Research Article

Big Data Analysis Technology for Artificial Intelligence Decision-Making Platform Construction and Application

Liangfang Liu\(^1\) and Zhiyi Hu\(^2\)

\(^1\)School of Information Engineering, Zhongshan Polytechnic, Zhongshan, Guangdong 528400, China
\(^2\)University of Electronic Science and Technology of China Zhongshan Institute, Zhongshan, Guangdong 528400, China

Correspondence should be addressed to Liangfang Liu; 2009013@zsc.edu.cn

Received 11 July 2022; Revised 27 July 2022; Accepted 2 August 2022; Published 31 August 2022

Academic Editor: R. Mo

Copyright © 2022 Liangfang Liu and Zhiyi Hu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the development of the information age, it is the opportunity and challenge for enterprises to apply big data analysis technology to make decisions and better solve the major problems of global and sustainable development. Decision-making is crucial for enterprises, and a correct decision can improve the development potential and competitiveness of enterprises. However, traditional decision models have certain limitations, and it is difficult to handle the massive, polymorphic, and changing decision data. To address these problems, we propose a combination of temporal data anomaly detection and width learning big data analysis technology for building an intelligent decision platform to assist enterprises to better solve major decision problems.

First, the goal of time series data anomaly detection is to correctly determine whether the data points in each moment of the time series are abnormal. The variation of time series data is affected by various factors, and the fluctuation of data caused by some nonanomalous factors can increase the difficulty of anomaly detection. To address the above problems, we propose an anomaly detection algorithm for time series data based on time series decomposition method. In this algorithm, the time series are decomposed by STL method and HP filtering method according to whether the time series is periodic or not, and then the components of the time series that are relevant to anomaly detection are retained and anomaly detection is performed on the processed time series using a cyclic model. Then, based on the call text and signaling data in enterprise decision-making, an improved width learning model called coding width learning is proposed. The coding width learning model is used to identify decision problems and make comprehensive decisions to improve the model training time and the accuracy of identification. At the same time, an integrated learning method with parallelized training is proposed for width learning in order to further improve the efficiency of coding width learning and prevent the potential memory explosion problem. Finally, the experimental results show that the proposed anomaly detection method effectively improves the anomaly detection performance of the model, and its performance is better than the existing time series anomaly detection algorithm based on variation self-encoder; combined with the improved width learning model, it can make fast decisions and analysis without using scenarios and targets, and help and guide the work of related personnel.

1. Introduction

The current rapid development of big data technology has added new impetus to the development of various industries, but also makes the enterprise management decision-making environment change dramatically, which has a greater impact on decision-making data and decision-making participants [1]. At the same time, the difficulty of identifying decision-making information has increased, and the traditional decision-making methods are in urgent need of innovation, which can easily affect the normal operation of enterprises once mishandled. In this regard, the leadership should establish data awareness, apply big data to management decision-making, ensure that the human resources management approach is more comprehensive and diversified, can fully respond to environmental changes, and innovate data mining and analysis technology to collect more information related to management decision-making,
relying on many new ways to avoid and reduce decision-making risks.

The meaning of big data is divided into broad and narrow sense. Broad sense represents the collection of the data, which not only can analyze and process the data quickly and accurately, but also has important implications for decision-making; narrow sense represents the combination of massive information with each other. Usually, big data has the following characteristics. First, it has large capacity, there are several storage units available in the data integration, the minimum is TB, and then in this unit there are PB and EB; second, processing efficiency is fast. Compared with the previous processing technology, big data technology has significant advantages in processing efficiency, which can quickly find favorable resources in the huge amount of data and show enterprise capabilities; third is the type of diversity, such as text, images, and video [2]. Under the influence of this feature, it can provide strong support for data capacity expansion, and when the capacity becomes large, it can provide many facilities for more people to collect relevant information, so that the data needs of more groups can be fully satisfied, and it can also promote the efficiency of data processing and analysis, and provide positive auxiliary effects for scientific management decisions.

Decision-making as the central line of enterprise management, including strategic decisions and related decisions, belongs to a highly dynamic and complex management behavior, which has a decisive role in the state of enterprise operation and development, as shown in Figure 1:

(1) The Influence of the Decision-Making Environment. The information content in the era of big data is in dynamic change, in which the volume of data has changed dramatically, from TB, PB to EB in the past, the storage volume has also increased rapidly, the decision-making environment has been greatly affected, and the decision-making mode has changed under the data drive [3]. According to the current situation of the application of big data in decision-making, it can be seen that the information processing efficiency of most enterprises is low, which affects the effectiveness of big data.

(2) The influence of Decision-Making Data. Through long-term development, data have been greatly improved in terms of type, quantity, and structure [4]. In business operation, data collected through the information platform should be uncluttered and organized, during which it should be purposefully selected and screened, then the data and information should be continuously optimized, and finally the existing information processing system should be comprehensively upgraded. In the fickle information dissemination environment, strong technical assistance can be provided for real-time data processing and focus on the connection between big data and information, thereby mining information closely related to the enterprise to achieve the goal of sound development.

(3) The impact of Decision-Making Participants. On the one hand, the role of participants has changed. After the birth and application of big data, traditional decision-making schemes cannot adapt to the development of the times and the development of enterprises, so decisions need to be made on the basis of more refined and reasonable analysis. For the senior leaders, they should change the previous wrong decision-making methods, not simply according to past experience to give orders, but to comprehensively collect data and information, and combined with the actual situation of the enterprise, the deployment of tasks as the key elements of decision-making, carefully arranged and arranged to ensure the maximum use of human resources.

(4) Decision-Making System Impact. The system mainly includes two elements: one is the basis for decision-making, and the other is the decision-making process. In the past, enterprise decision-making was mainly based on the data of internal information system and report data, with strong one-sidedness and subjectivity, which could only reflect the operation and financial management of the enterprise [5]. With the rapid development of the network, enterprises can quickly and easily collect information from other enterprises so as to understand the market price fluctuations, market demand, consumer evaluation, and other information. And the above information is applied to the decision, to make the decision more objective and comprehensive, which can help help enterprises clear the development direction, timely avoid market risks, and enhance their core competitiveness.

At the same time, enterprises can also create decision management systems supported by big data and create integrated systems corresponding to different departments to fully demonstrate the practicality, expandability, and comprehensive functions [6]. Relying on the integrated system, enterprises can open up channels to collect relevant data sources, grasp user behavior and feedback, and track and collect user behavior as a basis for optimizing product design, which helps products better meet consumer expectations, increase sales, and gain more economic benefits. In addition, the content and form of decision-making are increasingly complex due to big data.

In enterprise decision-making, the content of data analysis should be closely focused on promoting the full play of big data technology. Data analysis results are affected by the ability of employees, such as staff analysis ability is weak, it is easy to collect a lot of worthless information, and cannot use the data correctly, resulting in a waste of data resources, while employees with strong analytical ability can accurately grasp the valuable information to promote the effectiveness of data resources [7]. In this regard, a platform should be created for data analysis during enterprise development to achieve higher efficiency with lower cost as possible [8].

The main contributions of this paper are as follows. In this paper, we propose an artificial intelligence decision-
making platform in which anomalous data detection methods and width learning methods are combined to better utilize big data technologies to solve major decision-making problems for enterprises. The model can obtain information related to time series from time series and learn the patterns of normal data. Then, based on previous research, the code width learning network algorithm is proposed and applied to multi-scene and multi-objective decision-making tasks. Combined with the corresponding parallelized training algorithm, the feasibility and efficiency of the algorithm are discussed at the level of algorithmic principles. Finally, it is experimentally verified that the proposed method can detect abnormal data, improve the accuracy and reliability of decision-making, and assist the relevant personnel in decision analysis.

2. Related Works

2.1. Current Status of Big Data Technology Research. To respond to users’ needs in a short time, accurately complete data analysis tasks, and visualize the results to users is not encountered in traditional data analysis and processing. At present, a lot of researches are aimed at solving the problems faced by big data in various stages of generation, collection, storage, analysis and mining, and visualization.

Among them, big data collection is everywhere, and its sources cover finance, medical, Internet, transportation, communication, education, scientific research, and other fields [9]. Visual presentation: In order for users to better understand the results of data analysis and mining, the knowledge or patterns mined need to be visually displayed to users at the terminal in a friendly and easy-to-understand way to provide advice or support for user decision-making.

Big data analysis and mining is an important technology to transform massive, complex, high-speed, and low-density big data into knowledge or patterns for human production and life services. Second, the clustering algorithm module uses the density k-means algorithm as the basis for selecting the initial clustering centers; finally, the CUDA architecture and MPI message passing interface are used to achieve parallelism and reduce the time overhead of the algorithm [10]. On this basis, some researchers have focused their research on textual big data on the semantics of big data and given a constraint model based on clinical document standards and user use case consistency, which solves the semantic loss problem in the process of traditional medical big data document division.

On this basis, some researchers have worked on overlapping community structures and discovered community structures based on complete subgraph percolation, which has been successfully applied in biological, information, and social networks; further, some researchers have proposed a new community discovery method using aggregated hierarchical clustering techniques, which can reveal both network hierarchies and overlapping community structures [11]. Some researchers choose a series of important video clips to represent the original video and then use the features of the original video to smooth the video clips to get a smoother video summary. Based on this, some researchers construct a video hypergraph model and use hypergraph sorting to classify videos according to different contents, and finally generate video summaries by function optimization.

In order to analyze the semantics of mobile data to detect anomalies in mobile object activities, some researchers have studied both temporal and spatial aspects of mobile object trajectory data. For the mobility prediction problem in mobile data, a new evolutionary algorithm has been proposed, which predicts the next movement of a mobile user in a personal communication system through three stages: movement pattern mining, movement rule extraction, and
In order to better and fully utilize the value of data, it is usually necessary to govern the huge amount of data. One of the more important tasks is anomaly data detection. In many data mining and statistical literature, anomalies are also referred to as inconsistencies.

Almost all anomalous data detection methods nowadays create models of the normal patterns of the data and then calculate or measure data anomalies based on their own deviations from these models. The principle of anomaly data detection can be broken down into two specific sub-problems [13]. The first problem is the problem of how to define anomalous data for a given dataset. The second problem is how to select an effective anomaly data detection method for anomaly data based on the characteristics of the dataset.

The basic problems of anomalous data detection have been effectively solved from different perspectives. The global anomaly detection model gives a binary label to the data to be tested whether it is anomalous or not. The local anomaly detection model calculates the anomaly score of each data object, which indicates the probability that each object is anomalous data [14].

Since this method appeared before the emergence and popularity of computer technology, it does not take into account the various problems encountered in the actual use of data representation, computational efficiency, and computational complexity in the process of anomaly detection [15]. Despite the problems of this method, the idea of mathematical modeling is very useful in many computational scenarios. One of the more commonly used statistical model-based anomaly detection methods is to detect anomalous data by detecting extreme univariate values.

The distance-based anomaly data detection method is an anomaly data detection method that can span various data domains, which uses the nearest neighbor distance to define anomaly measurement criteria.

Later, techniques based on unsupervised representation learning began to emerge, such as subspace feature selection methods, neural networks, and stream learning methods. Subspace-based feature selection methods reduce the impact of irrelevant features by finding subsets of features that are relevant to the anomalous data, and then perform regular anomaly detection on these feature subsets. This approach usually separates subset selection and anomaly detection, which results in features that are irrelevant to the anomaly data being used for anomaly detection [16]. Therefore, this approach results in reduced accuracy and large bias in anomaly detection. Neural network and stream learning-based approaches focus on retaining regular information about the data, which is then used for learning tasks such as clustering and data compression.

In summary, in the field of anomaly data detection, the problem of anomaly detection for basic data types or low-dimensional data is relatively mature. However, there are still many problems in detecting anomalous data when facing high-dimensional data. In addition, with the continuous development of various industries, data have been fully accumulated, and the data in many fields have become very large.

In the early days of intelligent decision-making applications, relatively simple models such as linear regression algorithms and decision tree models were mainly used, and humans could easily understand the logic and make decisions directly. From the perspective of user needs, explainable AI methods can be classified into four categories: visual explanation for intuitive detection of the interior, exploration explanation from external perturbation, knowledge explanation based on user common sense, and causal explanation reflecting decision causality, which is expected to explain AI black box decisions into transparent ones [19]. In terms of application research, the application research of explainable AI in the fields of intelligent medical care, unmanned driving, intelligent finance, intelligent justice, etc., has also been carried out, with the goal of improving the reliability of AI-enabled system predictions or decisions.

Big data-driven AI systems should not be influenced by human subjectivity, but flawed data can lead to biased and unfair decisions, and adversarial data can lead to serious decision errors, which seriously affects the credibility of AI-enabled system applications and decisions. At the methodological level, research on how to address inequity has evolved through three phases. The first stage is perceived fairness, which investigates how to deal with protected attributes directly to obtain fairness, using differential treatment such as directly excluding protected attributes such as race and gender in the decision-making process.
In a multi-risk environment, AI-enabled systems raise uncertainty in both data and model dimensions. Data risk will lead to changes in prediction or decision results, model risk can present incorrect results for the inputs, further affecting the continuity of trust, and the above risks lead to increased uncertainty as shown in Figure 2. AI-enabled systems under multiple uncertainties make trustworthy decisions for information situations where the data sources are cross-domain and the data are no longer perfect, but may be biased or adversarial risk data, leading to unfair or incorrect results of the decision model. In the process of human-machine collaborative decision-making, the results at the case level need to be calibrated for trust, for the purpose of enhancing human or AI single decision-making.

In high-stakes decision-making, many human experts rely on the output of AI to form the final decision, forming human-machine teams. Research suggests that these human-machine teams may perform better than human-only or AI-only teams, and that in order to achieve this, humans must have a moderately calibrated level of trust; otherwise, the AI’s trust level will be mist-calibrated and the human-machine team will not perform as well as the human decision alone [20]. For trust calibration, some researchers have suggested the need for trust calibration that understands the capabilities of the system and the reliability of the system output. Thus, enhancing the understandability, responsiveness, and ability to resolve conflicting goals of AI-enabled systems may be far more meaningful than simply improving the accuracy of AI.

Metrics and standards for AI decision-making as research on AI-enabled system decision-making progresses: the problems of different principles and methods in industry applications will gradually be exposed, and metrics and standardization of decision-making are the way to go. On the one hand, we need to design AI decision-making from the perspective of developer training, testing and experimentation, deployment and operation, and supervision, and on the other hand, we need to develop standards and specifications from hardware, algorithms, and systems such as chips. Future research should focus on the metrics and standards of AI-enabled system decision-making, coordinate guidance, and regulation of decision-making, and promote the healthy and sustainable development of AI decision-making.

3. Algorithm Design

3.1. Abnormal Data Detection Method. In anomaly detection, whether the target data are anomalous at a certain moment is related to its timing information, contextual information, and the numerical information of the data itself. Some characteristics of the anomaly detection problem itself also make it different from the conventional classification problems, such as the fact that the positive and negative samples of the data corresponding to anomaly detection are usually severely unbalanced, and there are sometimes unlabeled anomalous data in the data.

Recent research work on time series anomaly detection has made more use of modeling methods related to deep learning. In this thesis, we focus on analyzing recurrent neural network-based and vibrational self-encoder-based anomaly detection algorithms.

In anomaly detection problems, there are usually far more normal data than anomalous data, so the anomaly detection problem cannot be solved using the conventional classification problem solution [21]. Due to this nature of the anomaly detection problem, solving the anomaly detection problem using unsupervised learning algorithms has become an idea for researchers. According to existing time series decomposition methods, such as STL methods, the time series can be decomposed into several combinatorial parts. The decomposition of the time series is represented using the following formula:

\[ x_t = \tau_t + c_t + s_t + i_t. \]  (1)

In the formula, \( x_t \) denotes the numerical vector of the time series at moment \( t \), \( \tau_t \) is the trend part, \( c_t \) is the periodic part, \( s_t \) is the seasonal part, and \( i_t \) is the irregular part. The retained part of the time series will be used as the object of anomaly detection model processing. The following explains the reasons for using the time series decomposition method to process the sequences. In many cases, the amount of data for training anomaly detection models is not sufficient, so anomaly detection models based on generative models do not learn the pattern information inside the data well. There may be noisy parts in the time series and some parts that are not related to anomaly detection, and these parts can interfere with anomaly detection to some extent. By decomposing the time series using the time series decomposition method, it will be less difficult for the model to perform anomaly detection on the processed time series, which is equivalent to the simplification of the problem and is beneficial for the model to perform better anomaly detection.

The complex distribution of data can be learned by unsupervised learning, the structural features of recurrent neural networks (RNN, LSTM) make them suitable for scenarios involving sequences, recurrent neural networks are able to obtain time series-dependent information from time series, and this thesis uses recurrent neural networks in time series anomaly detection. There have been research works applying LSTM to time series anomaly detection and achieving better results in some scenarios. In order to better obtain time series-dependent information, we consider adding more neural network processing layers for processing time series in VAE, but the LSTM layers have more parameters, and the presence of too many LSTM layers in the model will increase the difficulty of model training, so RNN layers are introduced. The overall structure of the anomaly detection model is consistent with that of VAE, and the algorithm performs model training in an unsupervised learning manner, using normal time series data to train the model and determine anomalies based on reconstruction errors [22]. The general process of solving the problem is to process the time series using the time series decomposition method and subsequently perform anomaly detection on the processed time series using the cyclic VAE model.

Figure 3 shows the general framework of D-R-VAE, which can be divided into 3 parts. The anomaly detection
model uses a combination of recurrent neural network and VAE. Each module is described below.

For nonperiodic time series, we utilize the HP filtering method. This decomposition method can decompose the time series data into two components, the trend part and the remaining part. We retain the trend part of the time series decomposed by the HP filtering method for subsequent anomaly detection. The trend part of the time series becomes insensitive to short-term fluctuations compared to the original time series, which is equivalent to removing the interference of noise and other factors to some extent. The sensitivity of the trend part to short-term fluctuations can be adjusted using the parameters in the decomposition function, and suitable parameters are selected for the HP filter decomposition function according to the nature of the dataset during the experiment.

D-R-VAE decomposes the time series into multiple components using the time series decomposition method and removes the unnecessary components of the series for the anomaly detection task. The cyclic VAE model performs anomaly detection on the retained time series, and a combination of VAE, LSTM, and RNN is used in the model.

Figure 4 shows the framework flowchart of the algorithm. The execution process of the algorithm can be divided into two main parts: firstly, the time series are decomposed using the classical time series decomposition methods in statistics and mathematics. Time series with obvious periodicity are decomposed by the STL method, and time series without periodicity are decomposed by the HP filtering method. After the sequence is decomposed by STL method, the remaining part is retained, and after the sequence is decomposed by HP filter method, the trend part is retained. The retained part of the time series is used as the new time series. At the end of the first part, the new sequence is divided into a set of fixed-length subsequences and further divided into a training set and a test set. In the second part of the algorithm, the D-R-VAE model performs anomaly detection on the processed time series [23]. The D-R-VAE model is first trained using normal time series, and subsequently the anomaly score output from each subsequence in the training set data after processing by the model is retained and used to calculate the anomaly threshold for the anomaly scores. In the testing phase, the time series in the test set are reconstructed using the D-R-VAE model, and the abnormal scores are calculated, and the abnormal scores are compared with the abnormal thresholds to determine whether they are abnormal or not.

The first judgment is based on the time series’ own attributes. When the origin of the time series is not clear, it is necessary to divide the multivariate time series into a set of univariate time series, visualize each univariate time series using a visualization function, and observe whether it is cyclical or not. If the periodicity cannot be judged with certainty, it is necessary to select an approximate period for the series based on the observation of the series and segment the time series according to this period, and if the distance of adjacent segments is less than the threshold, it is periodic.

The model is expanded to represent according to the length of the fixed-length window, and the structure and internal parameters of each unit after expansion are the same. The overall structure of the model is the same as that of VAE, the first neural network processing layer of both the encoder part and the decoder part of the model is the LSTM layer, and the processing layer for generating the parameters of the hidden variable distribution of the model is the RNN layer. Due to the introduction of LSTM and RNN, the D-R-VAE model can obtain the time-dependent information in the sequence when processing the time series data. Substituting the variables into the objective function of VAE, the objective function of the D-R-VAE model is

$$\sum_{t=1}^{w} (KL[q_{\phi}(z_t | x_t)] || p(z_t)).$$

(2)

The objective function is maximized during the model training. The right-hand term calculates the KL scatter of the two Gaussian distributions. In the anomaly detection problem, it is necessary to define an anomaly score for the object to be detected and determine the anomaly according to the anomaly score. For example, the data $x_t$ at time $t$ as the object to be detected are expressed by the following equation:

$$\{x_{t-w+1}, x_{t-w+2}, ..., x_t\} \rightarrow D - R - VAE$$

$$\{x_{t-w+1}, x_{t-w+2}, ..., x_t\} \ast s_t = mse(x_t, x_t').$$

(3)

The $s_t$ in the formula is the anomaly score. After the model training is completed, the anomaly scores corresponding to each moment in the training set are calculated for use in the anomaly threshold selection phase. Anomaly threshold selection: After the model training is completed, the model is used to reconstruct the time series in the training set, and we get a set of anomaly scores.
Anomaly thresholds are selected based on density. We place the anomaly scores into a one-dimensional space, and we also have a set of points distributed in the one-dimensional space. Anomalous data are data that are significantly different from normal data. The model is trained using normal data, so the reconstruction error will be larger when the model is reconstructed for the abnormal time series. Since the training data are all normal data, the scores in the set of abnormal scores corresponding to the training set are concentrated in a certain normal range, and the density of abnormal scores in the normal range is the largest [24]. However, there are a small number of abnormal scores in the training set that deviate from the normal range, and the density of these abnormal scores in the space is small. We select the threshold value among these abnormal scores. The smaller density of points at the locations of the anomaly scores leads to the distances between the points and their immediate neighbors, so the sum of the differences between the distances of the anomaly scores and their immediate neighbors in the one-dimensional space is used as a measure of point density.

3.2. Artificial Intelligence Decision-Making Platform. In order to better meet the requirements of accuracy and timeliness of enterprise decision-making, the width learning system is referred to as the training model, which finds the weight matrix by generating feature nodes and enhancement nodes and calculating the pseudo-inverse directly, which greatly reduces the prediction time [25–28].

The signaling features and the first 15 bits of the voice call are extracted as the text, the text is preprocessed accordingly to build a word vector space, and a decision method based on the width learning network is proposed, which together constitute the data preprocessing part. The proposed decision model is shown in Figure 5.

However, since the feature nodes are initialized in a completely random way for width learning, there may be a large number of inefficient node generation and a certain amount of feature information missing [29, 30]. To solve this problem and to learn the vector representation better, the denoising self-encoder adds noise to the original data in order to solve this problem:
The new variables with Gaussian noise are then fed to the noise reduction auto-encoder, which is set to zero with a certain probability, equivalent to the loss of some features of the data, to improve the noise immunity of the data, and to generate a new representation:

$$x_{\text{new}} = P(x + \text{noisy}(x, y)).$$  (5)

Next, we feed the learned features into the width learning network as coded vectors. In this way, the feature nodes can be represented as

$$Z_o^i = \phi(x_o w_o + \beta).$$  (6)

In deep learning, the common way of parallelized training is that at the beginning of each round of training, these processes or devices will first read the current parameter values uniformly and get the data. But width learning is different, because it does not back propagate the gradient for training, so we need to partition at the data level more, cut into smaller matrices to reduce the capacity, and calculate the pseudo-inverse. To better solve this problem and to try to further optimize the performance of the width learner, the integration learning approach is borrowed. The integrated learning essentially trains multiple weak learner models and combines the learning results of each learner in some way to obtain better results than a single weak learner. For the width parallelization algorithm, we can train multiple weak width learners with small data size and combine the learning results in an integrated way, which can well solve the memory explosion problem and improve the generalization performance of the model.

### 4. Experiments

#### 4.1. Abnormal Data Detection Experiment

**4.1.1. Experimental Parameter Settings.** Three algorithms are selected for comparison with the algorithms proposed in this thesis, and the benchmark algorithms are briefly described below: (1) LSTM-based anomaly detection algorithm, which uses LSTM to predict time series and determine anomalies based on the prediction error; (2) VAE-based anomaly detection algorithm, which uses VAE to reconstruct input variables and determine anomalies based on the reconstruction error; And (3) VAE-LSTM anomaly detection algorithm, which uses a combination of VAE and LSTM in the anomaly detection model and adapts the first fully connected processing layer of the encoder part and decoder part of VAE to the LSTM layer. The experimental environment is as follows: Intel i7 3.30 GHz, RAM 256 GB, and Nvidia 2080 GPU; operating system and software platform are Ubuntu 18.04, TensorFlow 1.12, and Python 3.8. The training process loss convergence curve and performance improvement are shown in Figures 6 and 7.

The experimental data in this paper were obtained from internal nonpublic data of the China Decision Evaluation Research Institute. The test datasets include (1) Server Machine Dataset (SMD), which is collected by the researchers from an Internet company and contains time series data from 3 sets of physical machines, for a total of 28 multivariate time series datasets from different machines, which need to be used independently; (2) Yahoo Benchmark Dataset (Yahoo), which contains 4 benchmark datasets, the first of which is selected to evaluate the model algorithm, and the remaining 3 datasets (synthetic datasets, where anomalies are easily identified by the model).

In the anomaly detection problem, there are usually more normal data than anomalous data in the dataset, and the data are unbalanced, so we choose the criterion for evaluating the algorithm performance when it is unbalanced. We use abnormal data as positive data and normal data as negative data, and classify the data judged by the model into four categories according to the nature of the data itself and the model’s judgment of the nature of the data. True positive (TP) is the abnormal data judged by the model as abnormal, false negative (FN) is the abnormal data judged by the model as normal, false positive (FP) is the normal data judged by the model as abnormal, and true negative (TN) is the normal data judged by the model as normal.

**4.1.2. Experimental Results and Analysis.** Firstly, we analyze the effects of parameters and window length in the objective function on the anomaly detection performance of the D-R-VAE model by increasing the parameters from 0.2 to 1.0 (each time increasing by 0.2), and evaluate the model performance on the SMD dataset with the $F_1$ score metric. The red curve in Figure 8 shows the experimental results. The model performance is optimal when the parameters are...
between [0.2, 0.6], and the model anomaly detection performance tends to decrease as the parameters are increased. For model training of the D-R-VAE model and anomaly detection using the model, a fixed-length sub-series is required as model input. The sub-series length is the time series window length, and the window length has an impact on the anomaly detection performance of the model. The blue curve in Figure 8 shows the experimental results of the D-R-VAE model with different window lengths for anomaly detection on the SMD dataset, and the anomaly detection performance of the model gradually improves as the window length increases from 80 to 140 and then stabilizes.

Table 1 shows the effect of time series decomposition on the anomaly detection performance of the D-R-VAE model prior to anomaly detection by the D-R-VAE algorithm, where the D-R-VAE model performs anomaly detection on the retained portion of the time series decomposed by a specific time series decomposition method, and the R-VAE algorithm uses the same model to perform anomaly detection on the original time series. The experimental results show that the performance of the D-R-VAE algorithm is significantly better than that of the R-VAE algorithm, thus indicating that the anomaly detection performance of the D-R-VAE model can be effectively improved by using the time series decomposition method to process the time series appropriately.

Figure 9 shows the visualization results of the D-R-VAE algorithm for anomaly detection on the SMD dataset, where the red dashed line corresponds to the anomaly threshold, the red points correspond to the anomalous points, and the blue points correspond to the normal points. The number and density of red abnormal points and blue normal points are based on the abnormality threshold above. It can be seen from Figure 9 that the D-R-VAE model performs better, the red anomalous points above the anomaly threshold are more intensive, and the number of red anomalous points above the threshold is significantly more than the blue anomalous points, thus indicating that the anomaly scores generated by the D-R-VAE model can better distinguish between anomalous and normal points.

4.2. Decision Platform Experiment
4.2.1. Experimental Parameter Setting. A total of about 500,000 pieces of labeled data are used to verify the algorithm effect. The data are divided into 6 categories according to the decision type. At the same time, the training set and the validation set were divided into 8:2 ratios according to the data time series, and the feature generation process was completed as described above.

First, the features used are modeled accordingly to fill in the missing values for continuous and discrete numerical
features, and a large number of missing fields are removed. Based on the a priori knowledge, statistical features are generated, such as the number of calls and the average call duration of each call in a certain period of time. The discrete features are combined to construct cross features; the continuous features are varied to transform the value range. The features are evaluated by the feature importance evaluation method mentioned above, and the feature selection is completed. The preprocessed feature set is imported into the code width learning model classifier, and the effect of the model is adjusted by adjusting the number of feature nodes and enhancement nodes.

To verify the effectiveness of code width learning network in enterprise decision-making, it is compared with several classical neural network algorithms including convolutional neural network (CNN), recurrent neural network (RNN), and deep confidence network (DBN) without using our proposed parallel algorithm.

4.2.2. Experimental Results and Analysis. First, select four groups of parameters with high accuracy and draw the AUC characteristic curve, as shown in Figure 10. It can help us better judge whether the classifier is effective enough and find the best decision data parameters in the model. It can be seen from Figure 10 that the AUC is greater than 0.9, which shows that the classifier has a good effect in enterprise decision-making. In this figure, the maximum value of AUC is 0.94. When 400 feature nodes and 2000 enhancement nodes are set, better results can be obtained. The experimental results also show that it is necessary to reasonably adjust the parameters of the width learning network. Increasing the number of feature nodes and enhancement nodes can help the module better interpret features and improve the accuracy of the model. However, if the number of feature nodes and enhancement nodes is too large, the accuracy will be reduced and the model will be over fitted. Using the existing data, the maximum accuracy can reach 94.38%.

Table 2 gives the experimental results of comparing the coding width learning model with the classical algorithm. From Table 2, we can see that the accuracy of our proposed coded width learning network and recurrent neural network is high, reaching 94.38% and 92.75%, respectively, while the original width learning is slightly lower compared to the improved one, but its accuracy is also high compared to most models. It reached 91.86%. From the perspective of recall, the coded width learning network has the highest recall rate of 93.46%. In terms of training time, the original width learning network is the fastest. The whole training process takes only 38.29 seconds, which is hundreds of times faster than other algorithms. Our proposed coded width learning network model takes slightly longer time due to the addition of coded feature extraction, but it also has obvious advantages over other models. In general, the code width learning network has higher accuracy and highest recall compared to several classical models, and the training time is very short. Therefore, it has excellent performance in business decisions that require very high timeliness and accuracy.

5. Conclusions

In the era of big data, enterprise management and decision-making behavior has changed greatly, mainly in terms of decision-making environment, data, participants and system, etc., which put forward higher requirements for the
leadership and managers, who should actively establish data awareness, rely on big data to comprehensively grasp the current market situation change pattern, and combine with their own situation to scientifically develop decision-making solutions. To this end, combined with big data analysis technology, this paper proposes an artificial intelligence decision-making platform to assist relevant personnel in solving major decision-making problems. First, for the problem of anomaly detection of time series data, the D-R-VAE algorithm is proposed to decompose the time series and retain the components of the time series related to anomaly detection. The algorithm uses the model structure of VAE to learn the normal patterns of time series data, the LSTM and RNN layers included in the model make the model more suitable for processing sequence-like data, and the model is able to utilize the time series-dependent information of time series in the process of processing data. Then, an artificial intelligence decision platform is constructed based on width learning, which includes a decision algorithm based on coded width learning network and a parallelized training algorithm based on width learning model. The former draws on denoising self-coding and width learning to achieve the requirement of high timeliness and high accuracy in decision-making; the latter combines width learning features with integrated learning methods for parallelized training to further improve efficiency and solve the memory explosion problem. Finally, the experiments show that the D-R-VAE model can effectively detect anomalous data and provide data reference for the subsequent decision platform; the width learning-based decision platform requires short practice time and high accuracy, and can meet the real-time decision requirements. In the future, we plan to use graph convolutional neural network for artificial intelligence decision platform building and application of big data analysis technology.

**Data Availability**

The datasets used during the current study are available from the corresponding author on reasonable request.

**Conflicts of Interest**

The authors declare that they have no conflicts of interest.

**References**


