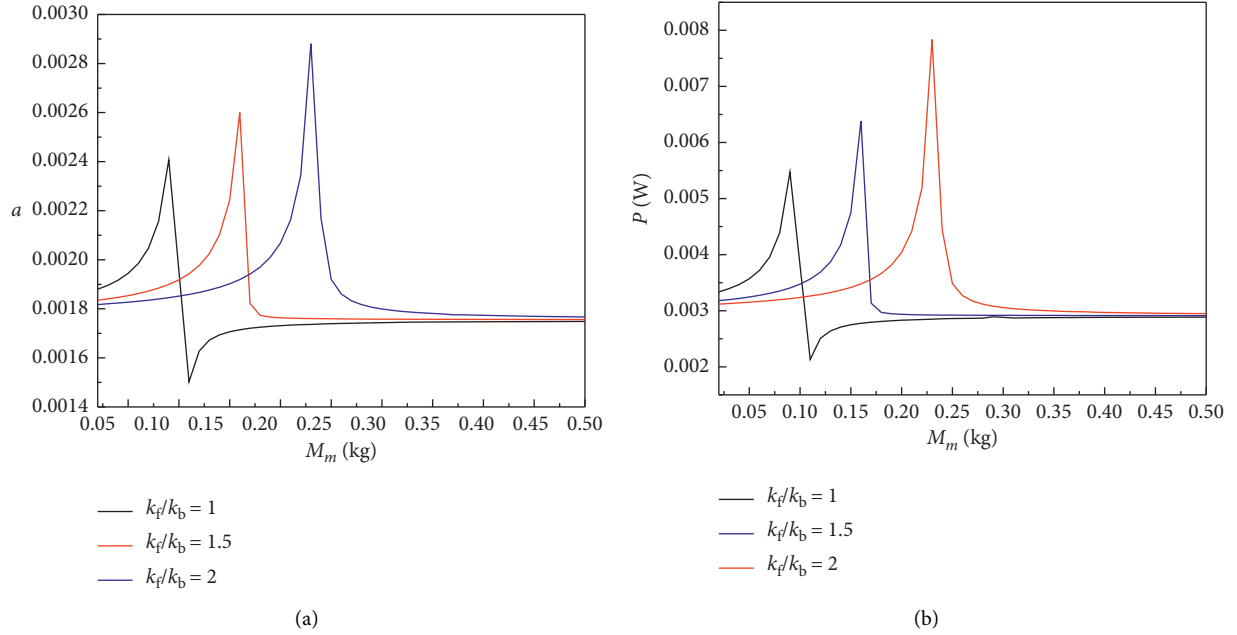

FIGURE 5: (a) Displacement amplitude and (b) output power amplitude versus the mass of the base amplifier M_m for excited frequency $\omega = 1.7$ with different value of the stiffness ratio k_f/k_b.

$$u = \left(\frac{\omega}{\sqrt{\omega_2 + \alpha^2}} \right) a, \qquad (29)$$

$$P = \frac{l^2 \theta_1^2 u^2}{C_p^2 R}. \qquad (30)$$

4. Results and Discussion

In this section, we numerically investigate the effects of the magnet spacing, the mass of the base dynamic magnifier M_m, the load resistance, the stiffness ratio of the k_f to k_b, and the mass ratio of the M_f to M_t on the dynamic characteristics of the BPEH + TDM system. The geometric and material
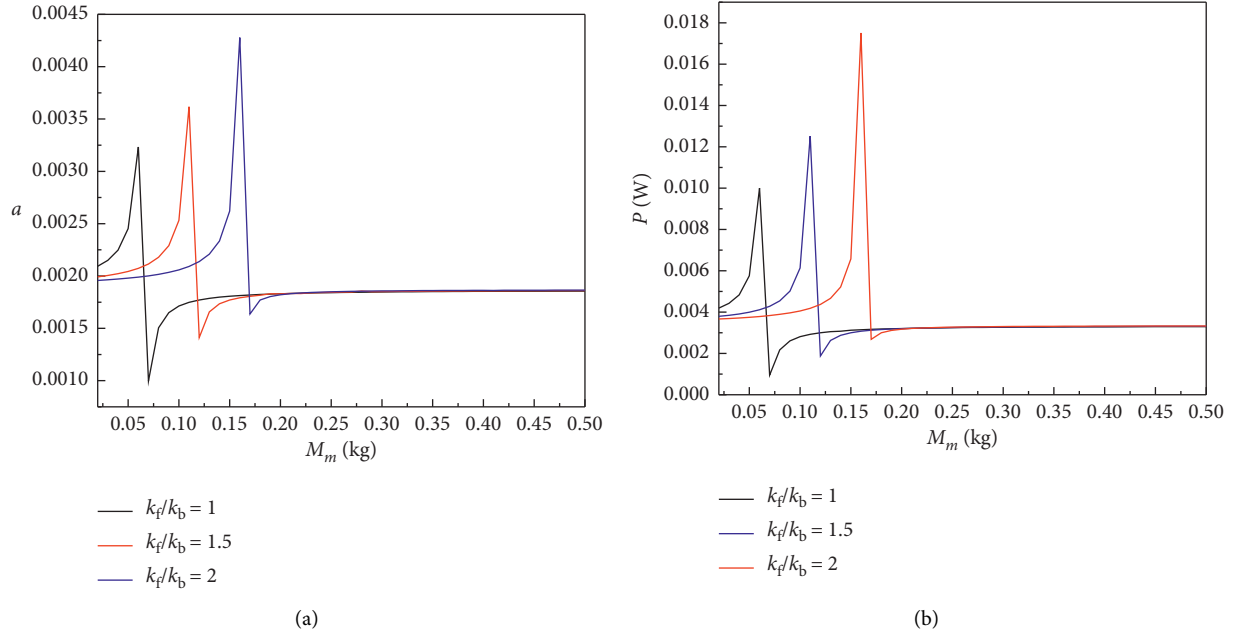
FIGURE 6: (a) Displacement amplitude and (b) output power amplitude versus the mass of the base amplifier M_m for excited frequency $\omega = 2$ with different value of the stiffness ratio k_f/k_b.
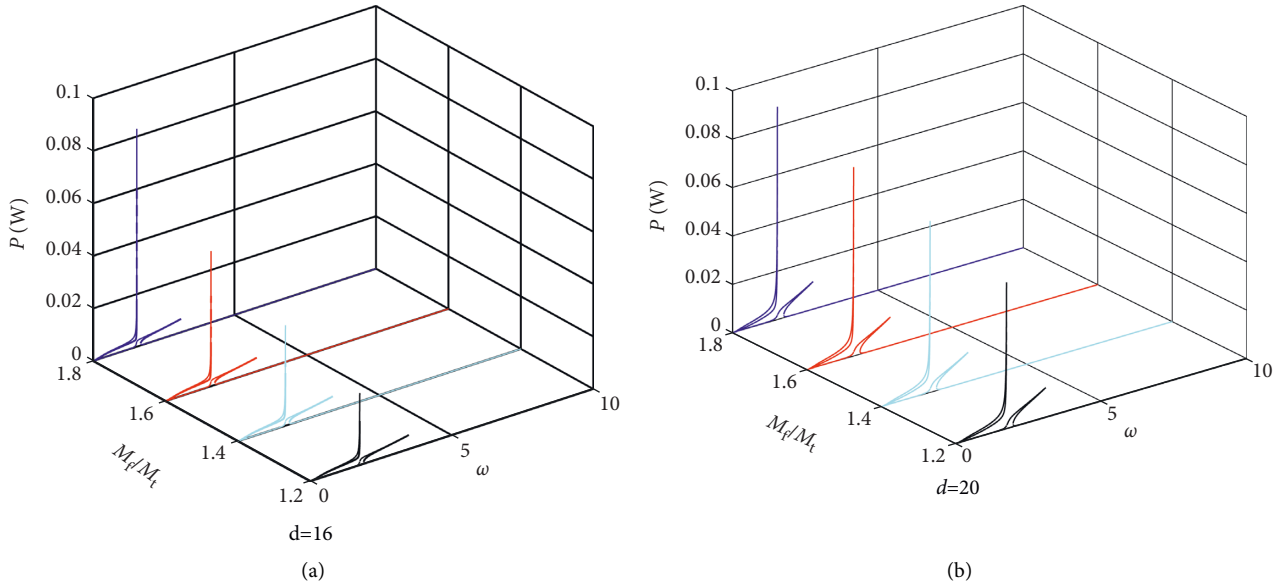
FIGURE 7: Power frequency response curve in different values of the mass ratio M_f/M_t for $M_m = 0.12$ kg when (a) $d = 16$ mm, (b) $d = 20$ mm.

properties are as follows [35]: $l = 75mm$, $b = 20mm$, $h_s = 0.2mm$, $Y_s = 70Gpa$, $\rho_s = 2700kg/m^3$, $M_t = 10 \times 10^{-3}kg$, $M_m = 0.18kg$, $M_f = 16.5 \times 10^{-3}kg$, $k_f = 10.2KN \cdot m$, $k_b = 15.8KN \cdot m$ $M_A = M_B = 1.22 \times 10^6 A/m$, $V_A = V_B = 1 \times 10^{-6}m^3$, $\xi_1 = 0.01$, $Y_p = 60.98Gpa$, $\rho_s = 7750kg/m^3$, $d_{31} = -1.71 \times 10^{-10}C/N$, $\varepsilon_{33}^s = 1.33 \times 10^{-8}F/m$.

In Figures 2 and 3, we define three bistable piezoelectric energy harvester (BPEH) calculation models, namely, BPEH + DM1 (BPEH with a dynamic amplifier placed between the fixed end of the piezoelectric beam and the base

structure), BPEH + DM2 (BPEH with a dynamic amplifier placed between the BPEH and the base structure), and BPEH + TDM (BPEH with DM1 and DM2 amplifiers). Figure 2 depicts variations of displacement and output power versus excited frequency for different calculation models when $d = 16$ mm, $M_t = 10$ g, $M_f = 16.5$ g, and $R = 300$ kΩ. It shows that among the three calculation models, the peak displacement and peak power of the interwell motion of BPEH + TDM are the highest, and its frequency bandwidth is also the widest. When magnet spacing d increases to 20 mm, it can be seen from Figure 3
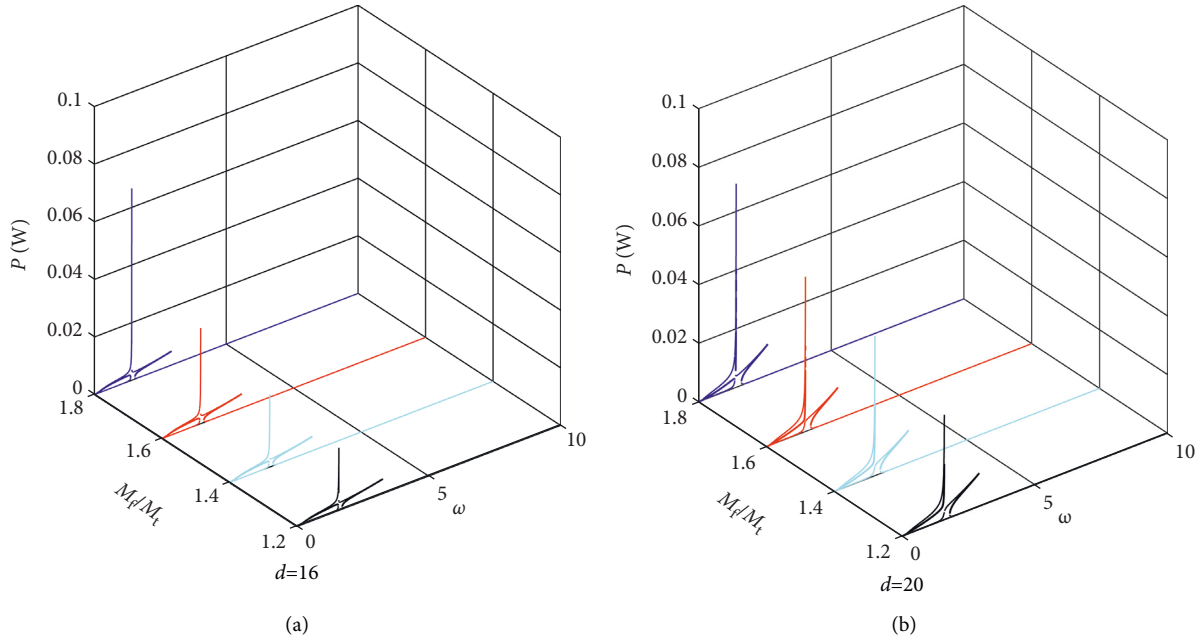
FIGURE 8: Power frequency response curve in different values of the mass ratio M_f/M_t for $M_m = 0.15$ kg when (a) $d = 16$ mm, (b) $d = 20$ mm.
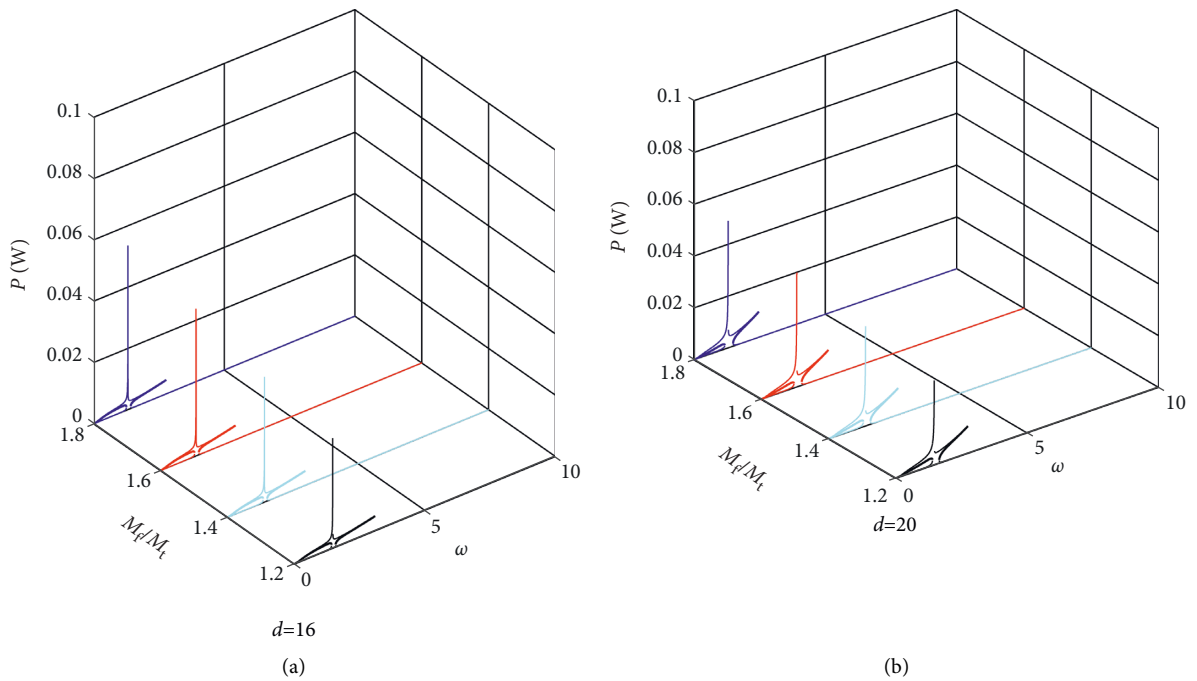

FIGURE 9: Power frequency response curve in different values of the mass ratio M_f/M_t for $M_m = 0.18$ kg when (a) $d = 16$ mm, (b) $d = 20$ mm.

that the peak displacement and peak output power of the three calculation models increase significantly, however, the interwell frequency bandwidth decreases.

Figures 4–6 show the steady-state amplitude response curves of the BPEH + TDM interwell motion displacement and output power with the variation of the base amplifier mass M_m for different stiffness ratios of k_f to k_b when excited frequency $\omega = 1.4$, $\omega = 1.7$, and $\omega = 2$. As can be seen from Figure 4, when excited frequency $\omega = 1.4$, the

displacement amplitude and output power amplitude of the BPEH + TDM first increase to extreme values as the mass of the base amplifier M_m gradually increases, then rapidly decreases, and finally, tend to be stable in a small range, and there exists an optimal mass of the base amplifier mass M_m value which maximizes the displacement amplitude and output power amplitude of the system, and the optimal M_m value increases with the stiffness ratio of k_f to k_b increasing.

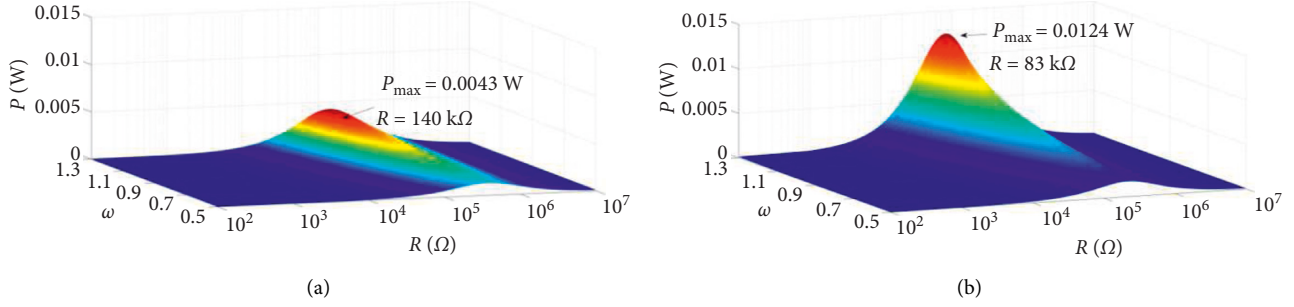(a)   (b)

FIGURE 10: Output power amplitude response of the system with different load resistance: (a) $d = 18\,\text{mm}$ and (b) $d = 20\,\text{mm}$.

Figures 5 and 6 show that when the excitation frequency increases, with the gradual increase of the M_m, the displacement amplitude and output power amplitude of the BPEH + TDM will first increase to the extreme value, then decreases sharply followed by a slight increase, and finally tend to be stable due to falling into the intrawell. It can also be seen from Figure 4 that with the increase of stiffness ratio k_f/k_b, the optimal value of M_m increases, and when M_m reaches the optimal value, the corresponding displacement amplitude and output power amplitude of the BPEH + TDM also increase with the stiffness ratio k_f/k_b increasing.

Figures 7–9 give the power amplitude variation curve with excited frequency for different values of the mass ratio M_f/M_t when $M_\text{m} = 0.12\,\text{kg}$, $M_\text{m} = 0.15\,\text{kg}$ and $M_\text{m} = 0.18\,\text{kg}$. Figures 7–9 show that when the base amplifier M_m and magnet spacing d are kept constant, the peak output power of the BPEH + TDM increases significantly as the mass ratio of the M_f to M_t increases and the excitation frequency at which the system generates peak power decreases. It can also be found from the results of Figures 7–9 that when magnet spacing d and mass ratio M_f/M_t remain unchanged, the peak output power of the BPEH + TDM decreases with the increase of M_m. However, the reduction rate of the peak power slows down as M_m becomes larger. When $d = 20\,\text{mm}$, $M_\text{f}/M_\text{t} = 1.8$, taking $M_\text{m} = 0.15\,\text{kg}$ as examples, the peak power of the BPEH + TDM is 0.069 W, which is 21.6% lower than that of $M_\text{m} = 0.12\,\text{kg}$. However, when M_m increases to 0.18 kg, the corresponding peak power of the BPEH + TDM is decreased by 14.3%, compared with that of $M_\text{m} = 0.15\,\text{kg}$.

Figure 10 shows the variation of output power amplitude with load resistance for magnet spacing $d = 18\,\text{mm}$ and $d = 20\,\text{mm}$. The results show the power amplitude tends to increase at the beginning and decrease afterwards with the increase of load resistance at each excited frequency. Each excitation frequency corresponds to an optimal load resistance to maximize the amplitude of power of the BPEH + TDM, and the optimal load resistance decreases with the increase of excitation frequency. The optimal resistance decreases with the increase of magnet spacing, but the corresponding peak power is significantly higher when the magnet spacing increases.

5. Conclusions

In this paper, based on the generalized Hamilton variational principle, considering the size effect and the

rotary inertia of the tip magnet, an electro-mechanical coupling equation of the BPEH + TDM system is obtained, and the analytical solution of the equation is obtained by using the harmonic balance method. The effects of magnet spacing, the mass of the base dynamic magnifier M_m, the load resistance, the stiffness ratio of the k_f to k_b, and the mass ratio of the M_f to M_t on the BPEH + TDM system are investigated and the following conclusions were obtained:

(1) Increasing the magnet spacing can improve the interwell output power amplitude of the BPEH + TDM system, but the interwell frequency bandwidth decreases.

(2) There exists an optimal mass of the base dynamic magnifier to maximize the output power of the BPEH + TDM system, and the optimal value of the base dynamic magnifier mass increases with the increase of stiffness ratio k_f/k_b.

(3) The peak output power of the BPEH + TDM system increases significantly as the mass ratio of the M_f to M_t increases, and the excitation frequency at which the system generates peak power decreases with increasing M_f/M_t. The peak output power of the BPEH + TDM decreases with the increase of M_m. However, the reduction rate of the peak power slows down when M_m is large.

(4) Compared with the BPEH + DM1 system which a dynamic amplifier is placed between the fixed end of the piezoelectric beam and the BPEH + DM2 system which a dynamic amplifier is placed between the BPEH and the base structure, the BPEH + TDM system can produce higher peak output power and wider interwell bandwidth.

In many cases, the excitation of piezoelectric energy capture devices is mostly random. In the future, to further explore the strategy of inducing the multistable energy harvester to vibrate on the high energy orbit for low-level random excitation is of great significance to improve the application of piezoelectric energy harvesting.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] M. Safaei, H. A. Sodano, and R. Anton, "A review of energy harvesting using piezoelectric materials: state-of-the-art a decade later," *Smart Materials and Structures*, vol. 28, no. 11, Article ID 113001, 2019.

[2] S. Zhou, M. Lallart, and A. Erturk, "Multistable vibration energy harvesters: principle, progress, and perspectives," *Journal of Sound and Vibration*, vol. 528, Article ID 116886, 2022.

[3] A. Erturk and D. J. Inman, *Piezoelectric Energy Harvesting*, John Wiley & Sons, 2011.

[4] A. Erturk and D. J. Inman, "An experimentally validated bimorph cantilever model for piezoelectric energy harvesting from base excitations," *Smart Materials and Structures*, vol. 18, no. 2, Article ID 025009, 2008.

[5] F. Qian, T.-B. Xu, and L. Zuo, "Design, optimization, modeling and testing of a piezoelectric footwear energy harvester," *Energy Conversion and Management*, vol. 171, pp. 1352–1364, 2018.

[6] X. D. Xie, A. Carpinteri, and Q. Wang, "A theoretical model for a piezoelectric energy harvester with a tapered shape," *Engineering Structures*, vol. 144, pp. 19–25, 2017.

[7] G. Miao, S. Fang, S. Wang, and S. Zhou, "A low-frequency rotational electromagnetic energy harvester using a magnetic plucking mechanism," *Applied Energy*, vol. 305, Article ID 117838, 2022.

[8] Z. Li, Y. Liu, P. Yin et al., "Constituting abrupt magnetic flux density change for power density improvement in electromagnetic energy harvesting," *International Journal of Mechanical Sciences*, vol. 198, Article ID 106363, 2021.

[9] C. Wei and X. Jing, "A comprehensive review on vibration energy harvesting: modelling and realization," *Renewable and Sustainable Energy Reviews*, vol. 74, pp. 1–18, 2017.

[10] H. Qing, S. Jeong, S. Y. Jeong, T. H. Sung, and H. H. Yoo, "Investigation of the energy harvesting performance of aLambda-shaped piezoelectric energy harvester using an analytical model validated experimentally," *Smart Materials and Structures*, vol. 30, no. 7, Article ID 75017, 2021.

[11] J. Jung, P. Kim, J.-I. Lee, and J. Seok, "Nonlinear dynamic and energetic characteristics of piezoelectric energy harvester with two rotatable external magnets," *International Journal of Mechanical Sciences*, vol. 92, pp. 206–222, 2015.

[12] A. H. Hosseinloo and K. Turitsyn, "Non-resonant energy harvesting via an adaptive bistable potential," *Smart Materials and Structures*, vol. 25, no. 1, pp. 15010–15018, 2015.

[13] Z. Fang, Y. Zhang, L. Xiang, H. Ding, and L. Chen, "Complexification-averaging analysis on a giant magnetostrictive harvester integrated with a nonlinear energy sink," *Journal of Vibration and Acoustics-Transactions of the ASME*, vol. 140, no. 2, Article ID 021009, 2017.

[14] Y. Peng, Z. Xu, M. Wang et al., "Investigation of frequency-up conversion effect on the performance improvement of stack-based piezoelectric generators," *Renewable Energy*, vol. 172, pp. 551–563, 2021.

[15] K. Fan, Q. Tan, Y. Zhang, S. Liu, M. Cai, and Y. Zhu, "A monostable piezoelectric energy harvester for broadband low-level excitations," *Applied Physics Letters*, vol. 112, no. 12, Article ID 123901, 2018.

[16] A. Kumar, S. F. Ali, and A. Arockiarajan, "Exploring the beneffits of an asymmetric monostable potential function in broadband vibration energy harvesting," *Applied Physics Letters*, vol. 112, no. 23, Article ID 123901, 2018.

[17] A. Erturk, J. Hoffmann, and D. J. Inman, "A piezomagnetoelastic structure for broadband vibration energy harvesting," *Applied Physics Letters*, vol. 94, no. 25, Article ID 254102, 2009.

[18] A. Erturk and D. J. Inman, "Broadband piezoelectric power generation on high-energy orbits of the bistable Duffing oscillator with electromechanical coupling," *Journal of Sound and Vibration*, vol. 330, no. 10, pp. 2339–2353, 2011.

[19] S. Fang, S. Zhou, D. Yurchenko, T. Yang, and W.-H. Liao, "Multistability phenomenon in signal processing, energy harvesting, composite structures, and metamaterials: a review," *Mechanical Systems and Signal Processing*, vol. 166, Article ID 108419, 2022.

[20] T. Yang and Q. Cao, "Dynamics and performance evaluation of a novel tristable hybrid energy harvester for ultra-low level vibration resources," *International Journal of Mechanical Sciences*, vol. 156, pp. 123–136, 2019.

[21] N. Tran, M. H. Ghayesh, and M. Arjomandi, "Ambient vibration energy harvesters: a review on nonlinear techniques for performance enhancement," *International Journal of Engineering Science*, vol. 127, pp. 162–185, 2018.

[22] R. L. Harne and K. W. Wang, "A review of the recent research on vibration energy harvesting via bistable systems," *Smart Materials and Structures*, vol. 22, no. 2, Article ID 023001, 2013.

[23] S. C. Stanton, C. C. McGehee, and B. P. Mann, "Nonlinear dynamics for broadband energy harvesting: investigation of a bistable piezoelectric inertial generator," *Physica D: Nonlinear Phenomena*, vol. 239, no. 10, pp. 640–653, 2010.

[24] S. C. Stanton, B. A. Owens, and B. P. Mann, "Harmonic balance analysis of the bistable piezoelectric inertial generator," *Journal of Sound and Vibration*, vol. 331, no. 15, pp. 3617–3627, 2015.

[25] Q. He and M. F. Daqaq, "Influence of potential function asymmetries on the performance of nonlinear energy harvesters under white noise," in *Proceedings of the International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, vol. 46391, Article ID V006T10A060, 2014.

[26] P. Kim, Y.-J. Yoon, and J. Seok, "Nonlinear dynamic analyses on a magnetopiezoelastic energy harvester with reversible hysteresis," *Nonlinear Dynamics*, vol. 83, no. 4, pp. 1823–1854, 2016.

[27] N. Yu, H. Ma, C. Wu, G. Yu, and B. Yan, "Modeling and experimental investigation of a novel bistable two-degree-of-freedom electromagnetic energy harvester," *Mechanical Systems and Signal Processing*, vol. 156, Article ID 107608, 2021.

[28] G. Sebald, H. Kuwano, D. Guyomar, and B. Ducharne, "Simulation of a Duffing's oscillator for broadband piezoelectric energy harvesting," *Smart Materials and Structures*, vol. 20, no. 7, pp. 75022–75038, 2011.

[29] G. Sebald, H. Kuwano, D. Guyomar, and B. Ducharne, "Experimental Duffing oscillator for broadband piezoelectric energy harvesting," *Smart Materials and Structures*, vol. 20, no. 10, pp. 102001–102010, 2011.

[30] H. Li, H. Ding, X. Jing, W. Qin, and L. Chen, "Improving the performance of a tri-stable energy harvester with a staircase-shaped potential well," *Mechanical Systems and Signal Processing*, vol. 159, Article ID 107805, 2021.

[31] X. Ma, H. Li, S. Zhou, Z. Yang, and G. Litak, "Characterizing nonlinear characteristics of asymmetric tristable energy harvesters," *Mechanical Systems and Signal Processing*, vol. 168, Article ID 108612, 2022.

[32] G. Wang, Y. Ju, W.-H. Liao, Z. Zhao, Y. Li, and J. Tan, "A hybrid piezoelectric device combining a tri-stable energy harvester with an elastic base for low-orbit vibration energy harvesting enhancement," *Smart Materials and Structures*, vol. 30, no. 7, Article ID 075028, 2021.

[33] L. Tang and J. Wang, "Size effect of tip mass on performance of cantilevered piezoelectric energy harvester with a dynamic magnifier," *Acta Mechanica*, vol. 228, no. 11, pp. 3997–4015, 2017.

[34] D. W. Man, D. H. Xu, X. C Kuang, X. F. Kang, Q. H. Xu, and Y. Zhang, "Analysis of Dynamic Characteristics of Tristable Piezoelectric Energy Harvester Based on the Modified Model," *Mathematical Problems in Engineering*, 2021.

[35] D. Man, H. Xu, G. Xu, D. Xu, L. Tang, and Q. Xu, "Dynamic characteristics analysis of tri-stable cantilever piezoelectric energy harvester with a novel-type dynamic amplifier," *International Journal of Heat and Technology*, vol. 40, no. 2, pp. 619–626, 2022.

Research Article

Quantitative Analysis of Broken Rotor Bars in Cage Motor Based on Energy Characteristics of Vibration Signals

Jie Shi [ID], Haifeng Shen [ID], and Zhenkai Ding [ID]

School of Electronics and Information Engineering, Suzhou University of Science and Technology, Suzhou, Jiangsu Province, China

Correspondence should be addressed to Haifeng Shen; 1195534012@qq.com

The rotor, as the power output device of a cage motor, is subject to a type of invisible fault, BRB, during long-term use. The conventional motor vibration signal fault monitoring system only analyzes the rotor qualitatively for the fault of BRBs and cannot evaluate the fault degree of BRBs quantitatively. Moreover, the vibration signal used for monitoring has nonstationary and nonlinear characteristics. It is necessary to manually determine the time window and basis function when extracting the characteristics of the time-frequency domain. To address these problems, this paper proposes a method for quantitative analysis of BRBs based on CEEMD decomposition and weight transformation for feature extraction and then uses the AdaBoost to construct a classifier. The method applies CEEMD for adaptive decomposition while extracting IMFs' energy as the initial feature values, uses OOB for contribution evaluation of features to construct weight vectors, and performs a spatial transformation on the original feature values to expand the differences between the feature vectors. To verify the effectiveness and superiority of the method, vibration signals were collected from motors in four BRB states to produce rotor fault data sets in this paper. The experiment results show that the feature extraction method based on CEEMD decomposition and weight transformation can better extract the feature vectors from the vibration signals, and the constructed classifier can accurately perform quantitative analysis of BRB fault.

1. Introduction

Induction motors are widely used in industrial production because of their economical, reliable, and ease of control. In the course of their use, they are subject to mechanical, thermal, and electrical stresses as well as environmental factors, which inevitably lead to failures [1]. The induction motors are subject to a wide range of failures. Among the various possible failures of induction motors, a broken rotor bar (BRB) is a kind of hidden fault [2]. This means that a single or small number of broken rotor bars do not have a noticeable effect on the function of the motor, as the current that should flow from the broken bar can be dispersed to the adjacent bars, thus ensuring normal motor operation. However, if they are not repaired at an early stage of failure, this can lead to excessive vibration, energy loss, and other problems. Also, under electrical and thermal stress, the number of BRB expands, and a sweeping failure may occur under centrifugal force, leading to motor shutdown and even causing injury to personnel and machinery.

Unlike the design for the stator, the design and manufacture of cage rotors have remained virtually unchanged for many years, and rotor failures now account for about 5–10% of all induction motor failures [3]. There are two types of cage rotors: cast and fabricated. Previously, cast rotors were only used on small machines. With the advent of cast ducted rotors, cast rotors can even be used for motors under the 3,000 kW, and the field of application of cage motors continues to grow [4]. Therefore, the monitoring of rotor health and the quantitative analysis of BRB faults in motors in using is a very important and challenging task.

Several condition monitoring techniques have been developed to monitor the health of motors, including air gap torque monitoring, noise measurement, thermal

monitoring, partial discharge measurement, instantaneous angular velocity, instantaneous power, surge testing, vibration monitoring, and current monitoring [5]. Vibration monitoring and noise monitoring have been extensively investigated for their ability to detect both electrical and mechanical faults. When a fault occurs, the vibration characteristics of the machine will change [6]. Fault diagnosis based on vibration signals is a well-established field covering a wide range of techniques, mainly using time, frequency, and time-frequency feature extraction. Acoustic noise monitoring is performed by investigating the noise spectrum, and the acoustic signal shows the acoustic stress waves emerging from the energy release due to microstructural changes in the material or structure of the motor in different states [7], whereas noise monitoring techniques are sensitive to environmental noise during use, filtering, signal identification, and noise reduction become critical at the pre-processing stage, resulting in the technique being far less effective than laboratory results during application, so this monitoring technique is often used as a secondary monitoring technique [8].

As a multifactor coupled rotating system, the vibration signal is a good representation of the motor's operating conditions under different faults. When a motor has a BRB fault, the vibration signal will produce torque pulsations and speed oscillations [9]. At present, the research work on BRB faults feature recognition is mainly based on this. Sabbaghian-Bidgoli and Poshtan combined the wavelet packet decomposition (WPT) and the Hilbert transform to propose the Hilbert–Huang transform (HHT) to detect the characteristic frequency of the fault [10]. The results show that this hybrid technique exhibits better performance than the conventional wavelet packet transform. Moon and Dae used the finite element method to analyze the effect of vibration on motor life and performance due to thermal deformation of the frame caused by heat during motor operation [11]. Liu et al. used empirical mode decomposition (EMD) thresholding to denoise the vibration signals and applied probabilistic neural networks (PNN) to classify them [12]. The signal is decomposed by ensemble empirical mode decomposition (EEMD) with curve coding techniques to achieve stable identification of motor faults [13]. Ke et al. used a genetic algorithm to optimize the white noise amplitude in EEMD for bearing fault diagnosis in electric motors [14]. Miceli et al. used time-frequency analysis methods to apply axial and radial vibration signals for BRB faults diagnosis [15]. Xiao et al. proposed a feature extraction method combining complete ensemble empirical mode decomposition (CEEMD) and composite multiscale basic scale entropy (CMBSE), which can extract weak signal features under strong noise background [16]. A specific Zhao-Atlas-Marks distribution was proposed by Climente-Alarcón et al. to extract the harmonic components associated with BRB faults [17]. All of these research works have achieved effective detection of BRB faults using vibration signals. The above methods solve the problem of fault diagnosis of motors in specific environments and are able to identify BRB faults in motors, but they are qualitative fault analysis and cannot achieve a quantitative diagnosis of breakage and rotor health discrimination, which is not conducive to the

development of inspection and maintenance plans for continuously operating equipment.

The amplitude, frequency, and axial trajectory of the vibration signal can reflect the characteristics of the fault [18]. When using common methods such as correlation analysis and Fourier transform for signal feature mining, the analysis and processing of nonlinear nonperiodic signals can only extract features in the time or frequency domain, resulting in a large amount of information being lost. The torque output of the motor system is often affected by load changes, which makes nonstationary and nonlinear characteristics of vibration signal more significant [19]. The wavelet transform can be used to improve this phenomenon through the choice of the time window in signal analysis, but how to determine the appropriate wavelet basis and the number of decomposition layers is an important issue. On the other hand, the wavelet bases are explicitly chosen in the wavelet transform process and cannot be adaptively adjusted during the decomposition process according to the signal characteristics [20]. Unlike the wavelet transform, empirical mode decomposition is an adaptive processing technique that can be applied to the analysis of complex signals based on the inherent characteristics of the signal [21]. Although it avoids the choice of decomposition layers and wavelet bases and has multiresolution analysis capability, while there are sudden changes or disturbances in the signal, part of the timescale will be lost, leading to a severe modal mixing phenomenon [22]. As an important problem in EMD, the improvement method of modal mixing has been one of the current research hotspots. Among the many improvement algorithms, EEMD is a noise-friendly method, known for its ability to suppress modal mixing, but its noise suppression mechanism will produce great residual noise, which creates obstacles for further feature extraction [23, 24].

To address the above problems and, considering the complexity of the rotor fault signal, to improve the stability of feature extraction and ensure the accuracy of quantitative analysis, this paper starts from the vibration signal of the motor, extracts the energy parameters of different layers by adaptive decomposition through CEEMD, and adjusts the recognition weights of each layer of energy through the random forest out-of-bag (OOB) estimation method. Finally, An AdaBoost strong classifier is used to quantitatively identify the BRB faults.

2. Complete Ensemble Empirical Mode Decomposition

2.1. Empirical Mode Decomposition. Empirical mode decomposition is an adaptive algorithm that decomposes a signal into a series of intrinsic mode functions (IMFs) based on the characteristics of the signal itself and is particularly suitable for the analysis of nonlinear and nonstationary signals. The decomposition process is only based on the original signal data, and the IMFs are separated layer by layer from high to low according to the oscillation law of the signal. Finally, the remaining component with the longest oscillation period is regarded as the residual. The IMFs should satisfy two conditions: (1) the number of extremes

and the number of crossing zeros must be equal or differ by at most once, and (2) at any point, the envelope defined by the local maximum and the envelope defined by the local minimum have an average value of zero.

For a given signal $x(t)$, the EMD method first extracts the local extremes to form the upper and lower envelopes and then interpolates the local maximum and local minimum, respectively. The average of the upper and lower envelopes is then calculated at $m(t)$, and the average is subtracted from the signal using $h(t) = x(t) - m(t)$. The signal $x(t)$ is replaced by $h(t)$, and the above two steps are repeated until $h(t)$ satisfies the two conditions of the IMFs; then the process is completed. The residual $r(t)$ is defined as $r(t) = x(t) - h(t)$ and considered as a new signal continuously decomposed by the above steps. If the residual satisfies the stopping criteria, the decomposition process ends. The original signal $x(t)$ will eventually be decomposed into multiple IMFs and a final residual $r(t)$ as shown in the following equation:

$$x(t) = r(t) + \sum_{i=1}^{n} IMF_i(t). \tag{1}$$

Although the EMD can be adaptively adjusted to the inherent characteristics of the signal, it has two important drawbacks: (1) the overshoot or undershoot of the spline fitting method may lead to large errors, which in turn affect the structure of the IMFs, and (2) if there are sudden changes or perturbations in the signal, the EMD will lose some of its timescale, which in turn will cause severe modal mixing.

2.2. Ensemble Empirical Mode Decomposition.
The ensemble empirical mode decomposition algorithm is an improved algorithm that effectively solves the modal mixing problem in EMD. The algorithm adds N nonrepeating Gaussian white noises of rank L to the original signal and employs EMD to modal decompose the noise-added signal, taking the average value of the corresponding IMFs as the final result. The addition of white noise enriches the spectral content of the signal and can effectively prevent modal mixing. The algorithm assumes that there is a sufficient number of noise additions to ensure that the final noise energy is zeroed when the signal is reconstructed, but in practice, due to the limitation of computing resources, the number of noise additions is often insufficient to achieve noise cancellation, making the IMFs contain residual noise. At the same time, the introduction of noise breaks the energy distribution between the layers of the original signal, making the fault features extracted from the IMFs no longer reliable.

2.3. Complete Ensemble Empirical Mode Decomposition.
To address the problems in the application of EMD and EEMD, Colominas et al. proposed a complete ensemble empirical mode decomposition (CEEMD) [25]. The method adds noise to each decomposition stage and uses the original signal minus the noise to construct the complementary signal, canceling out the problem of noise present in the signal by taking the mean of the two signals' EMD. Not only does it solve the mode mixing problem, but it also provides an accurate reconstruction of the original signal. The operational steps are as follows:

Step 1. Generate a random Gaussian white noise $n(t)$ with a noise level of L and construct a noise-added signal $s_a(t)$ and its complementary signal $s_b(t)$:

$$\begin{cases} s_a(t) = x(t) + n(t), \\ s_b(t) = x(t) - n(t). \end{cases} \tag{2}$$

Step 2. Decompose $s_a(t)$ and $s_b(t)$ with EMD to get the layers IMF_a^i and IMF_b^i and take the average value of both as IMF^i:

$$IMF^i = \frac{IMF_a^i + IMF_b^i}{2}. \tag{3}$$

Step 3. Repeat steps 1 and 2 N times but add a different white noise each time.

Step 4. Calculate the average value of the corresponding IMFs as the final result. The original signal $x(t)$ will be decomposed into the same form as the EMD, as shown in equation (1).

3. Feature Weight Transformation

3.1. Feature Extraction.
The objective of feature extraction is to determine a feature vector that characterizes the variability between samples based on a given set of samples so that the samples can be correctly classified [26]. The vibration signal contains a large amount of information that characterizes the state of the motor rotor, from which a feature vector can be extracted for state classification. However, due to the complex nonsmooth, nonlinear nature of the vibration signal, although CEEMD has adaptively decomposed the signal into IMFs in different frequency domains, the features are still not obvious and cannot be directly used for quantitative analysis of the BRB fault. Therefore, we need to further extract the fault features from the IMFs.

The core idea of CEEMD decomposition is to eliminate modal mixing by adding white noise to the signal to be decomposed and to eliminate the effect of noise on the original signal by complementary signals. Therefore, IMF itself has antinoise characteristics, and there is no need for filtering operation in feature extraction. The IMFs are essentially a series of central envelopes of similar frequencies obtained by subtracting the signal to find the difference, characterizing the degree of oscillation between the different modes. The vibration signal of a motor has a strong structural response, and this vibration is an inherent property of the mechanical state of the motor. When the motor rotor changes in response to different BRB faults, the response of the vibration signal and the energy distribution between the IMFs will also change. By analyzing the energy distribution of IMF in each layer, the number of BRBS of the motor can be determined. Lu et al. used the Teager–Kaiser energy factor to analyze the main oscillation frequency of the stator harmonics to quantify the BRBs in a specific state,

which indirectly demonstrates that the occurrence of BRBs affects the energy distribution between the IMFs of the motor vibration signal distribution [27].

The energy characteristic formula of the IMFs is shown in the following equation:

$$E_i = \frac{\sum_{j=1}^{M} |c_j|}{\sum_{1}^{k} \sum_{j=1}^{M} |c_j|}. \tag{4}$$

Where $i = 1, 2, \ldots, k$, k is the number of IMF components, M is the data length of the IMF component, c_j is the j-th data of the IMF component, and E_i is the proportion of the i-th IMF component in the total energy, and its value is taken in the interval $[0, 1]$. If E_j is used directly as the eigenvalue to form the eigenvector characterizing the state of the motor, this will lead to local stacking in the characterization space, making the eigenvectors with clear differences extremely similar. Therefore, the eigenvectors should first be logarithmically transformed with respect to each other to expand the characterization space of the eigenvectors. To further improve the ability of the feature vectors to characterize the state of BRB, this paper uses the random forest out-of-bag (OOB) estimation method to measure the contribution of features in state identification and uses this as a basis to transform the weights of the features to expand the interclass variation of the samples.

3.2. Feature Importance Metric. The random forest OOB estimation method is an integrated machine learning method using decision trees as base classifiers [28, 29]. Its training steps are as follows:

Step 1. The training sample set and OOB data set are generated from the original sample set by the bootstrap resampling method

Step 2. Train the decision tree as the base classifier using the training data set and test its OOB error using the OOB data set, denoted as $errOOB_1$

Step 3. Add noise to the feature Q of all samples in the OOB data set and calculate the OOB error of the decision tree again, denoted as $errOOB_2$

Step 4. Assuming there are K decision trees in the forest, the importance of feature Q is $Q_p = (errOOB_2 - errOOB_1)/K$

Q_p is possible to characterize the importance of a feature because, if the accuracy of the OOB data drops significantly when random noise is added (i.e., $errOOB_2$ goes up), this indicates that the feature has a significant impact on the prediction of the sample, which in turn indicates that it is of high importance. Each feature is judged in the random forest training process, with the most widely used measure of feature importance being "ranked importance." This is based on the change in the classification error rate of the random forest model before and after a feature label is changed in the OOB data [30].

If a feature is used for classification and the results are less different from those of random classification, the

contribution of this feature to the classification effect is considered to be minimal. Weakening these features empirically has little effect on classification accuracy. Q_p is a good measure of how well the features contribute to the classification effect of a sample, and using them as weights to scale and transform the feature values can expand the interclass differences or spatial distances of the samples and improve the effectiveness of the samples when classified or clustered. This is certainly a good way to deal with samples with small differences in feature vectors between classes.

4. AdaBoost Quantitative Fault Classifier

Integration learning is an important area of research in machine learning. Instead of trying to obtain a single optimal classifier, it trains a set of weak classifiers from existing data and then combines the weak classifiers into a strong classifier according to a certain strategy. AdaBoost is one of the representative algorithms of integrated learning [31], which can take a simple, coarse prediction method with low accuracy and construct a complex prediction method with high accuracy by specific rules [32]. The aim of this paper is to quantitatively analyze the BRB status of motor rotors. The task is achieved not only by simple qualitative analysis of faults but also by further classification of small difference samples, which cannot or is difficult to achieve with a single classifier alone. Therefore, we can construct multiple weak classifiers, use threshold control to discriminate whether the sample belongs to that small class, and integrate the weak classifiers into strong classifiers through strategy control to finally achieve accurate classification.

There are two main aspects of integration learning: the update strategy of sample weights and the combination strategy of base classifiers. The feature weight transformation using random forest OOB estimation described in Section 3.2 is a sample weight update strategy and can be considered as a pre-processing of samples. The use of decision trees as base classifiers can maintain the consistency of the classification system, and the steps for composing a strong classifier are as follows:

Step 1. Given the data set $S = \{(X_1, Y_1), \ldots (X_i, Y_i), \ldots, (X_n, Y_n)\}$, where $X_i = (IMF_1\prime, IMF_2\prime, \ldots, IMF_8\prime)$ is the feature vector, Y_i is the classification flag, and n is the number of training samples

Step 2. Initialize the sample weight distribution vector for k iterations $D_k(i) = 1/n$

Step 3. Train the base classifier $T_k(X_i) = \arg\min_T \sum_{t=1}^{k} D_k(i)[T_k(X_i) \neq Y_i]$ and its classification error rate $\varepsilon_k = \sum_{i=1}^{n} D_t(i)[T_k(X_i) \neq Y_i]$, where $[T_k(X_i) \neq Y_i]$ is the indicator function that takes on a value of 1 when $T_k(X_i) \neq Y_i$ and 0 otherwise

Step 4. Calculate the weights of the base classifier $\alpha_k = 1/2 \ln 1 - \varepsilon_k/\varepsilon_k$

Step 5. Update the sample weights $D_{k+1}(i) = D_k(i)^{-\alpha_i}/\sum_{k=1}^{m} D_k(i)$ to obtain the strong classifier $H(X) = \text{sign}(\sum_{t=1}^{k} \alpha_k T_k(X))$
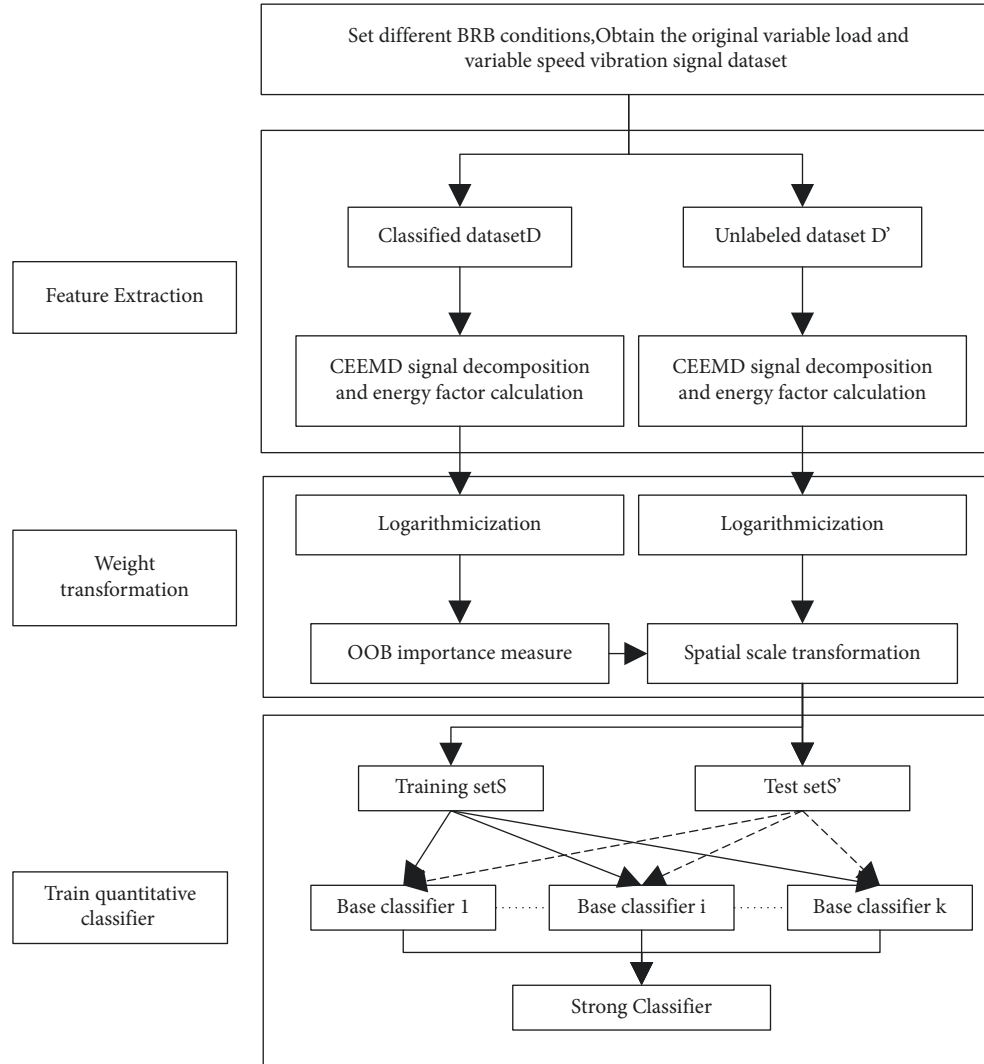
FIGURE 1: Flow chart for vibration signal analysis.

Although the AdaBoost classification algorithm has a certain sample weight update mechanism, a large number of base classifiers are required to support a reliable classification effect. Using out-of-bag estimation to pre-process the data according to the importance of weight transformation can reduce the number of base classifiers required by the quantitative classifier and improve the discriminative efficiency of the model.

5. BRB Quantitative Analysis

The flow of quantitative analysis of BRB faults in cage motor rotors based on the energy characteristics of vibration signals is shown in Figure 1. The process is divided into three main stages, namely feature extraction, weight transformation, and classifier training.

In the first stage, by manually setting the operating state of the fault simulation motor and performing three times mean filtering during the signal acquisition process, the motor vibration signals in different states are obtained and decomposed using CEEMD to obtain a series of IMFs, and

the energy factor of each layer is calculated as the primary feature for quantitative rotor analysis.

The second stage logarithm the primary features and uses OOB estimation to obtain a weight vector for the feature space scale transformation using the OOB estimation method for the feature importance measure on the annotated data sets.

The third stage is the training of the AdaBoost rotor broken bar quantitative classifier, where k base classifiers are trained from the training set to combine into a strong classifier, and the performance is evaluated using the test set.

6. Experimental Analysis and Discussion

Figure 2 shows a cage motor rotor vibration signal acquisition system consisting of M1, three-phase cage asynchronous motor; M2, magnetic powder brake; C1, inverter; C2, tension controller; S1, accelerometer; S2, photoelectric tachometer; and a computer. The experiments simulate 0–3 BRBs states by manually punching damage to the motor rotor. In order to obtain the vibration signals of the motor in
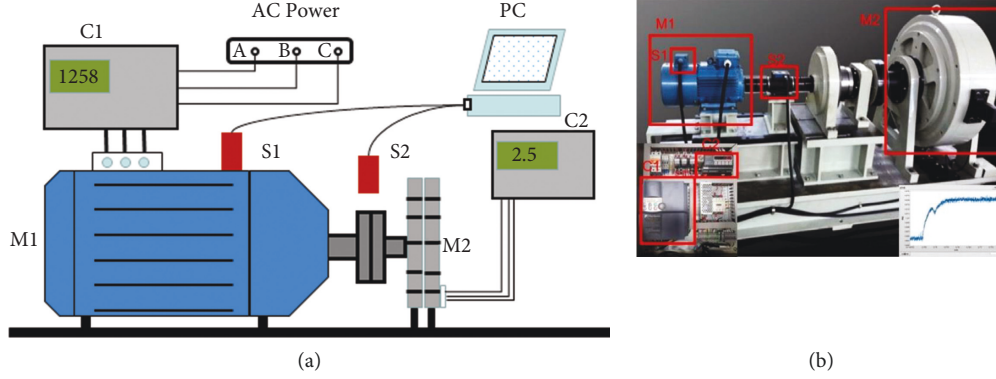
Figure 2: Signal acquisition system for broken rotor bars in cage motors: (a) sketch map and (b) physical map.

different states, in addition to setting the corresponding faults, the vibration signals under variable load and variable speed operation are obtained by controlling the tension controller and the frequency converter, respectively.

6.1. CEEMD Performance Analysis.

In order to verify the advantages of CEEMD over EMD and EEMD in signal decomposition, five sets of actual collected motor BRB vibration signals were decomposed, and the root mean square error (RMSE), mean absolute error (MAE), and Nash–Sutcliffe efficiency coefficient (NSE) are the evaluation indexes of the decomposition accuracy, and the three error evaluation indexes were calculated as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y - y_i)^2},$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y - y_i|, \qquad (5)$$

$$\text{NSE} = 1 - \frac{\sum_{i=1}^{n} (y - y_i)^2}{\sum_{i=1}^{n} (y - \overline{y})^2},$$

where y is the original value, y_i is the decomposition value, n is the number of original values, $\overline{y}$ is the average of n original values, and $NSE \le 1$, and the closer to 1, the higher the decomposition accuracy. The decomposition effects of EMD, EEMD, and CEEMD are shown in Table 1. Comparing the data performance of the three, it can be found that the decomposition error and algorithm stability of EEMD are the worst because EEMD adds random noise to the original signal, which is not completely controllable and the number of noise additions cannot be increased indefinitely to eliminate the influence of noise, while CEEMD also adds noise, but the noise energy is removed from the decomposition components through the signal complementary mechanism. The CEEMD also adds noise, but the noise energy is removed from the decomposition component by a signal complementary mechanism, which solves the modal mixing problem while perfectly maintaining the actual energy distribution of the signal.

Table 1: Comparison of the decomposition effects of EMD, EEMD, and CEEMD.

	No.	RMSE	MAE	NSE
EMD	1	0.04251	0.03330	0.99995
	2	0.03790	0.02887	0.99997
	3	0.04737	0.03955	0.99995
	4	0.04325	0.03504	0.99993
	5	0.06053	0.04897	0.99994
	Standard deviation	0.00862	0.00763	$1.48E-5$
EEMD	1	0.37228	0.36575	0.99624
	2	0.40699	0.39852	0.99628
	3	0.81135	0.80659	0.98661
	4	0.29647	0.28544	0.99681
	5	1.48709	1.48375	0.96350
	Standard deviation	0.49619	0.498743	0.01428
CEEMD	1	0.01279	0.00944	1.0000
	2	0.01185	0.00907	1.0000
	3	0.01066	0.00822	1.0000
	4	0.01537	0.01112	0.9999
	5	0.01822	0.01275	0.9999
	Standard deviation	0.00302	0.00180	$5.47E-5$

Table 2 shows the data when using the OOB estimation method to sample the five subsets generated by the variable load data sets for the weight metric, with the sampling process being artificially set to have an unbalanced number of samples across the BRBs. The data in the table shows that the features have the same ranking in terms of their ability to characterize the state on different subsets and that the unbalanced data does not affect the ability of OOB to measure the importance of the features, while the standard deviation (SD) with minimal importance of each feature indicates not only the stability of the OOB algorithm but also the ability of the selected features to characterize the state of the motor rotor in a very stable manner.

6.2. Analysis of Variable Load Conditions.

A set of vibration signals in four states was randomly selected from the variable load data sets, and the four types of signals were first decomposed using CEEMD, and the IMFs were sorted from highest to lowest according to the signal frequencies, as shown in Figure 3, where S is the original signal, IMF is the

TABLE 2: Out-of-bag estimated feature weighting metrics.

No.	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6	IMF7	IMF8
1	0.18293	0.16138	0.12307	0.12893	0.13046	0.10212	0.08357	0.08754
2	0.18618	0.16315	0.12426	0.12886	0.12973	0.10406	0.07993	0.08383
3	0.18623	0.15852	0.12397	0.12872	0.13185	0.10502	0.08087	0.08482
4	0.18976	0.16221	0.11840	0.12673	0.13012	0.10100	0.08440	0.08738
5	0.19356	0.16351	0.12032	0.12678	0.12009	0.09960	0.08670	0.08944
SD	0.00406	0.00199	0.00254	0.00114	0.00474	0.00221	0.00273	0.00226
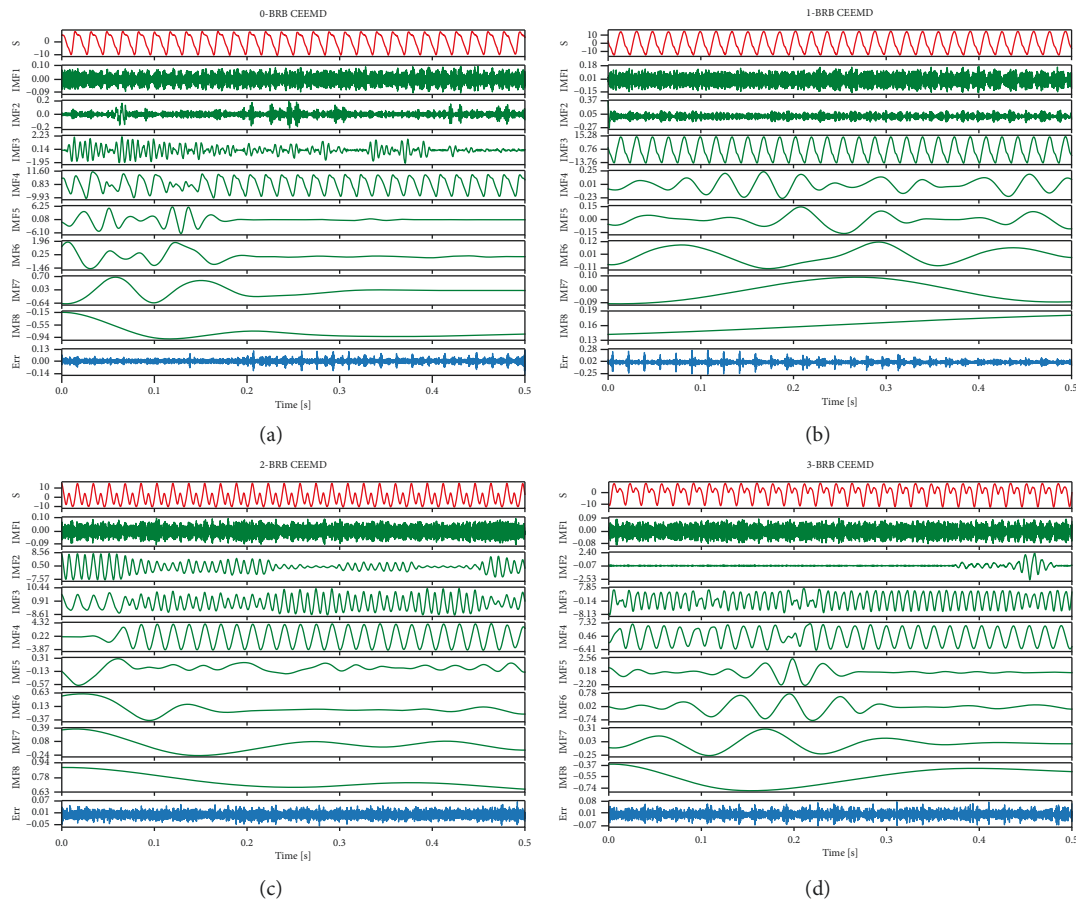
FIGURE 3: IMF under CEEMD of vibration signals for four rotor states under variable load conditions: (a) 0 BRB CEEMD decomposition results, (b) 1 BRB CEEMD decomposition results, (c) 2 BRB CEEMD decomposition results, and (d) 3 BRB CEEMD decomposition results.

decomposed modal function of each layer, and Err is the final decomposition residual. Although there is no significant difference between the original signal S at different BRB states, the difference in the energy distribution of the vibration signal at different BRBs can be directly observed by the decomposition of each IMF, and the lower the frequency the more obvious the difference.

The energy features of the intrinsic modal functions of each layer are calculated as signal features and form the input feature vector. The importance of the feature vectors is calculated as weights after logarization to obtain the weight transform vectors of the feature values. The initial feature vectors are spatially transformed according to the weights to form the feature vectors used for state identification. The energy distribution of each IMF layer before and after weighting is shown in Figure 4. To increase visualization the energy distribution is associated with the [0, 255] thermal interval, where each row is a complete feature vector, 0–400 corresponds to 0–3 BRB, and each BRB subset has 100 samples. As shown in Figure 4(a), the original feature energy distribution is more extreme, with IMF1, IMF2, IMF3, IMF7, and IMF8 distributed in the 0–50 range, while IMF4, IMF5, and IMF6 occupy the 150–200 interval. The huge uneven distribution of features makes the intersample capability degrade, which will further lead to the failure of BRB quantitative analysis. In contrast, as shown in Figure 4(b), the feature vectors have been logarithmically and the weight redistributed by importance; the energy distribution is more even; and the intra- and interclass differences of the samples have been reflected.
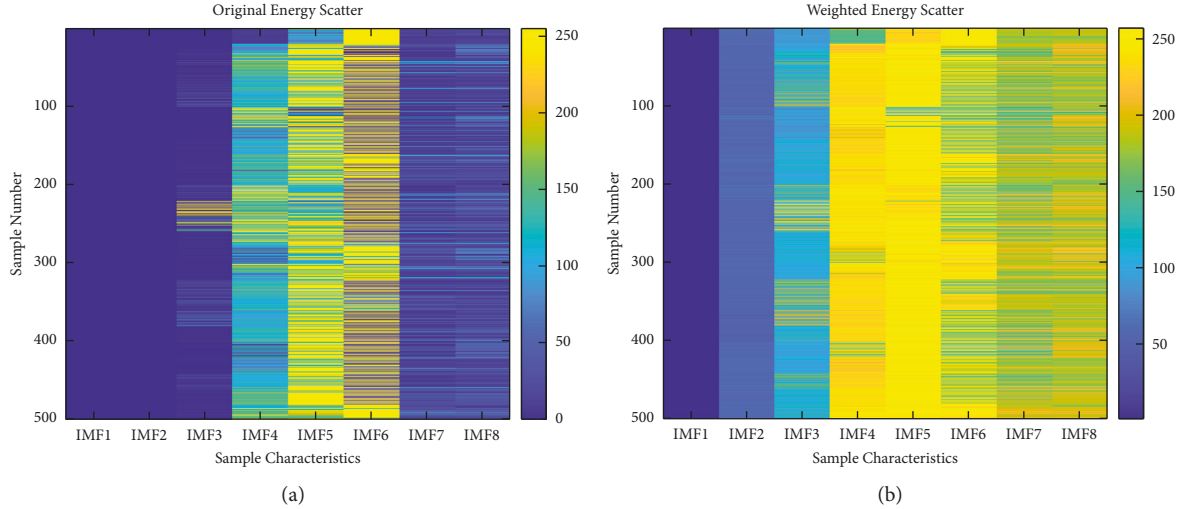
FIGURE 4: Energy distribution before and after IMF weighting for variable load conditions: (a) original IMF energy distribution and (b) weighted IMF energy distribution.

Randomly select 0.5 s in the 1BRB data set and analyze according to the above process to obtain their feature weight, as shown in Table 3. From the table, it can be seen that for the vibration signal data sets collected at different time periods, the ranking of feature importance by weights is not affected by load changes and the feature importance measurement by weights has good stability.

Similarly, select time windows of 0.25 s, 0.5 s, 0.75 s, and 1 s, respectively, in the 1BRB data set. The analysis was carried out according to the above process, and their weights were derived as shown in Table 4, where the eigenvalues $x_1 \sim x_8$ are the energy characteristics of the IMF in descending order of frequency. As can be seen from the table, the ranking of the importance of the weights on the features did not change for vibration signals acquired in different length time windows. It can be seen that the extraction of fault features by CEEMD for signal decomposition is not affected by the time window, does not require manual determination of the basis function, and is highly adaptive.

Based on the weights obtained in Tables 3 and 4 for different time periods and different time windows, the joint average of 0.17655, 0.16247, 0.12481, 0.13366, 0.12039, 0.09513, 0.08668, and 0.10030 were taken as the final weights determined by this method. The intra- and interclass differences of the feature vectors after the weighted transformation were further analyzed. Considering that the feature space corresponding to the weighted transformed feature vectors has been changed and there is scale variation, it is obviously inappropriate if the general Euclidean distance method is chosen to measure the spatial distance between samples at this point. Therefore, we select the vector mode of $Xs = [1, 1, 1, 1, 1, 1, 1, 1]$ as the base in the original feature space and the vector mode of $Xs' = [0.17655, 0.16247, 0.12481, 0.13366, 0.12039, 0.09513, 0.08668, 0.10030]$ as the base in the transformed feature space and use the normalized ratio between the vector mode of the feature vectors and the base in each space as the index to measure the effect of the spatial transformation. The normalized ratios before and after the weighting transformation are shown in Figure 5. The overall improvement of the vector modes of the weighted feature vectors with respect to the baseline vectors is evident, and the original interclass differences are still maintained, which indicates that the weighting transformation can achieve a uniform distribution of features without breaking the expression ability of the features.

Traditional classification algorithms use a supervised learning model that relies on the manual selection of the sample data set. To demonstrate the ability of feature extraction to characterize the samples, this paper uses unsupervised learning to cluster the feature vectors that have undergone a weight weighting transformation by different clustering algorithms with the clustering accuracy shown in Table 5 for the CEEMD weighting transformation. In order to further verify the advanced nature of the proposed method, the fault characteristics extracted from CEEMD and discrete wavelet analysis without weight transformation are compared. According to the description of reference [27], the best parameters of the wavelet method are selected: the wavelet base is db3 and the number of decomposition layers is 8. The clustering accuracy of the features extracted by the different methods is shown in Table 5. The clustering effect of the CEEMD weight transform is the best in terms of both individual clustering algorithms and overall average accuracy, with an average improvement of 4.3% compared to the other methods. In contrast, the accuracy of the feature vectors extracted by the other methods was average, with a large difference in accuracy on different clustering algorithms, and even a failure. The failure phenomenon may occur because the CEEMD decomposition shows modal mixing and the wavelet bases of wavelet decomposition have inconsistent time windows under different signals. These problems lead to lopsided feature extraction, resulting in huge deviations in performance under different clustering algorithms.

TABLE 3: Weights corresponding to the feature values of the data set for different time periods.

Feature	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
0.5 s Data set 1	0.18196	0.16391	0.12252	0.12918	0.11843	0.09120	0.08841	0.10439
0.5 s Data set 2	0.17606	0.16068	0.12002	0.14404	0.11226	0.09302	0.08957	0.10435
0.5 s Data set 3	0.17056	0.16008	0.12291	0.14097	0.12642	0.09119	0.08708	0.10078

TABLE 4: Weights corresponding to the eigenvalues of the data set under different time windows.

Feature (s)	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
0.25	0.16635	0.16490	0.13442	0.13192	0.11311	0.10632	0.07856	0.10441
0.5	0.18196	0.16391	0.12252	0.12918	0.11843	0.09120	0.08841	0.10439
0.75	0.17617	0.16245	0.12817	0.13147	0.12357	0.09089	0.09111	0.09617
1	0.18277	0.16139	0.12311	0.12889	0.13053	0.10208	0.08363	0.08762
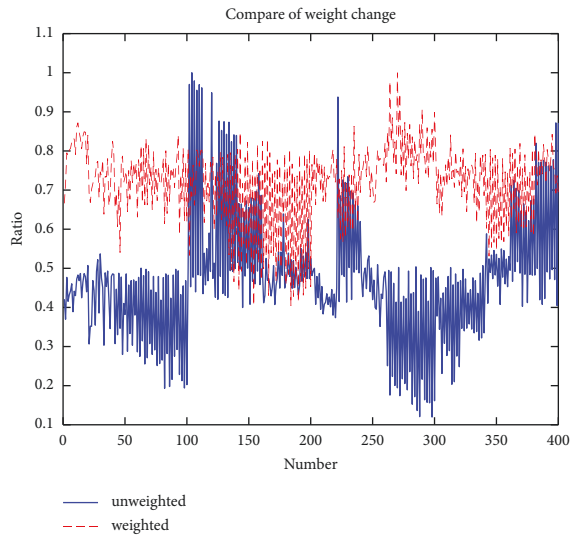
FIGURE 5: Comparison of sample distances before and after weight transformation for variable load conditions.

The ultimate aim of feature weight transformation is to enable better quantitative analysis of rotor break bars. The unsupervised learning approach, although good at analyzing intra- and interclass differences in samples, has the problem of not being able to quantify the effect in practical engineering use. Supervised learning is therefore required for model evaluation and application of classification models to practical engineering. Figure 6 shows the test results of training a classification model using a ratio of 7:3 to divide the data set to generate a training set and a test set. The spatial distribution of the four different samples can be seen in the figure, and although there is still some interclass coupling between the 0 BRB and 3 BRB samples, the classifier still achieves an accuracy of 100%.

6.3. Analysis of Variable Speed Working Conditions.

The data sets collected under variable speed conditions were analyzed following the process in Section 6.2. Its weighted vector obtained by OOB estimation is [0.38491275, 0.19737257, 0.071427953, 0.082986883, 0.07176596, 0.072956693, 0.063359313, 0.055217877], and the weighted pre- and post-IMF interlayer energy distribution is shown in Figure 7. The

TABLE 5: Clustering accuracy of features extracted by different methods.

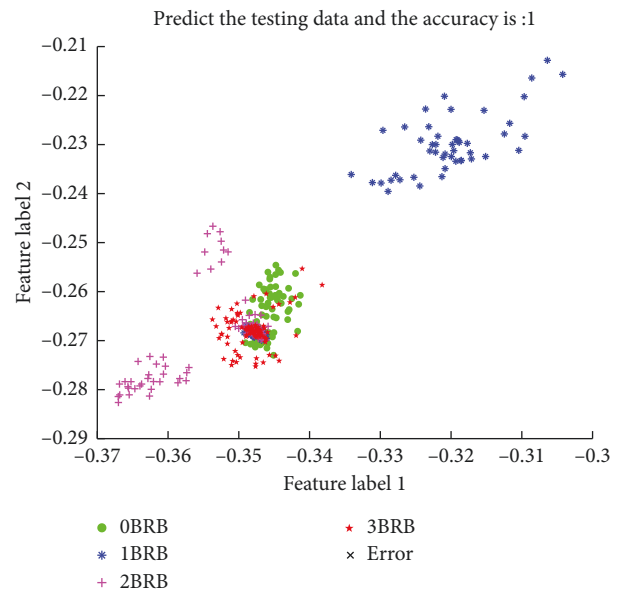| Clustering methods	CEEMD weighting transformation (%)	CEEMD (%)	Wavelet method (%)
K-means	37	36.75	36.75
Mean-shift	24.5	20.5	25
AGG	35.25	24.5	20.5
GMM	35	30.75	30.75
Spectral clustering	34.75	32.5	32
Average	33.3	29	29

FIGURE 6: Classification effect of the AadaBoost broken bar quantitative classifier for variable load conditions.

results of the AdaBoost quantitative classifier are shown in Figure 8. The coupling between the 0 BRB, 1 BRB, and 2 BRB samples is still present, but the coupling is significantly lower compared to the data under variable torque conditions, and the classification accuracy is 99.7%.

The results of the above analysis show that the method proposed in this paper is able to identify broken strips and quantify the number of broken strips accurately under different operating modes of the motor.
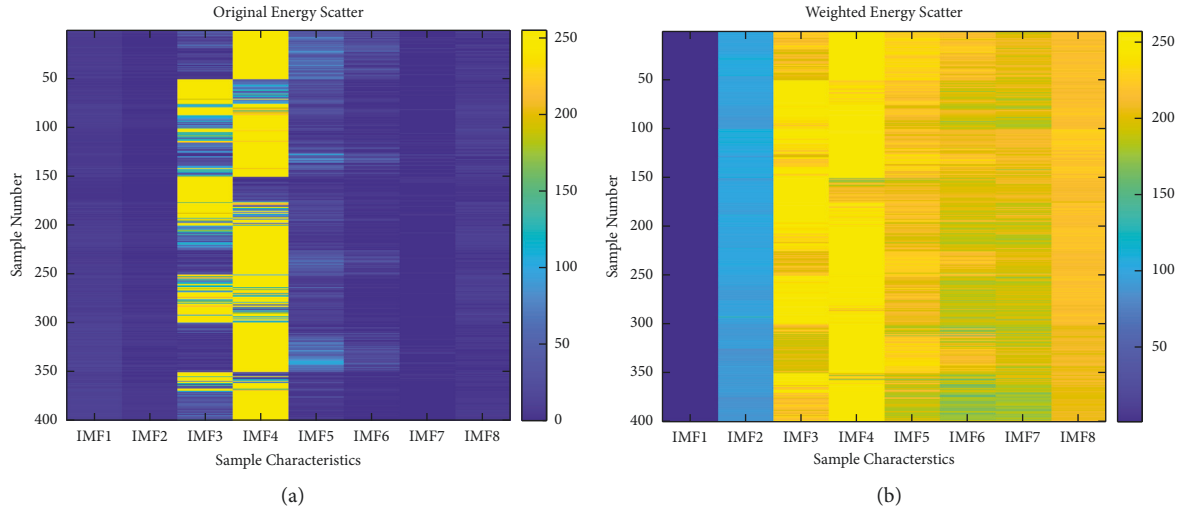
(a)

(b)

FIGURE 7: Energy distribution before and after IMF weighting for variable speed conditions: (a) raw IMF energy distribution and (b) weighted IMF energy distribution.
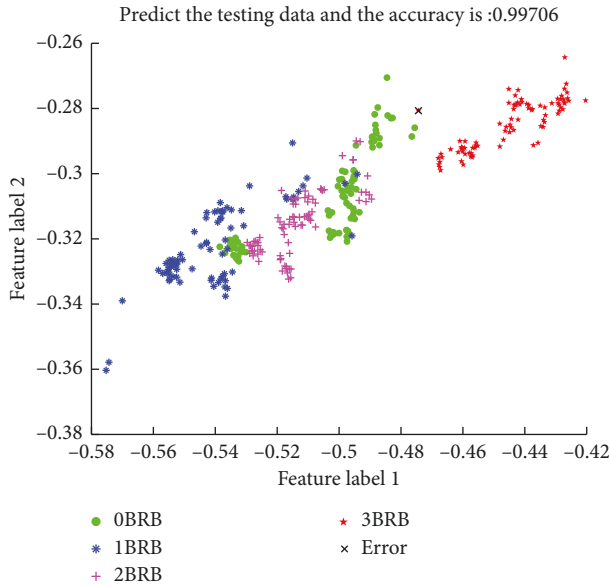

FIGURE 8: Classification effect of AadaBoost broken bar quantitative classifier for variable speed conditions.

7. Summary

This paper proposes a fault feature extraction method that uses CEEMD to decompose the vibration signal, uses the IMF energy factor of each layer as the initial feature, and uses the OOB estimation error to measure the contribution of the feature to the sample classification to construct the feature importance vector and further expands the differentiation between features by logarithmization and weight transformation. The feasibility of using this method and the high discrimination in fault feature extraction is verified based on the experimental data set, and the higher accuracy of this method is demonstrated by comparing it with existing feature extraction methods. Finally, an AdaBoost integrated learning method is used to construct a quantitative classifier

for the number of broken rotor bars to achieve accurate identification of the number of broken rotor bars. The following conclusions can be drawn:

(1) In this paper, the CEEMD decomposition technique is successfully applied to the fault feature extraction of electric motors, solving the previous problems of manually determining the time window and basis function when feature extraction is performed through the time-frequency domain, as well as the problem of residual noise introduced by EEMD during decomposition to destroy the energy distribution.

(2) In this paper, the concept of weight transformation is proposed to measure the importance of eigenvalues in characterizing the rotor state of a motor, and it is applied to the spatial transformation of eigenvectors to improve the discrimination between different eigenvectors.

(3) The quantitative classifier of broken rotor strips constructed by the AdaBoost method in this paper can achieve an accurate quantitative analysis of the number of broken rotor strips, which can be used by engineers to grasp the degree of rotor failure in time and reasonably arrange the inspection and maintenance schedule.

8. Feature Work

The solution proposed in this paper is only capable of quantitative analysis of faults with broken rotor strips. The identification and quantitative analysis of faults such as stator insulation damage, bearing wear, and rotor eccentricity are not yet possible, which will be a further research direction for us. In addition, the current solution is based on the analysis of existing samples and cannot identify faults outside the data set, and it will be a challenging and meaningful task to achieve a monitoring system that grows to learn from unknown faults.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] K. Kudelina, B. Asad, T. Vaimann, A. Rassõlkin, A. Kallaste, and H. Van Khang, "Methods of condition monitoring and fault detection for electrical machines," *Energies*, vol. 14, 2021.

[2] D. Camarena-Martinez, M. Valtierra-Rodriguez, J. P. Amezquita-Sanchez, L. Granados, J. R.-T. René de, and G.-P. Arturo, "Shannon Entropy and *K*-means method for automatic diagnosis of broken rotor bars in induction motors using vibration signals," *Shock and Vibration*, vol. 2016, Article ID 4860309, 10 pages, 2016.

[3] S. K. Gundewar and P. V. Kane, "Condition monitoring and fault diagnosis of induction motor," *Journal of Vibration Engineering*, vol. 9, pp. 1–32, 2020.

[4] L. Frosini, "Novel diagnostic techniques for rotating electrical machines—a review," *Energies*, vol. 13, 2020.

[5] P. Gangsar and R. Tiwari, "Signal based condition monitoring techniques for fault detection and diagnosis of induction motors: a state-of-the-art review," *Mechanical Systems and Signal Processing*, vol. 144, Article ID 106908, 2020.

[6] A. Abid, M. T. Khan, and J. Iqbal, "A review on fault detection and diagnosis techniques: basics and beyond," *Artificial Intelligence Review*, vol. 54, pp. 3639–3664, 2021.

[7] O. Alshorman, F. Alkahatni, M. Masadeh et al., "Sounds and acoustic emission-based early fault diagnosis of induction motor: a review study," *Advances in Mechanical Engineering*, vol. 13, 2021.

[8] G. B. Lucas, B. Albuquerque de Castro, M. A. Rocha, and L. A. André, "A new acoustic emission-based approach for supply disturbances evaluation in three-phase induction motors," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–10, 2021.

[9] Z. M. Wang, H. Li, Z. Dong, F. Gu, and A. D. Ball, *Vibration Signature Analysis for Broken Rotor Bar Diagnosis in Induction Motors Based on Cyclic Modulation Spectrum*, Springer, Berlin, Germany, 2021.

[10] F. Sabbaghian-Bidgoli and J. Poshtan, "Fault detection of broken rotor bar using an improved form of hilbert-huang transform," *Fluctuation and Noise Letters*, vol. 17, Article ID 1850012, 2018.

[11] B. Y. Moon and S. C. Dae, "A study on vibration reduction by heat deformation of casting frame motor," *Journal of Mechanical Science and Technology*, vol. 36, 2022.

[12] D. Liu, H. Zeng, Z. Xiao, L. J. Peng, and O. P. Malik, "Fault diagnosis of rotor using EMD thresholding-based de-noising combined with probabilistic neural network," *Journal of Vibroengineering*, vol. 19, pp. 5920–5931, 2017.

[13] D. Liu, Z. Xiao, X. Hu, C. Zhang, and O. P. Malik, "Feature extraction of rotor fault based on EEMD and curve code," *Measurement*, vol. 135, 2018.

[14] Z. Ke, D. Chong, and X. Bao, "Adaptive suppression of mode mixing in CEEMD based on genetic algorithm for motor bearing fault diagnosis," *IEEE Transactions on Magnetics*, vol. 58, pp. 1–6, 2022.

[15] R. Miceli, Y. Gritli, A. Oscar Di Tommaso, F. Filippetti, and C. Rossi, "Vibration signature analysis for monitoring rotor broken bar in double squirrel cage induction motors based on wavelet analysis," *Compel-the International Journal for Computation and Mathematics in Electrical and Electronic Engineering*, vol. 33, pp. 1625–1641, 2014.

[16] X. Liu, X. Yang, F. Shao, W. Liu, F. Zhou, and C. Hu, "Composite multi-scale basic scale Entropy based on CEEMDAN and its application in hydraulic pump fault diagnosis," *IEEE Access*, vol. 9, Article ID 60564, 2021.

[17] V. Climente-Alarcón, A. A.-D. José, V.-S. Francisco, and P.-P. Ruben, "Vibration transient detection of broken rotor bars by PSH sidebands," *IEEE Transactions on Industry Applications*, vol. 49, pp. 2576–2582, 2013.

[18] D. K. Soother and J. Daudpoto, "A brief review of condition monitoring techniques for the induction motor," *Transactions of the Canadian Society for Mechanical Engineering*, vol. 43, pp. 499–508, 2019.

[19] D. Shang, X. Li, M. Yin, and F. Li, "Vibration suppression method based on PI fuzzy controller containing disturbance observe for dual-flexible manipulator with an axially translating arm," *International Journal of Control, Automation and Systems*, vol. 20, 2022.

[20] B. Akhil Vinayak, K. Anjali Anand, and G. Jagadanand, "Wavelet-based real-time stator fault detection of inverter-fed induction motor," *IET Electric Power Applications*, vol. 14, 2019.

[21] Y. Wu, Z. Zhang, L. Yang, and Q. Sun, "Open-circuit fault diagnosis of six-phase permanent magnet synchronous motor drive system based on empirical mode decomposition energy Entropy," *IEEE Access*, vol. 9, Article ID 91137, 2021.

[22] X. Liu, B. Lin, and H. Luo, "Bearing faults diagnostics based on hybrid LS-SVM and EMD method," *Measurement*, vol. 59, pp. 145–166, 2015.

[23] T. A. Shifat and J.-W. Hur, "EEMD assisted supervised learning for the fault diagnosis of BLDC motor using vibration signal," *Journal of Mechanical Science and Technology*, vol. 34, 2020.

[24] K. Gao, X. Xu, J. Li, S. Jiao, and N. Shi, "Application of multilayer denoising based on ensemble empirical mode decomposition in extraction of fault feature of rotating machinery," *PLoS One*, vol. 16, 2021.

[25] M. A. Colominas, G. Schlotthauer, and M. E. Torres, "Improved complete ensemble EMD: a suitable tool for biomedical signal processing," *Biomedical Signal Processing and Control*, vol. 14, pp. 19–29, 2014.

[26] D. K. Soother, J. Daudpoto, N. R. Harris et al., "The importance of feature processing in deep-learning-based condition monitoring of motors," *Mathematical Problems in Engineering*, vol. 2021, Article ID 9927151, 23 pages, 2021.

[27] N. Lu, Z. Xiao, and O. P. Malik, "Feature extraction using adaptive multiwavelets and synthetic detection index for rotor fault diagnosis of rotating machinery," *Mechanical Systems and Signal Processing*, vol. 52, pp. 393–415, 2015.

[28] K. U. Mahadeo and R. Dhanalakshmi, "Effects of random forest parameters in the selection of biomarkers," *Computer Journal*, vol. 64, pp. 1840–1847, 2021.

[29] Y. Zhou, S. Tian, J. Chen, Y. Liu, and C. Li, "Research on classification of open-pit mineral exploiting information based on OOB RFE feature optimization," *Sensors*, vol. 22, 2022.

[30] Z. Khan, N. Gul, N. Faiz, A. Gul, W. Adler, and B. Lausen, "Optimal trees selection for classification via out-of-bag

assessment and sub-bagging," *IEEE Access*, vol. 9, Article ID 28591, 2021.

[31] X. Jiang, Y. Xu, W. Ke, Y. Zhang, Q. Zhu, and Y. He, "An imbalanced multifault diagnosis method based on bias weights AdaBoost," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–8, 2022.

[32] Y. Xu, C. Zhang, Q. Zhu, and Y. He, "A novel local selective ensemble-based AdaBoost method for fault detection of industrial process," in *Proceedings of the IEEE 9th Data Driven Control and Learning Systems Conference (DDCLS)*, pp. 1388–1393, Liuzhou, China, November 2020.

Retraction

Retracted: Experimental Study on Hydraulic Pulsation Features of Intelligent Variable Valve System for Auto Energy Saving

Computational Intelligence and Neuroscience

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Y. Xu, J. Chen, and Z. Liu, "Experimental Study on Hydraulic Pulsation Features of Intelligent Variable Valve System for Auto Energy Saving," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 5679520, 9 pages, 2022.

Research Article

Experimental Study on Hydraulic Pulsation Features of Intelligent Variable Valve System for Auto Energy Saving

Yuliang Xu,[1] Jiadui Chen [iD],[2] and Zhenghong Liu[1]

[1]*School of Mechanical Engineering, Guiyang University, Guiyang 550005, China*
[2]*Key Laboratory of Advanced Manufacturing Technology and Ministry of Education, Guizhou University, Guiyang 550025, China*

Correspondence should be addressed to Jiadui Chen; jdchen1@gzu.edu.cn

The electrohydraulic valve system can realize continuous variable valve timing and lift using the flexibility of liquid. However, the existing electrohydraulic valve systems usually rely on a high-speed solenoid valve to control the on and off states of the hydraulic circuit, which pushes up the system cost. This paper introduces a continuous hydraulic variable valve timing and lift system with adjustable volume (CVVTL). Unlike the other electrohydraulic valve systems, the CVVTL does not need a high-frequency servo valve to control each valve but adjusts the valve timing and lift by controlling the system volume. However, the maximum operating speed of hydraulic variable valve systems is limited by the inherent pressure fluctuation. To relax the limit, the authors further studied the pressure fluctuation features of the CVVTL system under various conditions and summarized the harms of the fluctuation. After identifying the causes of pressure fluctuation of the CVVTL system, the authors came up with systematic countermeasures to system pressure fluctuation.

1. Introduction

Under the premise of ensuring the automobile power performance, improving the engine work efficiency is an effective means to reduce energy consumption and improve the greenhouse effect. The automotive industry applies several technological solutions to improve the engine work efficiency, such as direct fuel injection, engine downsizing, exhaust gas recirculation (EGR), and variable valve trains (VVT) [1–5]. Among them, the VVT can greatly boost the engine performance by adjusting the timing, lift, and duration of the valve according to the engine working conditions. The variable valve actuation can be achieved using mechanical, electromagnetic, and electrohydraulic valve mechanisms [6–11].

Now, many mechanical valve mechanisms are being used in engines, namely, Honda's Variable Valve Timing and Lift Electronic Control (VTEC), Mitsubishi Innovative Valve timing Electronic Control system (MIVEC), Toyota's Variable Valve Timing with intelligence (VVT-i), and Porsche's VarioCam [12, 13]. Nevertheless, the discontinuous change of valve timing offered by the above mechanical valve mechanisms does not support continuous adjustment of valve timing. For example, in VTEC technology, there are only two or three types of distribution cam profiles, which can only meet the ideal distribution requirements of two or three speed conditions of the engine. It is impossible for an engine with such a mechanical valve mechanism to gain reasonable gas distribution under every working condition.

Electrohydraulic and electromagnetic valve mechanisms can achieve continuous valve timing, lift, and duration at all operating speeds and loads. But the electromagnetic valve systems face several disadvantages: large impact of valve seating, huge size of the electromagnet, serious emission of heat, frequent electromagnetic responses, high energy consumption, as well as complex and expensive system [14–17]. For electrohydraulic valve systems, a high-frequency servo valve is often needed to control the value, such as Fiat's Multiair Valve-Lift System [18] and Lotus's electrohydraulic fully variable valve train (EHFVVT) [19, 20].

To solve the defects of the existing VVT technologies, this paper proposes a continuous hydraulic variable valve timing and lift system with adjustable volume (CVVTL). In the system, the advance angle, retard angle, and duration angle of the valve are adjusted continuously by controlling the time that the oil flows into and out from the valve cylinder. The valve lift is adjusted continuously by controlling the liquid volume flowing into the valve cylinder. The CVVTL system eliminates the need for a high-frequency servo valve, which is required by other electrohydraulic valve systems to control each valve. Hence, the system can work at a higher engine speed than the other electrohydraulic valve systems. However, it was found that the oil pressure of the CVVTL system fluctuates during the high-speed operation, which limits the maximum operating speed of the system. Therefore, the authors further examined the phenomenon, causes, harms, and countermeasures of pressure fluctuation in the CVVTL system. Except Xie et al. [21], virtually no scholar has reported the pressure fluctuations of hydraulic variable valve systems.

2. Structure and Working Principle of the CVVTL

2.1. Structure of the CVVTL. As shown in Figure 1, the CVVTL is composed of a cam, a cam oil cylinder, a valve oil cylinder, a valve assembly, a phase regulator, a lift regulator, a seating buffer, an oil supply system, etc. Among them, the phase regulator and lift regulator each consists of an oil cylinder, a piston, a spring, a gag lever post, and an adjustment device. The oil supply system involves an oil tank, an oil pump, a check valve, a relief valve, and a pipeline.

2.2. Working Principle of the CVVTL. The working principle of CVVTL is as follows [22].

2.2.1. Valve Timing Adjustment. Under the effect of the spring, the phase regulator piston lies at the very front of the phase regulator cylinder, while the cam works in the base circle segment for the gag lever post of the phase regulator at any position in the adjustment range. To adjust the valve timing, it is only necessary to change the position of the gag lever post.

When the cam lift is coming, the cam cylinder begins to pump oil. Since the phase regulator spring has a smaller pretightening force than the valve spring, the oil pumped from the cam cylinder would firstly enter the phase regulator cylinder, until the regulator piston is stopped by the gag lever post. At this stage, the valve keeps still. Therefore, the valve is opened after the adjustment of the gag lever post equals zero, that is, the valve advance angle is decreased. As the cam continues rotating, the oil pressure gradually increases in the system. Once the valve cylinder can overcome the pretightening force of the valve spring, the valve is opened, and the valve opening is gradually increased until the end of cam lift.

The above working process is reversed when the cam works in its fall cure. Since the valve spring has a greater
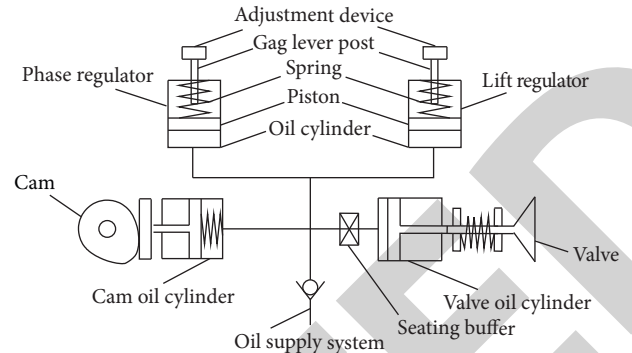

Figure 1: System structure.

force than the phase regulator spring, the oil of the valve cylinder would firstly return to the cam cylinder, until the valve is seated. After that, the cam continues rotating, and the phase regulator spring drives the oil in the phase regulator to return to the cam cylinder, until the cam returns to the base circle segment. Therefore, the valve is closed before the adjustment of the gag lever post equals zero, that is, the valve retard angle is decreased.

Through the above procedure, the system continuously adjusts the advance angle, retard angle, and duration angle of the valve by controlling the time that the oil flows into and out of the valve cylinder.

2.2.2. Valve Lift Adjustment. Under the effect of the spring, the lift regulator piston lies at the very front of the lift regulator cylinder, while the cam works in the base circle segment for the gag lever post of the lift regulator at any position in the adjustment range. To adjust the valve lift, it is only necessary to change the position of the gag lever post.

When cam works in its lift, and the system oil pressure is large enough to overcome the pretightening force of the valve spring, i.e., to open the valve, the lift regulator piston will enter the idle state, for the lift regulator spring has a larger pretightening force than that of the valve spring. In this case, the oil pumped from the cam cylinder will only flow into the valve cylinder to open the valve. Once the system pressure is large enough to overcome the pretightening force of the lift regulator spring, the oil pressure will move the lift regulator piston, and the oil will start to flow into the lift regulator oil cylinder, until the piston is stopped by the gag lever post of the lift regulator. Therefore, the liquid volume entering the valve cylinder will decrease and so will the valve lift.

The above working process is reversed when the cam works in its fall cure. Since the lift regulator works in the opening of the valve, the lift adjustment has no impact on valve timing. It is apparent that the system can continuously adjust the valve lift by controlling the liquid volume entering the valve cylinder and lift regulator. The volume is controlled by properly adjusting the position of the gag lever post of the lift regulator.

Table 1 shows the optimal intake valve opening angle (IVO), intake valve closing angle (IVC), and valve lift

Table 1: Optimal gas distribution parameters at different engine speeds.

Engine speed (r/min)	IVO (deg)	IVC (deg)	Valve lift (mm)
1500	82	302	7.24
2500	83	310	7.71
3500	79	318	7.81
4500	73	335	8.54
5500	77	348	8.76

parameters of an in-line four-cylinder gasoline engine at different engine speeds. It can be seen that the optimal parameters vary at different engine speeds.

For the CVVTL, the simultaneous adjustment of the valve timing and lift can be achieved, as the system adjusts the phase regulator piston and lift regulator piston to their reasonable positions, according to the working conditions of the engine. Therefore, the CVVTL will be more intelligent because it can meet the gas distribution requirements of the engine in more conditions.

2.3. CVVTL Prototype.

This paper develops a four-cylinder CVVTL prototype of the engine cylinder head (Figure 2). In the prototype, the timing and lift of the intake valves are adjusted by the CVVTL. For the exhaust valves, only the timing is adjustable. There are one cam cylinder, one phase regulator, and one lift regulator for the intake valves of a cylinder and one cam cylinder and one phase regulator for the exhaust valves of a cylinder.

All intake cam cylinder pistons, which are driven by the intake cam, are laid with a spacing of 90°. All exhaust cam cylinder pistons, which are driven by the exhaust cam, are laid with the same spacing. The intake cam and exhaust cam are mounted on the same camshaft. The angle between the two cams meets the requirement of engine gas distribution. The intake cam cylinders, exhaust cam cylinders, camshaft, intake phase regulators, intake lift regulators, as well as exhaust phase regulators are integrated in the driving assembly.

The intake phase regulators and exhaust phase regulators are controlled by the phase adjusting device, while the intake lift regulators are controlled by the lift adjusting device. The intake and exhaust valve cylinders are integrated in the intake valve cylinder assembly and the exhaust valve cylinder assembly, respectively. The intake valve cylinder assembly and the exhaust valve cylinder assembly are connected with the oil passage assembly via the driving assembly.

The camshaft is driven by the servo motor with the synchronous belt, with a transmission ratio of 1 : 1. The valve motion was measured by the laser displacement sensor. During the experiments, both the phase adjusting device and lift adjusting device were adjusted manually.

3. Fluctuation Features and Harms

3.1. Fluctuation Features

3.1.1. Different Engine Speeds.

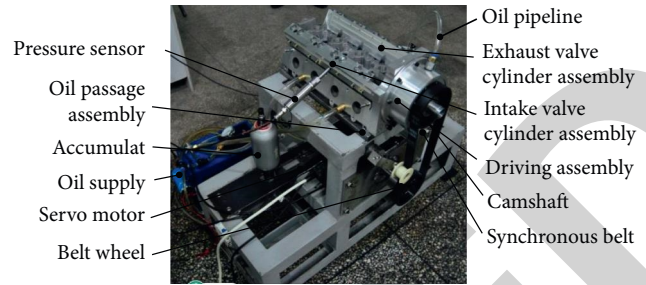Figure 3 shows the system pressure curves at different engine speeds, when neither

Figure 2: CVVTL prototype.

valve timing nor valve lift is adjusted. When the engine speed was less than 1,716 r/min, the system pressure almost had no fluctuation. When the engine grew continuously beyond that speed, the maximum pressure gradually increased, while the minimum pressure gradually decreased. The faster the engine speed, the more obvious the system pressure fluctuated.

With the rise of engine speed, the crank angle that the system arrives at its minimum pressure became smaller. When the engine moved at 5,150 r/min, the minimum pressure was smaller than the initial pressure, and the second smallest pressure was very close to the initial pressure. It could be predicted that, with further growth of engine speed, the crank angle and the minimum pressure would continue to decrease, and the duration angle and the number of troughs of the system pressure curve would increase.

Once the system pressure fell below the initial pressure, the valve would be out of control. Then, the oil supply subsystem would automatically replenish oil to the system, causing the hydraulic pressure to rise in the system. In this case, the CVVTL would lose its gas distribution capability. Therefore, the CVVTL should work in the engine speed range, which ensures that the system pressure is always greater than the initial pressure. To support the CVVTL operations at a high engine speed, it is necessary to effectively control and mitigate the fluctuation of system pressure.

3.1.2. Different Adjustment Quantities of the Phase Regulator.

Figure 4 shows the system pressure curves at three different adjustment quantities of the phase regulator (1 mm, 2 mm, and 3 mm) and the engine speed of 4,006 r/min. As can be seen from the figure, as the adjustment quantity of the phase regulator increased, the maximum system pressure dropped, yet the system pressure fluctuated more significantly. When the adjustment quantity of the phase regulator was greater than 3 mm, the system pressure fell below the initial pressure before the valve opened, and the valve timing could not be effectively adjusted by the CVVTL. Therefore, the adjustment quantity of the phase regulator should be less than 3 mm at 4,006 r/min.

3.1.3. Different Adjustment Quantities of the Lift Regulator.

Figure 5 shows the system pressure curves at three different adjustment quantities of the lift regulator (1 mm, 2 mm, and 3 mm) and the engine speed of 4,006 r/min. As can be seen
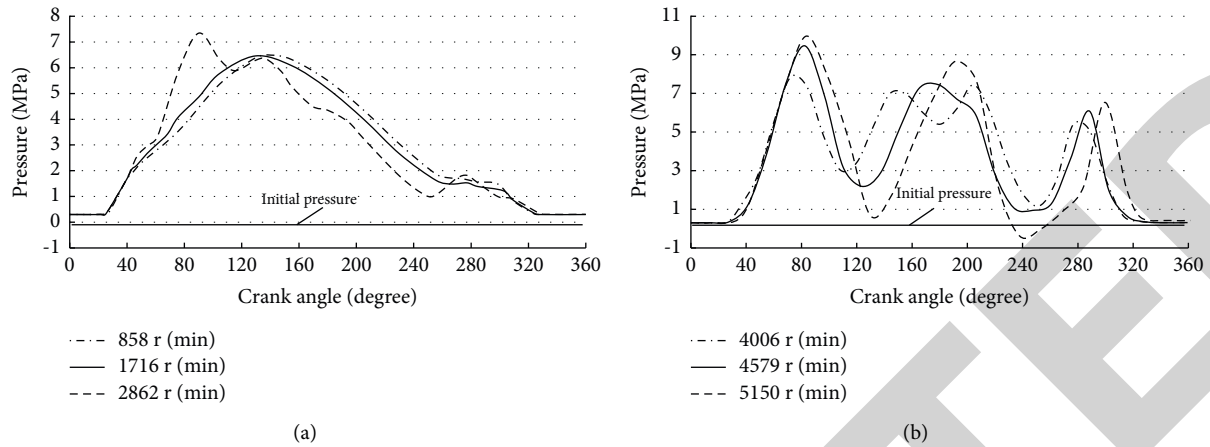
(a)

(b)

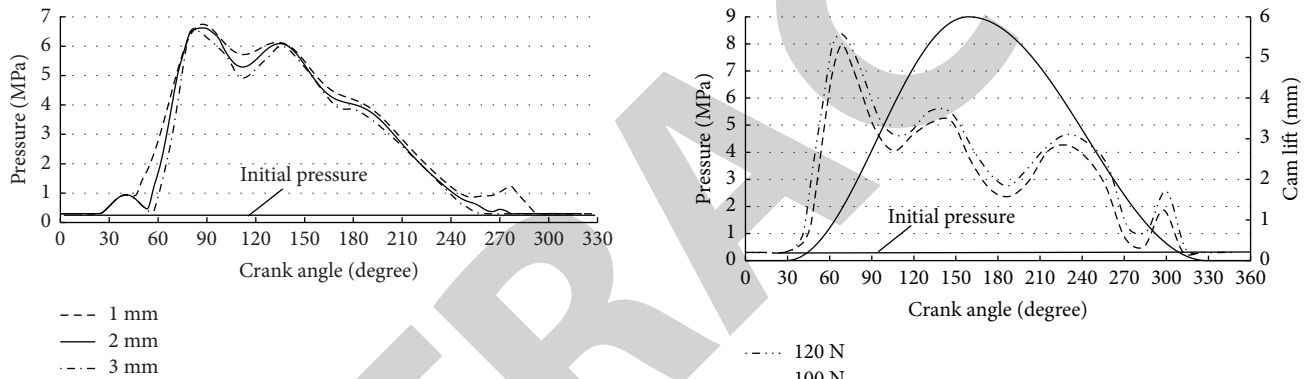Figure 3: System pressure curves at different engine speeds.

Figure 4: System pressure curves of different adjustment quantities of the phase regulator.

Figure 5: System pressure curves of different adjustment quantities of the lift regulator.

from the figure, as the adjustment quantity of the lift regulator increased, the maximum system pressure dropped and appeared later, the system pressure fluctuated more significantly, and the minimum system pressure declined. If the adjustment quantity of the lift regulator continues to rise, the system pressure would fall below the initial pressure.

Figure 6: System pressure curves at different pretightening forces of the valve spring.

3.1.4. Different Pretightening Forces of the Valve Spring. Figure 6 shows the system pressure curves at different pretightening forces of the valve spring at the engine speed of 4,006 r/min. With the growing pretightening force of the valve spring, the system pressure increased, and the trough point of system pressure (i.e., the position of minimum system pressure) moved away from the initial pressure. Hence, the minimum system pressure could be improved by reducing the pretightening force of the valve spring. If that force is too large, more drive energy would get lost, resulting in a greater stress and a more severe wear to hydraulic components.

3.1.5. Different Masses of Valve Components. Figure 7 shows the system pressure curves at different masses of the valve components at the engine speed of 4,006 r/min. With the growing mass of the valve components, the maximum system pressure was on the rise, and the system pressure fluctuated more apparently. Using light valve components can improve the system pressure curve and allow the CVVTL to work at a faster speed.
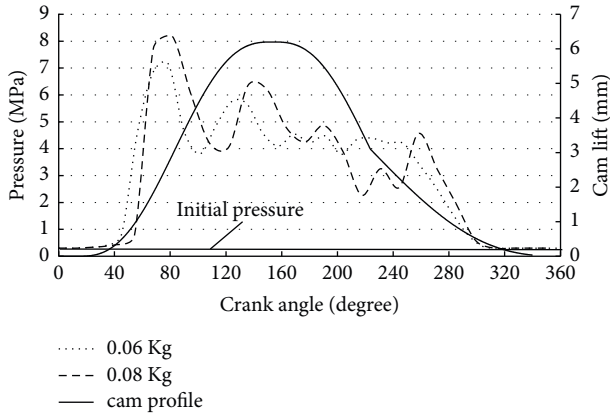
Figure 7: System pressure curves at different masses valve component.

3.2. Fluctuation Harms. The foregoing analysis shows that, when the CVVTL operates faster than the maximum allowable rotational speed, the pressure fluctuation will cause the system pressure to fall below the initial pressure. As the engine speed continues to increase, the system pressure will oscillate more prominently, the valve at the trough point of system pressure will decrease, and the system will eventually enter the negative pressure state. The pressure fluctuation will bring the following harms to the system:

(1) Once the system pressure falls below the initial pressure, the valve movement will be out of control, making it hard to adjust the intake air volume of the engine. In this case, the CVVTL cannot meet the engine gas demand.

(2) When the system pressure is less than the initial pressure, the oil supply subsystem will automatically replenish the oil to the system, pushing up the system pressure. If too much oil is replenished, the system pressure will exceed the reasonable design range, as the cam returns to the base circle, and the hydraulic force acting on the valve cylinder piston will surpass the valve spring preload. If this occurs, the valve will not be seated, and the engine will not work properly.

(3) When the system enters the negative pressure state, the air dissolved in the hydraulic oil will precipitate in the form of bubbles. The ensuing cavitation will
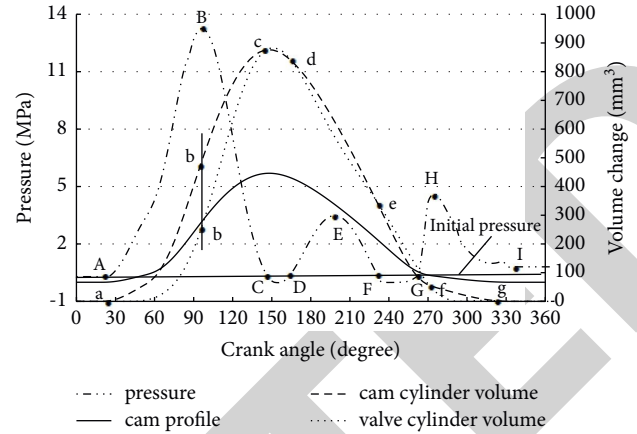

Figure 8: System pressure curve and volume change curves at the engine speed of 5,500 r/min.

compromise the CVVTL. What is worse is that fatigue peeling will take place on the metal parts. Repeated negative pressures will seriously damage the system. In addition, the volume elastic modulus of the oil will nosedive, further weakening the system stiffness. Then, the system vibration will be exacerbated during the high-speed operation.

4. Fluctuation Causes

The CVVTL adjusts the valve timing and lift by changing the total volume of the system. Without considering the risk of leakage, the system pressure should be constant and equal to the initial pressure, when the total volume of the system remains constant, that is, the volume change of the cam cylinder equals the total volume change of the valve cylinder, the phase regulator cylinder, and the lift regulator cylinder. If the volume change of the cam cylinder is greater than the said total volume change, the total system volume would shrink, and the oil would be compressed. Then, the system pressure would rise gradually, surpassing the initial pressure. If the volume change of the cam cylinder is smaller than the said total volume change, the total system volume would expand, and the oil volume would grow. Then, the system pressure would fall below the initial pressure. Mathematically, the above change trends can be expressed as

$$\begin{cases} V_c = V_v + V_p + V_l, & p = p_0, \\ V_c > V_v + V_p + V_l, & p > p_0, \\ V_c < V_v + V_p + V_l, & p < p_0, \end{cases} \tag{1}$$

where V_c, V_v, V_p, and V_l are the volume changes of the cam cylinder, the valve cylinder, the phase regulator cylinder, and the lift regulator cylinder, respectively; p is the system pressure; p_0 is the initial pressure.

Figure 8 shows the system pressure curve and volume change curves at the engine speed of 5,500 r/min. Note that

Curve 1 is about system pressure; Curve 2 is about the volume change of the cam cylinder; Curve 3 is about the total volume change of the valve cylinder, the phase regulator cylinder, and the lift regulator cylinder; Curve 4 is the cam profile.

In the AB segment, Curve 2 was above Curve 3, and the gap gradually increased. The largest gap between them appeared at point B. This means, in the AB segment, the volume reduction of the cam cylinder is greater than the total volume increment of the valve cylinder, the phase regulator cylinder, and the lift regulator cylinder. Hence, the total system volume gradually decreases, and the oil is compressed. Then, the system pressure would surpass the initial pressure, and gradually approach the maximum pressure, corresponding to the AB segment of Curve 1.

In the BC segment, Curve 2 was still above Curve 3; yet, the gap gradually narrowed. The volume reduction of the cam cylinder equals the said total volume increment at point C. This means, the oil volume compressed in the AB segment gradually expands and returns to the initial volume at point C. Therefore, the system pressure would gradually fall from the maximum pressure at point B to the initial pressure, corresponding to the BC segment of Curve 1.

In the CD segment, Curve 2 was below Curve 3, and the gap gradually widened before shrinking. That is, the volume reduction of the cam cylinder is smaller than the said total volume increment. Hence, the system pressure is less than the initial pressure and would gradually fall and then increase to the initial pressure, corresponding to the CD segment of Curve 1. Then, the system would be replenished by the oil supply subsystem.

In the DE segment, Curve 2 was above Curve 3, namely, the volume reduction of the cam cylinder is greater than the said total volume increment. Note that the gap first gradually increases and then narrows slowly. In this case, the oil would be compressed again and then expanded. Therefore, the system pressure would gradually increase from the initial pressure to the peak, before falling from the peak to the initial pressure. This process corresponds to the DEF segment of Curve 1.

The case of EF segment is the same as that of the CD segment: the system pressure is below the initial pressure and would decrease slowly before a gradual increase, corresponding to the FG segment of Curve 1. Then, the system would be replenished by the oil supply subsystem again.

The case of FG segment is similar to that of the DE segment. Since much oil has been replenished to the hydraulic system, the system pressure would pick up sharply before a gradual decline. In a working cycle, two replenishments happen to the system. Therefore, the system pressure would be larger than the initial pressure when the cam returns to the base circle segment.

The above analysis reveals an important cause of oil pressure fluctuation: the volume reduction of the cam cylinder is not equal to the total volume increment of the valve cylinder, the phase regulator cylinder, and the lift regulator cylinder.

5. Countermeasures

With the increase of engine speed, the hydraulic system pressure of the CVVTL will oscillate rather violently, producing a huge pressure shock. Then, the CVVTL operation will be seriously limited, as the engine operates at a fast speed. The system pressure fluctuation can be controlled in two aspects: firstly, prevent excessively large system pressure, or the valve may fly out. Secondly, prevent excessively low system pressure (keep the system not too far below the initial pressure) in the working segment of the cam. Once the system pressure falls below the initial pressure, the oil supply subsystem would start oil replenishment, which affects the next cycle of the engine. If the system enters the negative pressure state, the valve would go out of control, and too much oil would be added to the hydraulic system, making it impossible for the CVVTL to work properly. The system pressure fluctuation must be effectively controlled to ensure the adaptability and reliability of the CVVTL.

5.1. Improve the Natural Frequency of the System. The system vibration causes the system oil pressure to vary. This vibration can be reduced by increasing the natural frequency of the system. Thus, this paper suggests curbing the pressure fluctuation of the CVVTL by improving the system natural frequency. The natural frequency of the CVVTL can be expressed as

$$f = \sqrt{\frac{k}{m}}, \tag{2}$$

where f is the natural frequency of the CVVTL, k is the system stiffness, and m is the total mass of moving components.

Formula (2) shows that the natural frequency of the CVVTL can be improved by increasing the system stiffness and reducing the total mass of moving components. By the definition of hydraulic system stiffness, the stiffness of the CVVTL can be expressed as

$$
\begin{aligned}
k &= \frac{F}{\delta} \\
&= \frac{\Delta p \cdot A}{\Delta V/A} \\
&= \frac{\Delta p}{\Delta V} \cdot A^2,
\end{aligned}
\tag{3}
$$

where F is the load of the cam piston, δ is the displacement of the cam piston under load, Δp is the increment of system pressure, ΔV is the volume change of hydraulic oil, and A is the cross-sectional area of the cam piston.

The elastic modulus of the hydraulic oil E can be expressed as

$$E = \frac{\Delta p}{\Delta V/V}, \tag{4}$$

where V is the total volume of the hydraulic oil in the system.

Formula (4) can be rewritten as

$$\frac{\Delta p}{\Delta V} = \frac{E}{V}. \tag{5}$$

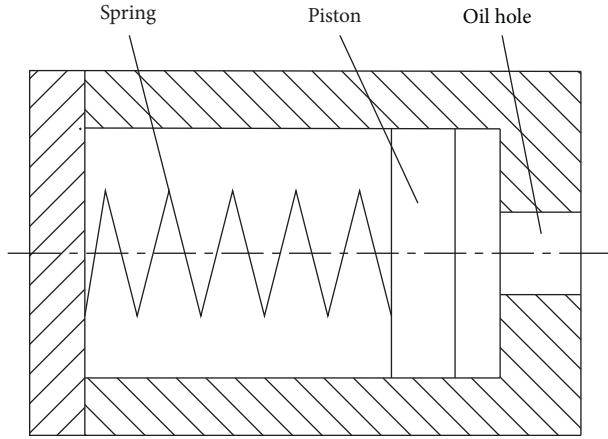Substituting formula (5) into formula (3), the system stiffness k can be expressed as

FIGURE 9: Structure of the buffer accumulator.

FIGURE 10: System pressure curves and valve lift curves at the engine speed of 5,500 r/min.

$$k = \frac{EA^2}{V}. \tag{6}$$

Formula (6) shows that, the system stiffness is proportional to the square of the cross-sectional area of the cam piston and the volume elastic modulus of the hydraulic oil in the system while inversely proportional to the total volume of that oil. Therefore, the system stiffness can be improved by expanding the cross-sectional area of the cam piston, reducing the total volume of the hydraulic oil in the system, and using a hydraulic oil with a large volume elastic modulus.

Without considering valve timing adjustment and lift adjustment, the relationship between the valve maximum lift and cam maximum lift can be expressed as

$$\frac{H_{cm}}{H_{vm}} = \sqrt{\frac{A}{A_V}}, \tag{7}$$

where H_{cm} is the cam maximum lift, H_{vm} is the valve maximum lift, and A_v is the cross-sectional area of the valve cylinder piston.

When the valve maximum lift and the cam maximum lift are satisfied, the system stiffness can be improved by using the piston with the largest possible cross-sectional area.

In the specific design, the total volume of the hydraulic oil in the system can be reduced by selecting the size of the oil chamber reasonably, filling the process holes formed in the processing and manufacturing processes, and properly decreasing the length of the oil line.

The volume elastic modulus of the hydraulic oil depends on the temperature and the air content of the oil. It is negatively correlated with the amount of air mixed in it and the temperature level. To reduce the air content, the oil should be directed into the oil tank slowly, such as to prevent the forming of spray or foam in the tank. To control the oil temperature, the hydraulic oil of the CVVTL should be replaced after each working cycle, for the replacement can take away the heat generated in the work of the oil. In this way, it is possible to rationalize the volume elastic modulus of the hydraulic oil.
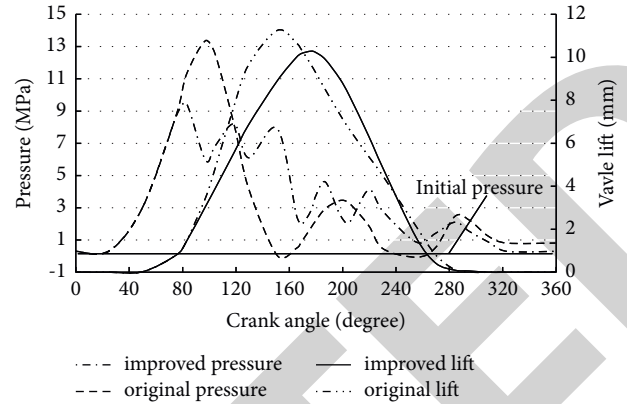
5.2. Parallel a Buffer Accumulator with the System. When the engine runs at a high speed, the system pressure remains high and fluctuates significantly. To ensure the normal operation of the system, it is a must to control the peak pressure and pressure fluctuation of the system. Paralleling a buffer accumulator with the system can both limit the maximum pressure of the system and improve the minimum operating pressure. The structure of the buffer accumulator is shown in Figure 9. The working principle of the buffer accumulator is as follows:

For the buffer accumulator, the pretightening force and stiffness of the spring are selected appropriately based on the needs of the system. The pretightening force must be greater than that of the valve spring so that the valve can open normally. Meanwhile, the stiffness must be smaller than that of the valve spring. When the hydraulic pressure on the buffer accumulator plunger is greater than the pretightening force of the buffer accumulator spring, the piston of the buffer accumulator would move, and the oil would flow into the buffer accumulator rather than into the valve cylinder. In this way, the system pressure would not increase very significantly, putting the maximum pressure under control. When the hydraulic pressure on the buffer accumulator plunger is smaller than the pretightening force of the buffer accumulator spring, the piston of the buffer accumulator would move to the initial position under the action of the spring force. In this case, the oil would flow back into the system from the buffer accumulator. The falling of system pressure would be slowed down, making the system pressure more stable.

Figure 10 shows the system pressure curves and valve lift curves at the engine speed of 5,500 r/min. Note that Curve 1 is the valve lift curve of the CVVTL without the parallelization of the buffer accumulator; Curve 2 is the valve lift curve of the CVVTL paralleled with a buffer accumulator; Curve 3 is the system pressure curve of the CVVTL without the parallelization of the buffer accumulator; Curve 4 is the system pressure curve of the CVVTL paralleled with a buffer accumulator.

As shown in Figure 10, Curve 3 had a greater maximum hydraulic pressure, fewer fluctuation numbers, and a larger
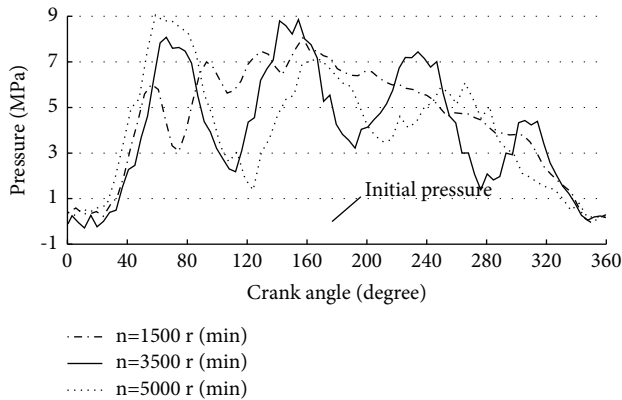
FIGURE 11: System pressure curves at different engine speeds.

amplitude than Curve 4. In Curve 3, two troughs were less than the initial pressure, and the valve lift peaked at 11.25. In Curve 4, the maximum valve lift was 10.28, smaller than that in Curve 3. These results demonstrate that the CVVTL can effectively control the pressure fluctuation and the maximum valve lift and improve the maximum allowable working speed of the CVVTL after the system is paralleled with a buffer accumulator.

5.3. Choose Reasonable Design Parameters. The design parameters directly bear on the CVVTL performance, including but not limited to the mass of valve components, the pretightening force and stiffness of the valve spring, the parameters of the lift regulator, the parameters of the phase regulator, and the length and diameter of the hydraulic pipe. Choosing the reasonable design parameters would significantly improve the system pressure fluctuation.

5.4. Design Reasonable Cam Profile. Cam, the power source of the hydraulic system, drives the system by the cam cylinder piston. The motion rule of the cam profile affects the fluctuation of the system pressure and the movement law of the valve. Therefore, the system pressure fluctuation can be improved by designing the cam profile more rationally. Possible design measures include selecting a cam profile with a continuous acceleration curve (e.g., a high-power cam profile), reducing the maximum acceleration of the cam profile, and widening the positive speed cam profile. In considering the system working characteristics of the system, the design method of cam profile should not only meet the basic requirements of inflation efficiency, contact stress, and smooth operation, etc. but it can also meet the special design requirements of valve adjustment matching and oil compression compensation.

According to the above analysis results, the mechanism parameters of the system are optimized. The recaptured system pressure fluctuation test curves are shown in Figure 11. Compared with Figure 3, the fluctuation range of oil pressure is significantly reduced. When the engine speed reaches 5500 r/min, the fluctuation range is 1.3~8.9 MPa, both higher than the initial oil pressure of the system.

6. Conclusions

In the proposed CVVTL, the system pressure fluctuates more violently with the increase of the engine speed. The minimum (trough value) of system pressure equals the initial pressure when the engine speed reaches the maximum allowable value. Further increase in engine speed would throw the system into the negative pressure state. Then, the CVVTL will lose its distribution capability. The system pressure fluctuation is affected by multiple factors such as the adjustment quantity of the phase regulator, the adjustment quantity of the lift regulator, the pretightening force and stiffness of the valve spring, and the mass of the valve components. It limits the maximum operating speed of the CVVTL.

The fluctuation of the oil pressure is primarily caused by the fact that the volume reduction of the cam cylinder is not equal to the total volume increment of the valve cylinder, the phase regulator cylinder, and the lift regulator cylinder.

The fluctuation of system pressure can be effectively controlled by improving the natural frequency of the system, paralleling a buffer accumulator with the system, choosing reasonable design parameters, and designing reasonable cam profile.

The fluctuation range of oil pressure is significantly reduced to 1.3~8.9 MPa, which increases the applicable range of CVVTL to 5,500 r/min.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this study.

Acknowledgments

References

[1] M. Haas and M. Rauch, "Electro-hydraulic fully variable valve train system," *MTZ worldwide*, vol. 71, no. 3, pp. 16–21, 2010.

[2] T. V. Johnson, "Vehicular emissions in review," *SAE International Journal of Engines*, vol. 5, no. 2, pp. 216–234, 2012.

[3] K. R. Vos, G. M. Shaver, M. C. Joshi, and J. McCarthy, "Implementing variable valve actuation on a diesel engine at high-speed idle operation for improved aftertreatment warm-up," *International Journal of Engine Research*, vol. 21, no. 7, pp. 1134–1146, 2020.

[4] P. Zou, J. Liu, X. Zhou, Z. Chen, and B. D. X. J. Luo, "Effect of a novel mechanical CVVL system on economic performance