

Research Article

Application of the Neural Network Based on the Multilayer Perceptron Genetic Algorithm in Chinese-English Two-Way Translation

Yuxiu Yu 

School of Foreign Languages, University of Sanya, China

Correspondence should be addressed to Yuxiu Yu; yuxiuyu@sanyau.edu.cn

Received 11 March 2022; Revised 15 April 2022; Accepted 18 April 2022; Published 16 June 2022

Academic Editor: Yuan Li

Copyright © 2022 Yuxiu Yu. 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 increasing application of the perceptron genetic algorithm neural network in Chinese-English two-way translation, there are many translation problems to be solved. In order to solve the translation problem of Chinese-English parallel corpus, the multilayer perceptron method, genetic word alignment model (GA), language model, and neural network method (including the translation model and bilingual pretraining model) are designed, which are combined into the ga-mlp-nn combination model to measure the parallelism of Chinese and English sentences from different emphases. The results show that the ga-mlp-nn model has good performance in filtering high-quality parallel corpus. The final experimental results show that compared with a single system, the improved multisystem fusion method based on weight multiplication has achieved better results in the test set. In the last five groups of evaluation results, the system submitted in this paper ranks second and first in multiple datasets, which has a certain reference value for the research of corpus filtering.

1. Introduction

With the development and in-depth application of information technologies such as mobile Internet and mobile communication, the number of online users in major social media platforms and e-commerce platforms has increased sharply. As one of the important ways of information interaction between enterprises, platforms, and consumers, online comments came into being with a huge number. An online comment is the user's comment on a specific topic through the network platform. It is the user's feeling and evaluation based on their own use or experience of products/services [1]. Because the online comments are rich in consumers' feelings and expectations after the product/service experience, involving preferences, needs, and evaluation in terms of quality, design, price, and service, the commercial value is becoming increasingly prominent, which has attracted great attention from enterprises, platforms, and consumers. However, due to the cross-platform interaction and information diffusion effect, the comment information of some topics, especially hot issues, is growing explosively.

In the face of a large amount of comment information, it is often difficult to effectively distinguish the valuable comment information and correctly judge the true situation of products/services. In addition, due to the anonymity of users, the difference of motivation to evaluate products/services, and noncontact in the network platform, the quality of comment information is uneven and the authenticity of comments is uneven. In view of the overload and uneven authenticity of online comments, the cost of information screening is greatly increased and the efficiency is greatly reduced. It is not operable to manually screen the useful information of online comments [2]. Therefore, how to scientifically and effectively mine the useful information of online comments has become an important research topic, which is very important for the decision-making efficiency and effect of enterprises, platforms, and consumers.

Most of the existing studies on comment usefulness use the number of comments, text depth, and text emotion as the main factors affecting comment usefulness and rarely consider the impact of the professionalism of the comment content on comment usefulness. The innovative

contributions of this paper include the following: (1) The multilayer perceptron method and statistical method are used to screen and filter the expected value. (2) The GA MLP neural network is used as classifier for comment usefulness recognition. This paper creatively analyzes the influence of the proposed features on the usefulness of comments and obtains the best feature combination. And (3) compared with a single system, the improved multisystem fusion method based on weight multiplication has achieved better results in the test set. It provides a reference value for the translation of Chinese-English parallel corpora.

Based on this, this paper uses the knowledge adoption model and multilayer perceptron neural network to identify the information usefulness of online comments. This paper is divided into five parts. The first part is the research background, and the second part is the literature review to analyze the research results of the problem. The third part is the introduction of the multilayer backbone straight up genetic algorithm neural network and related systems. The fourth part is the specific experimental analysis. It shows how the multilayer perceptron genetic algorithm neural network deals with the translation of Chinese and English parallel corpora and compares it with the traditional methods. The fifth part is the round of the article.

2. Related Work

Theoretically, the usefulness of comment information depends on whether one thinks the that comment information is valuable and helpful to varying degrees [3]. In the existing studies, Yuanyuan et al. [4], Mudambi and Schuff [5], and Baek et al. [6] studied the factors affecting the usefulness of online comments from the text characteristics and found that the factors such as comment emotion, comment depth, comment grade, and reviewer reputation have a significant impact on the usefulness of comments. Chen and Xie constructed a normative model to judge the usefulness of comments and help consumers determine the best matching products [7]. Liu Wei and Pengtao and Zhang et al. constructed a theoretical model of two-way analysis based on the IAM (information acceptance model) to effectively identify the key factors affecting the usefulness of online comments on e-commerce platforms [8, 9]. Rui and Jian studied the impact of the inconsistency between the star rating of online comments and the average star rating of products on the usefulness of comments based on the attribution theory [10]. Liu et al. analyzed the impact of the number of comments, reviewer professionalism, reviewer reputation, and other factors on the usefulness of community online comments. In terms of the selection of influencing factors of comment usefulness, most of the existing studies consider the internal factors such as the star rating, number of comments, depth of comments, comment emotion, and external factors such as the identity, professionalism, and reputation of the comment publisher but few studies analyze the influencing factors from the professionalism of the comment content [11–18]. Because the comment information of each field contains its specific domain words, there are serious

domain barriers. At present, there is no general method to effectively extract the factors related to the comment quality from the comment text. Existing studies show that compared with the feature extraction method without considering domain words, the feature extraction method based on the domain dictionary can enhance the expression ability of text features and improve the classification performance [19]. Therefore, this paper proposes a feature extraction method based on the domain dictionary to construct a measurement index of the comment text quality, so as to improve the recognition effect of comment usefulness.

In the existing research, there are a large number of theories that can be used to explain the usefulness of online comments, one of which is the knowledge adoption model (GA) [20]. The model holds that the information receiver's perception of information usefulness is the direct decision of knowledge adoption, and for a given information, the determinants of perceived information usefulness include the quality of the information itself and the credibility of the information source. Many scholars have studied the usefulness of online comments based on the GA model. For example, Cheung et al. and Huang et al. used the GA theory to study the usefulness of online comments [21, 22]. Erkan and Evans analyzed the usefulness of online comments and its effect on purchase intention on the basis of the GA theory [23]. It is not difficult to see that the GA theory has achieved good results in explaining the usefulness of online comments. In order to more comprehensively analyze the usefulness of comment information, this paper uses the GA model to construct the characteristics of the comment usefulness classification model from two aspects: the quality of the comment text and the credibility of the comment source.

At present, many researchers construct the judgment comment usefulness as a text binary classification problem. The solution to this kind of problem can adopt the text classification method based on machine learning, mainly including the support vector machine (SVM), decision tree, neural network, and Bayesian [24, 25].

3. Construction of the Multilayer Perceptron Genetic Algorithm Neural Network System (GA-MLP-NN)

Although a single perceptron cannot solve the XOR problem, it can realize the segmentation of complex space by combining multiple perceptrons. The two layers of perceptron are combined according to a certain structure and coefficients. The first layer of perceptron realizes two linear classifiers to divide the feature space, and an XOR operation can be realized by adding a layer of perceptron on the output of the two perceptrons. That is, it is composed of multiple perceptrons. Multilayer neural networks are also trained by the gradient descent algorithm, which is an algorithm dedicated to finding the extreme point of loss function to minimize the value of loss function. The so-called "learning" is to improve the model parameters

in order to minimize the loss through a large number of training steps.

The methods used in the system submitted in this paper can be divided into three categories: multilayer perceptron method, statistical method, and neural network method. The overall architecture of the system is shown in Figure 1. The multilayer perceptron method mainly designs a series of rules to filter corpus whose quality obviously does not meet the requirements. The statistical methods include the Zipporah system, genetic word alignment model, and language model. The purpose of filtering is achieved by statistical feature information on a large number of clean corpus. The neural network method includes a translation model and bilingual pretraining model. The model with strong generalization ability is trained on the clean corpus, and then, the translation corpus is filtered. Finally, according to the results of different methods, the excellent methods are weighted and fused to obtain the final clean corpus.

3.1. Multilayer Perceptron. The equivalent filtering methods of the sentence length ratio, maximum sentence length, and unique sentence are used to filter the corpus. According to his work, this paper formulates four rules:

- (1) According to the length filtering rule, the sentence pairs with the length of more than 80 words at the source or the target end will be scored 0 points; otherwise, they will be scored 1 point
- (2) According to the length ratio restriction rule, if the length ratio between the source and target sentences exceeds 1.7, 0 point will be recorded; otherwise, 1 point will be recorded
- (3) Language identification rules: langid (<https://github.com/saffsd/langid.py>) is used to identify the source and target languages. If the language is incorrect, 0 point will be recorded; otherwise, 1 point will be recorded
- (4) For the deduplication rule, record 1 point for the first occurrence of repeated sentences; otherwise, record 0 point

With the help of the abovementioned four rules, a four-dimensional feature can be obtained for a given sentence pair and the value of each dimension is 0 or 1.

3.2. Statistical Method

3.2.1. Zipporah System. As a part of the fusion system and achieved good results, the Zipporah system is a fast and scalable system, which can select “good data” of any size from a large number of translation data pools. For the training of neural machine translation model, the principle is as follows: firstly, the sentence is mapped to the feature space, which contains two features: adequacy score and fluency score. Then, logistic regression is used for binary classification and the categories are “good data” and “bad data,” respectively. Finally, equation (1) is used for normalization to obtain the score of parallelism

$$S_Z(x) = \frac{x - x_{\min}}{x_{\max} - x_{\min}}, \quad (1)$$

where x is the score of the Zipporah system.

3.2.2. Genetic Word Alignment Model (GA). There are few word alignments for nonparallel sentence pairs, so this paper considers using word alignment for corpus filtering. First, use the fast align (https://github.com/clab/fast_align) word alignment tool to train on the translationless Chinese-English parallel corpus provided by the 16th National Machine Translation Conference (CCMT 2020), and then, predict the translation corpus. The word alignment score of sentence pairs can be obtained directly. In the fast align tool, the word alignment score is calculated by logarithmic summation of word alignment probability, so the longer the sentence, the smaller the word alignment score, which means that the system prefers short sentences. In order to reduce the impact of sentence length on the word alignment score, equation (2) is used to calculate the parallelism score in this paper:

$$S_{\text{avg}} = \frac{2S_{\text{align}}}{l_{\text{source}} + l_{\text{target}}}, \quad (2)$$

where S_{align} is the word alignment score of sentence pairs and l_{source} and l_{target} are the length of the source and target sentences, respectively.

After the word alignment scores of sentence pairs are processed according to equation (2), they are sorted according to the scores from high to low. After statistics, it is found that the number of sentence pairs with word alignment scores greater than or equal to -4.5 is about 4 million, about 100 million words. In this paper, it is determined that the quality of these sentence pairs is good and their scores after normalization should be high, so equation (3) is designed to normalize the scores:

$$S'_{\text{avg}} = \frac{-4.5}{S_{\text{avg}} - 4.5}. \quad (3)$$

3.2.3. Language Model. Because the language model can filter out nongrammatical data, this paper considers using the language model to filter the corpus. This paper selects the corpus without translation to generate the language model and uses the language model to calculate the perplexity (p) score of the dataset to be filtered.

Specifically, the srilm (<https://github.com/BitSpeech/SRILM>) tool is used to train a 5-gram language model for Chinese and English materials on the bilingual corpus without translation and this language model is used to calculate the confusion score of Chinese and English sentences in the bilingual corpus to be filtered. This paper uses two scoring strategies: the sentence level confusion score and word level confusion score.

In order to standardize the two piecewise functions to distinguish the differences between them, this paper standardizes the confusion score. In the standardized operation,

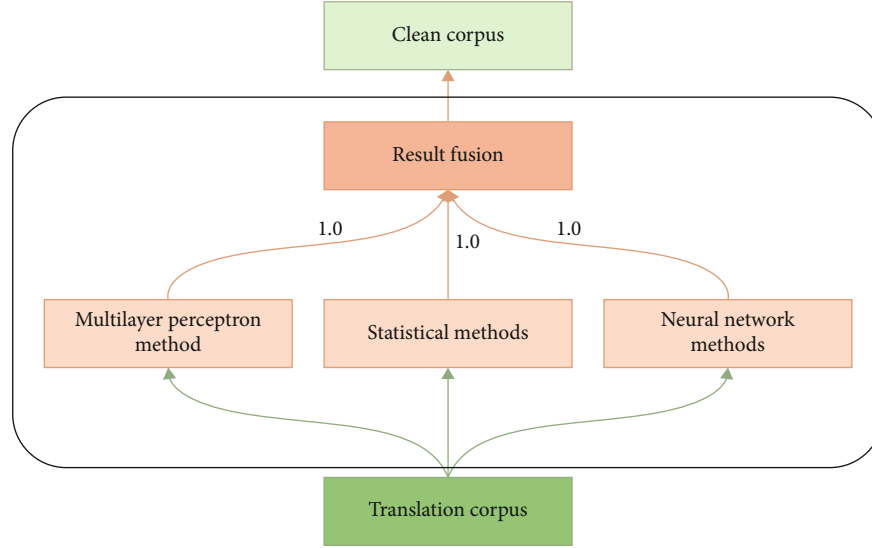


FIGURE 1: Construction diagram of the neural network system of the multilayer perceptron genetic algorithm.

a series of piecewise functions are designed according to experience. For the sentence level confusion score of Chinese corpus to be filtered, the normalized piecewise function designed in this paper is shown in formula (4) as follows:

$$S_1^L(x) = \begin{cases} 1 - 0.001p, & p \leq 100, \\ 1 - \left((p - 100) \frac{0.4}{900} + 0.1 \right), & 100 < p \leq 1000, \\ 1 - \left((p - 1000) \frac{0.3}{9000} + 0.5 \right), & 1000 < p \leq 10000, \\ 0.1, & p > 10000. \end{cases} \quad (4)$$

For the sentence level confusion score of the English corpus to be filtered, the designed normalized piecewise function is shown in equation (5) as follows:

$$S_2^L(x) = \begin{cases} 1 - 0.003p, & p \leq 100, \\ 1 - \left((p - 100) \frac{0.4}{900} + 0.3 \right), & 100 < p \leq 1000, \\ 1 - \left((p - 1000) \frac{0.15}{9000} + 0.7 \right), & 1000 < p \leq 10000, \\ 0.1, & p > 10000. \end{cases} \quad (5)$$

In addition, this paper considers the word level confusion score, calculates the word average confusion score of each sentence on the Chinese-English dataset and the word average confusion score on the overall dataset, and designs two piecewise functions to normalize the difference between

them. Therefore, when calculating the average word confusion score of the overall dataset, this paper ignores sentences with a confusion degree of more than 10000.

For the word level confusion score of Chinese corpus to be filtered, the designed normalized piecewise function is shown in equation (6) as follows:

$$S_3^L(x) = \begin{cases} 1, & p \leq 0, \\ 1 - 0.003p, & 0 < p \leq 100, \\ 1 - \left((p - 100) \frac{0.6}{900} + 0.3 \right), & 100 < p \leq 1000, \\ 0.01, & p > 10000. \end{cases} \quad (6)$$

Finally, each parallel sentence pair will get 4 feature scores.

3.3. GA-MLP-NN Neural Network. The neural network is a highly parallel information processing system, which has very strong adaptive learning ability, does not depend on the mathematical model of the research object, and has good robustness to the changes of system parameters and external disturbances of the controlled object. GA-MLP-NN is a deep neural network, which overcomes the weakness that the perceptron cannot recognize linear inseparable data and can get better expression effect. The model is shown in Figure 2.

3.3.1. Translation Model. Based on the following assumption, if sentences a and B are parallel sentence pairs, then, the semantics of a and B are similar; when a is translated into a' , the semantics of a' and B are still similar. In order to realize the abovementioned assumption, we should first train an English-Chinese translation model, then use the translation model to translate English sentences into corresponding translations, and finally calculate the similarity between the translation and the reference translation. For

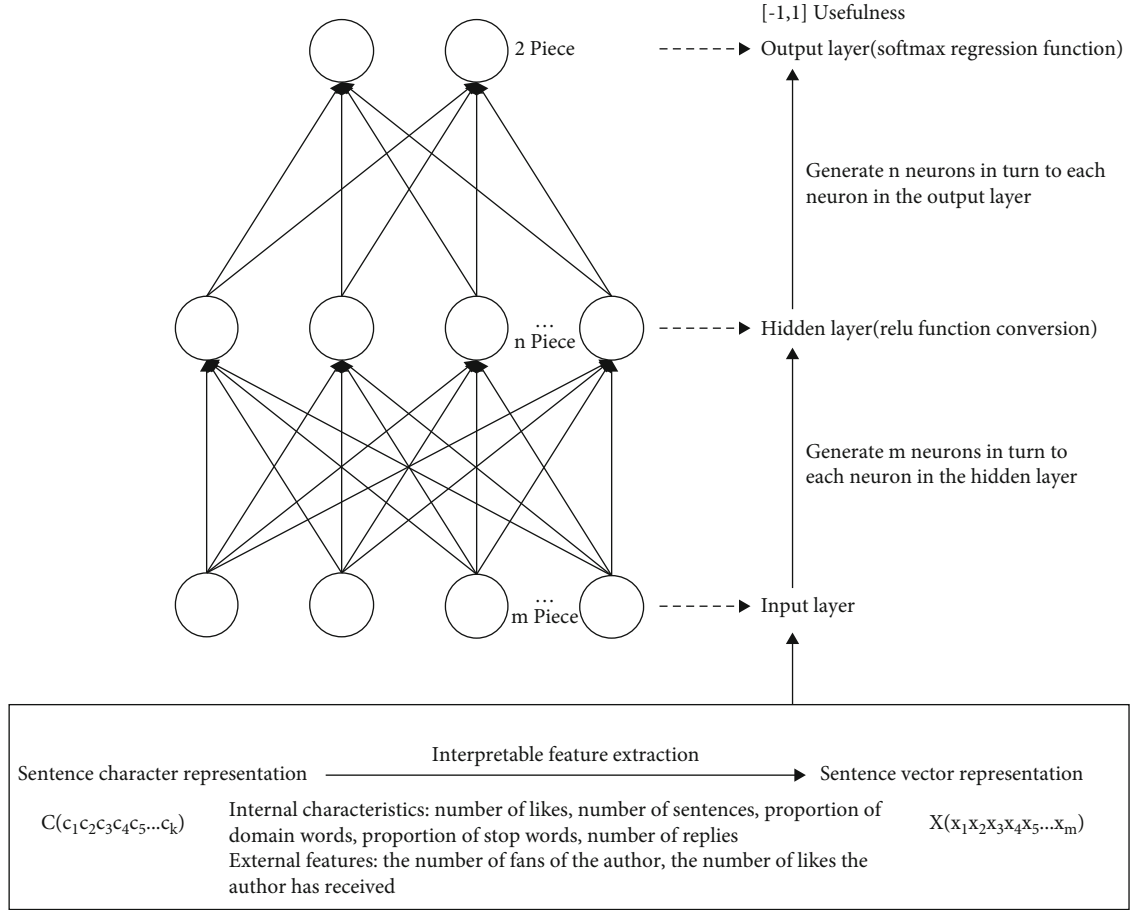


FIGURE 2: Schematic diagram of the GA-MLP-NN composite model.

similarity calculation, this paper uses two indicators: word-based editing distance and cosine similarity based on a pre-trained word vector, and finally forms a two-dimensional similarity feature.

3.3.2. Model Design. According to the abovementioned introduction, if you want to calculate the similarity between the translated translation and the reference translation, you should first get the translated translation, so you need a translation model. This paper adopts THUMT (<https://github.com/THUNLP-MT/THUMT.git>), an open-source neural machine translation tool of Tsinghua University. The system has less dependence and simple training and is suitable for rapid training of the neural machine translation system.

The training set data comes from the parallel corpus provided by the CCMT 2020 Chinese-English translation task, which is segmented and lowercased, and the sentence pairs with a length of more than 150 words are filtered to form about 10 million pairs of training data. The development set is the development set specified by the CCMT 2020 Chinese-English parallel corpus filtering task.

The main training parameters are selected by default and run for about 20 rounds. The five models with the highest bilingual translation evaluation results (BLEU) in the development set are saved; then, the models are averaged and

fused into a final model, with the direction of English \rightarrow Chinese, which is recorded as M_0 . Then, M_0 is used to decode the Chinese and English sentences of parallel sentences with translation to obtain the corresponding Chinese translation.

3.3.3. Word-Based Editing Distance. This index is essentially the editing distance, but the granularity for calculating the matching degree of two sentences is words, not a single character. Let a' and b be Chinese sentences after two word segmentation, where a' is the translation of the English source sentence a ; then, the editing distance $L_{a',b}(|a'|, |b|)$ can be calculated iteratively by equation (7) as follows:

$$L_{a',b}(i, j) = \begin{cases} \max(i, j), \min(i, j) = 0, \\ \min(L_{a',b}(i-1, j) + 1, \\ L_{a',b}(i, j-1) + 1, \\ L_{a',b}(i-1, j-1) + 1_{(a'_i \neq b_j)}), \\ \text{others,} \end{cases} \quad (7)$$

where $|a'|$ and $|b|$ are the number of words in a' and b sentences, respectively, $L_{a',b}(i, j)$ is the distance between i words

before a' sentence and j words before b sentence, and a'_i and b_j are the i word of a' and the j word of b , respectively.

In the calculation process, a' is regarded as the translation and b as the reference. In translation data, b as the target does not necessarily correspond to the source. When a does not correspond to b , the distance between a' and b is large. It is considered that this sentence pair is less parallel to a and b , so the sentence pair can be filtered. On the contrary, a small distance means that the similarity between the translation model and the actual reference is high and the parallelism between the source sentence and the target sentence is high. According to the editing distance, the parallelism score of the final sentence pair is shown in equation (8) as follows:

$$s_{a,b} = 1 - \frac{L_{a',b}(|a'|, |b|)}{\max(|a'|, |b|)}. \quad (8)$$

3.3.4. Cosine Similarity. Since the translation model M_0 can translate the English source sentence a into the corresponding Chinese translation a' , the semantic similarity between a' and b can be calculated only with the help of Chinese word vectors. The reason why two separate sets of Chinese and English word vectors are not used in this paper is that the language differences will cause the deviation of semantic space. This leads to inaccurate semantic similarity calculation. The data used for training the Chinese word vector is the same as the Chinese data in the machine translation training set. The training tool adopts the gensim (<https://radimrehurek.com/gensim/models/word2vec.html>) toolkit. The training window takes 5, and the words with word frequency lower than 5 are removed. Considering the high pressure of similarity calculation, the dimension is taken as 128 dimensions and trained for 10 rounds and the model is finally saved as M_1 .

For a and b sentence pairs, a' is the Chinese translation of a ; then, the parallelism score of the sentence pair can be obtained by using M_1 and cosine function, as shown in equation (9) as follows:

$$s_{a,b} = \cos(a', b|M_1). \quad (9)$$

3.3.5. Bilingual Pretraining Model. Considering that the pretraining model contains a lot of semantic knowledge, this paper uses the sentence-BERT (sense bidirectional encoder representations from transformation) model to fine tune the Chinese-English monolingual corpus given by CCMT2020; Chinese and English sentence vectors are obtained. However, the sentence vectors obtained in this way may have the problem of misalignment in the vector space between different languages, that is, sentences with the same meaning in different languages are mapped to different positions in the vector space. Therefore, when evaluating the parallelism between sentences in two different languages, this paper uses the ratio of the square of the Mahalanobis distance as the measurement index.

The Mahalanobis distance represents the covariance distance of data, which is an effective method to calculate the similarity of two unknown sample sets. Using the Mahalanobis distance is equivalent to the method of data conversion to eliminate the correlation and dimensional difference between different feature dimensions in the sample, so that the Euclidean distance can effectively measure the distance from the sample points to the distribution in the new distribution. Suppose that vector x represents μ . If the covariance matrix is a multivariable random vector of Σ , the Mahalanobis distance to the center is calculated as shown in (10) as follows:

$$d^2(x) = (x - \mu)^T \Sigma^{-1} (x - \mu) = \left\| \sum (x - \mu) \right\|_2^{-2}. \quad (10)$$

Here, $\Sigma^{-1/2}$ means finding the $-(1/2)$ power of each element in Σ .

In this system, firstly, each sentence vector is standardized so that it follows a random distribution with a mean value of 0. For each recentered Chinese-English sentence vector pair (l_1, l_2) , three cases in the change space are considered, such as formula (11) as follows:

$$\begin{aligned} e_1 &= \sum^{-1/2} [l_1, 0], \\ e_2 &= \sum^{-1/2} [0, l_2], \\ e &= \sum^{-1/2} [l_1, l_2], \end{aligned} \quad (11)$$

where e_1 , e_2 , and e represent the vectors of splicing vectors $[l_1, 0]$, $[0, l_2]$, and $[l_1, l_2]$, respectively, in the Markov space. Through the abovementioned three cases, the following ratio of the squares of Markov distances can be used to measure the parallelism between two language sentences, as shown in formula (12) as follows:

$$m = \frac{\|e\|_2^2}{\|e_1\|_2^2 + \|e_2\|_2^2}. \quad (12)$$

If two sentences have the same meaning, the possibility of the sentence to vector e in the Markov space should not be less than the probability of the vector of isolated single sentences e_1 and e_2 in the Markov space. The greater the value of m , the higher the parallelism between the two sentences.

Finally, normalize the m value and use equation (13) to measure the parallelism between two sentences:

$$m' = 1 - m. \quad (13)$$

That is, the smaller the m' , the higher the parallelism between the two sentences.

TABLE 1: BLEU value corresponding to each unit system.

System	BLEU
Stochastic system 0	16.59
Stochastic system 1	16.93
Rule	16.75
Zipporah	15.45
Word alignment model	17.04
Language model	16.92/16.38
Translation model	17.26/17.19
Bilingual pre-training model	17.18
Domain classifier	14.82

TABLE 2: CNN-based domain two classifier performance.

Category	p	R	F_1
News	82.8	99.0	90.2
Non-news	98.8	79.5	88.1

4. Results and Discussion

4.1. Data Processing. The development set, training set, and test set of the corpus filtering system in this paper are the Chinese-English news test set from WMT 2018 and WMT 2019 (including 3981 sentences and 2000 original texts and corresponding reference translations, respectively), the Chinese-English parallel corpus without translation in CCMT 2020 (9.02 million Chinese-English sentence pairs), and the parallel corpus with translation in CCMT 2020 (34.32 million Chinese-English sentence pairs).

Among them, the Jieba (<https://github.com/fxsjy/jieba>) word segmentation tool is used for Chinese corpus and Moses (<https://statmt.org/moses/>) script word segmentation and lowercase processing are used for English corpus. Because the amount of data is too large to prevent video memory overflow during decoding, the translation data after lowercase is truncated. At the same time, in order to alleviate the problem of out of vocabulary (OOV) and improve the processing ability of the model for rare words and OOV, this paper uses the method based on subword segmentation to segment Chinese corpus and English corpus with byte pair coding (BPE, <http://github.com/resennrich/subword-nmt>). In addition, in order to prevent the memory shortage and long decoding time caused by loading and decoding 34 million sentences at one time, this paper divides the translation data, each containing 2 million pieces of data. Finally, remove the sentences with a length of more than 150 words, and then, remove the sentences with language errors.

4.2. Evaluation Method. After the translation corpus is scored, it is sorted according to the score from high to low, so as to realize corpus filtering. This paper selects parallel sentence pairs containing about 100 million words, uses Marian, a neural machine translation tool designated by CCMT 2020 organizers, takes the previously selected parallel sentence pairs as a training set, trains them on Marian, and

then tests them on the test set designated by CCMT 2020 organizers. The Bleu index commonly used in the field of machine translation is used as the evaluation index to evaluate the quality of the filtered corpus.

The final contestant shall provide the CCMT 2020 organizer with two filtered corpora of 100 million words and 500 million words. The CCMT 2020 organizer will take the corpora submitted by the contestants as the training set, use the Marian tool for training, ensure that all parameters are consistent during the training process, and conduct the test on the specified test set as the final result of the contestant.

4.3. GA-MLP-NN Single-System Experiment. Because of the specific relationship between each language system and Zipporah system, the translation system can be selected as the basis of each language system. The scores of translation data are sorted from high to low according to each system. It should be noted that if some systems have multiple scores, each score is added or multiplied to obtain the comprehensive score and the weight is 1.0. Then, use Marian, the machine translation tool provided by CCMT 2020 to train the neural machine translation system. Calculate the BLEU value between the translation results on the development set and the reference translation. Select the dominant features according to the corresponding Bleu value of each system, and try to combine the dominant features to get a better ranking.

Due to the limitation of computing resources, this paper only trains 10 rounds for each system and takes the highest Bleu value in the development set as the final score of the system. Refer to Table 1 for the score of each system. Among them, the random system randomly disrupts the data and similarly samples the parallel corpus of 100 million words, the random system 0 randomly disrupts the data only once, and the random system 1 randomly disrupts the data five times. In addition, in order to explore the influence of the domain on performance, this paper collects 1409 Chinese news samples and 1434 Chinese non-news samples from nontranslation parallel corpus, divides 200 news and 200 non-news samples into development sets, and trains a domain two classifier based on the convolutional neural network (CNN). As can be seen in Table 1, the results of random system 1 even exceed most systems. The best one is the similarity index between translation and reference based on the translation model. The domain classifier is the worst because the domain classifier is mainly used to select news corpus, and the results show that the proportion of news corpus in the test set may not be high, resulting in poor performance. It is noted that the top sentences in the corpus filtered by the translation model are not very sensitive to the sentence length, so a large number of sentences with moderate length are expected to rank first. Although the rule system can treat long sentences and short sentences indiscriminately, it cannot measure the degree of parallelism, so the effect is not prominent when it plays a role alone.

The domain classifier is used for the test of translation data, and the prediction probability of news data is taken as the score. For the performance of the domain 2 classifier, refer to Table 2. It can be seen that the performance of this

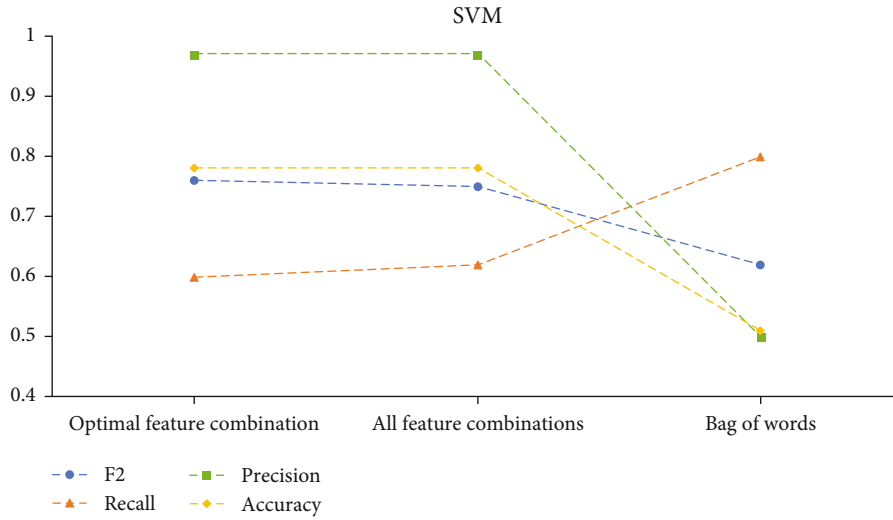


FIGURE 3: SVM result graph.

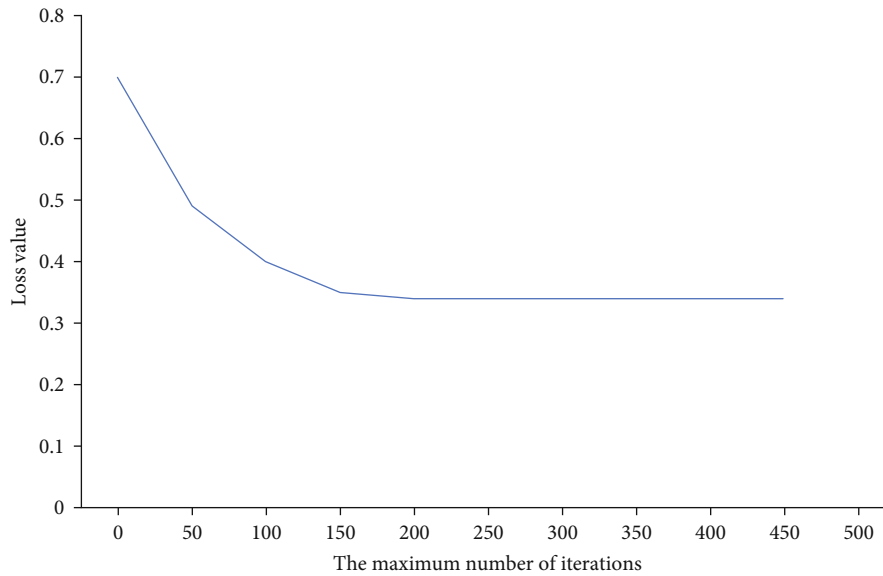


FIGURE 4: Schematic diagram of the number of iterations of GA-MLP-NN.

TABLE 3: Results of partial system integration.

System combination	BLEU	
	Addition	Multiplication
1,3	17.53	17.60
1,2,3	17.37	17.56
1,3,4	17.66	17.92
1,2,3,4	17.18	17.65
1,2,3,4,5	17.38	17.13
1,2,3,4,6	17.30	17.40

classifier is high, but in Table 1, the translation performance based on this classifier is very low, so it can be considered that in this task, the domain has little influence on the translation model. Therefore, the classifier is only used for verification, and it is not included in the final system in this paper.

4.4. Parameter Setting of the GA-MLP-NN Model. In the parameter setting of the GA MLP NN model, the process of superparameter optimization is as follows.

- (1) Select a set of superparameters (automatic selection)
- (2) Build the corresponding model
- (3) Fit the model on the training data and measure its final performance on the validation data
- (4) Select the next set of superparameters to try (auto select)
- (5) Repeat the abovementioned process
- (6) Finally, the performance of the model on the test data is measured

TABLE 4: Final evaluation results.

System	All test sets	WMT2020 news	BLEU			
			WMT2020 biomedical	IWSLT2020	CCMT2020	CCMT2019
1	16.8	15.3	9.1	17.7	13.6	18.4
2	16.0	15.8	9.3	16.1	13.9	19.0
3	14.2	14.3	6.8	14.3	12.9	17.5
4	14.2	14.2	6.8	14.3	13.1	17.1
5	13.6	15.7	8.3	12.0	12.8	18.8

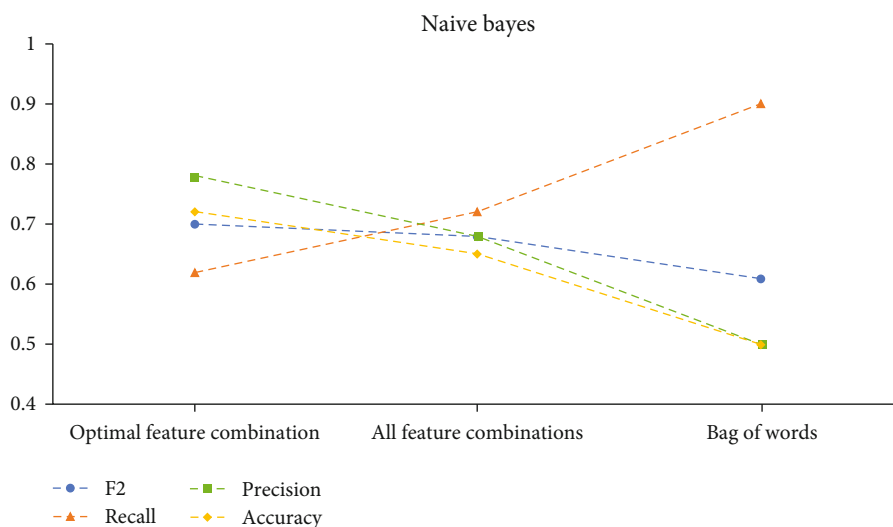


FIGURE 5: Naive Bayes result graph.

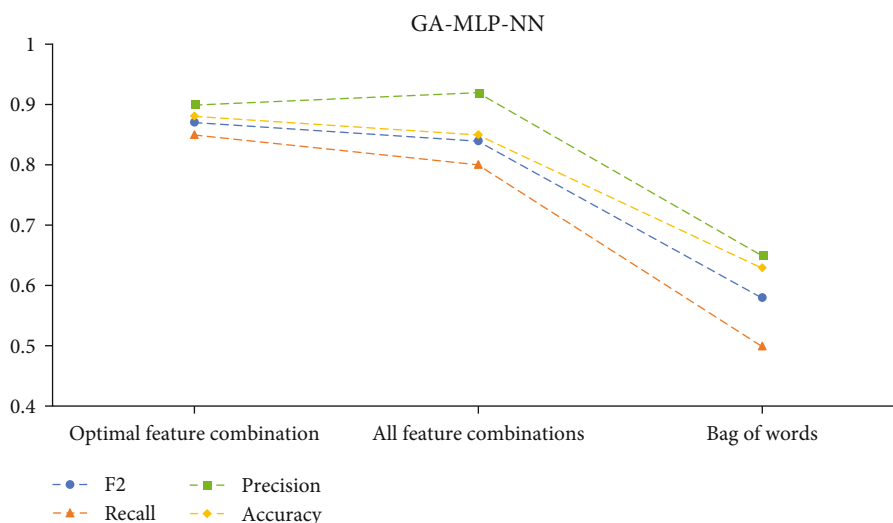


FIGURE 6: The result graph of the GA-MLP-NN combined model.

The key of this process is that given many sets of superparameters, the history of verification performance is used to select the next set of superparameters to be evaluated. It is necessary to systematically and automatically explore the possible decision space. According to experience, the recognition effect of the GA MLP NN classifier with the best performance is found.

The superparameter setting has an important impact on the recognition effect of the GA-MLP-NN classifier. Therefore, this paper has carried out many experiments on the super parameter setting of each group. Taking the number of iterations as an example, considering that the smaller the loss value of the model, the better the robustness of the model, the number of iterations of the model can be set through the loss value of the model. Specifically, the

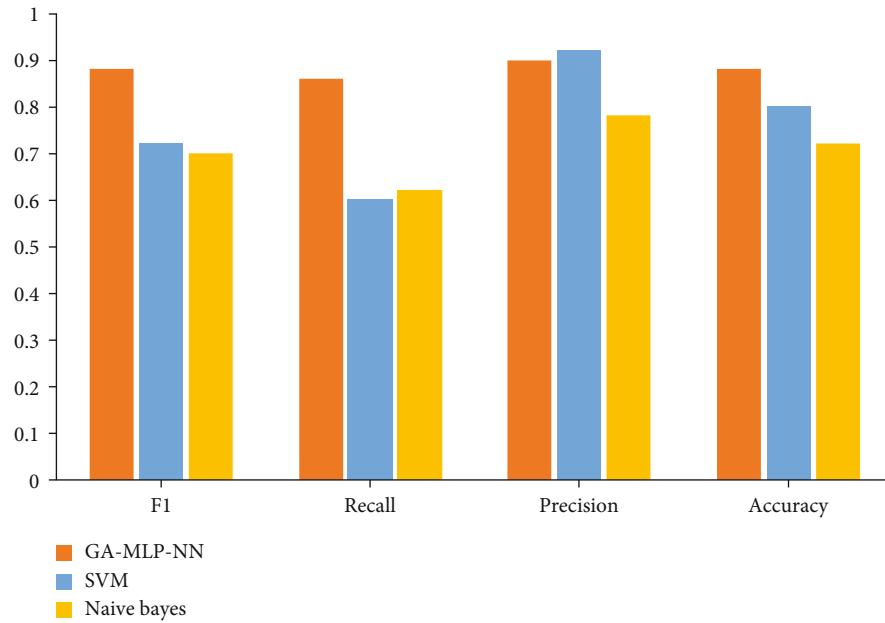


FIGURE 7: Comparison of the feature results of GA-MLP-NN and other models.

maximum number of iterations is set to 1, 50100, ..., 500 and the maximum number of iterations for loss value convergence is selected according to its corresponding loss value. As shown in Figure 3, the loss value does not decrease after the maximum number of iterations is greater than or equal to 300 and is stable at about 0.3. Therefore, the maximum number of iterations of the GA-MLP-NN model is finally set to 300, as shown in Figure 4.

4.5. GA-MLP-NN System Experiment after Fusion. In this paper, it is considered that the translation model system, genetic word alignment model system, language model system, and bilingual pre training model system are relatively potential systems. Therefore, the combination of these systems is given priority to the fusion test, that is, to form the GA-MLP-NN combination model. The method of multisystem fusion is relatively simple, that is, the scoring of each system is fused. Then, reorder. There are two fusion methods: multiply by weight and add by weight. In most cases, only the fusion with the weight of 1.0 is tried. Table 3 shows some experimental results. It can be seen that the overall performance of the fusion system exceeds that of the single system. The main reason for the better performance of the fusion system is that different systems measure the parallelism of sentence pairs from different starting points, so the sentence pairs can be evaluated more comprehensively after multisystem fusion, which also shows the effectiveness of method fusion.

4.6. Submission System. It is not found that the more integrated the systems, the better performance. After a large number of tests, it is found that the robustness and BLEU value of “1,3,4” combination are high. Considering the complexity of the system, this paper selects “1,3,4” combination as the main system, because the multilayer perceptron

method has been proved to be an effective means to improve the translation performance in WMT 2018 and WMT 2019 corpus filtering tasks. In addition, the pre training model has advantages in semantic extraction, so the combination of “1,2,3,4,6” is selected as the subsystem. The final evaluation results are shown in Table 4. It can be seen that the main system system2 submitted in this paper ranks second and first except the iwslt2020 dataset. Because the IWSLT2020 dataset is spoken language corpus, there are some differences between news corpus and spoken language corpus. This leads to the poor performance of the system on IWSLT2020 dataset, which also indicates that the training field will affect the filtering results.

4.7. Comparison between GA-MLP-NN and Classical Methods. In this paper, the recognition results of the GA MLP NN model are compared with those of classical methods such as naive Bayes, SVM, and support vector machine. The $f1f1$ value, recall, precision, and accuracy of the experimental results are analyzed and compared. The better the parameter value is, the higher the translation effect is. The results show that GA MLP NN model has good filtering performance for high-quality parallel corpus.

In order to further verify the recognition effect of the proposed method, the recognition results of the proposed method are compared with those of classical methods such as naive Bayes, SVM, and support vector machine. That is, the optimal feature combination (N_C , R_{RW} , N_F , and R_{SW}) and all features (N_C , R_{RW} , N_F , R_{SW} , N_G , and N_S) obtained based on the GA theory are used as feature representation, and the feature representation obtained by the word bag method is used to train the SVM classifier, naive Bayesian classic classifier model, and GA-MLP-NN neural network classifier. The $F1F1$ value, recall, precision, and accuracy of the experimental results are shown in Figures 5–6.

As can be seen in Figure 6, the $F1F1$ value, precision, and accuracy of the three classifiers trained by the feature representation based on the word bag method are lower than the recognition results of the classifiers trained by the other two feature representation methods based on the GA theory. The values of the three classifiers based on the optimal feature combination are better than those based on all feature combination and word bag method. Among them, the $F1F1$ value of the GA-MLP-NN classifier based on optimal feature combination is 3% higher than that based on all feature combinations and 33% higher than that based on the word bag method. The $F1F1$ value of the SVM classifier based on optimal feature combination is 2% higher than that based on all feature combinations and 11% higher than that based on the word bag method. The $F1F1$ value of naive Bayes based on the optimal feature combination is 1% higher than that based on all feature combinations, and 7% higher than that based on the word bag method. It can be seen that the selection of the optimal feature combination significantly improves the classification effect of the classifier and further verifies the feasibility and superiority of the method proposed in this paper.

As can be seen in Figure 7, GA-MLP-NN is much better than other classifiers in $f1f1$ value, recall, and accuracy. For the optimal feature combination of N_C , R_{RW} , N_F , and R_{SW} , the $F1F1$ value of the SVM classifier is 73.3% and the $F1F1$ value of the GA-MLP-NN classifier is 13.5% higher than that of the SVM classifier, reaching 86.8%. For the optimal feature combination of N_C , R_{RW} , N_F , and R_{SW} , the $f1f1$ value of the naive Bayesian classifier is 70% and the $f1f1$ value of the GA-MLP-NN classifier is 16.8% higher than that of the naive Bayesian classifier, which fully reflects the superiority of the GA-MLP-NN classifier trained by feature combination proposed in this paper.

5. Conclusion

In this paper, the ga-mlp neural network is used as a classifier for comment usefulness recognition. This paper creatively analyzes the impact of the proposed features on the usefulness of comments, measures the parallelism of Chinese and English sentences from different emphases, and obtains the best feature combination. Finally, the experimental results show that the $f1f1$ value of naive Bayes based on the optimal feature combination is 1% higher than that based on all feature combinations and 7% higher than that based on the word bag method. It can be seen that the selection of the optimal feature combination significantly improves the classification effect of the classifier and further verifies the feasibility and superiority of the method proposed in this paper. Compared with a single system, the improved multisystem fusion method based on weight multiplication has achieved better results in the test set. The application of the perceptron genetic algorithm neural network in Chinese-English two-way translation is solved. The system presented in this paper ranks second and first in multiple datasets, which has a certain reference value for corpus filtering research. However, this paper does not carry out the actual simulation verification of the adopted model and

there are still some limitations. Further research is needed in the future.

Data Availability

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

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This study is supported by the Hainan Provincial Social Science Project of Foreign Language Base: a study on the construction of Hainan's local translation talent team under the background of new liberal arts (HNSK(JD)21-41).

References

- [1] J. Wei, Z. Li, D. Yi, J. Jing, and W. Gang, "Analysis of the usefulness of online comments for user demand acquisition," *Chinese Journal of Computers*, vol. 36, no. 1, pp. 119–131, 2013.
- [2] S. Shin, N. Chung, Z. Xiang, and C. Koo, "Assessing the impact of textual content concreteness on helpfulness in online travel reviews," *Journal of Travel Research*, vol. 58, no. 4, pp. 579–593, 2019.
- [3] X. Liu, G. A. Wang, W. Fan, and Z. Zhang, "Finding useful solutions in online knowledge communities: a theory-driven design and multilevel analysis," *Information Systems Research*, vol. 31, no. 3, pp. 731–752, 2020.
- [4] H. Yuanyuan, Y. Qiang, and L. Yijun, "Research on the influencing factors of the usefulness of online reviews based on film review data," *Journal of Management Science*, vol. 13, no. 8, pp. 78–88, 2010.
- [5] S. M. Mudambi and D. Schuff, "Research note: what makes a helpful online review? A study of customer reviews on Amazon.com," *MIS Quarterly*, vol. 34, no. 1, pp. 185–200, 2010.
- [6] H. Baek, J. Ahn, and Y. Choi, "Helpfulness of online consumer reviews: readers' objectives and review cues," *International Journal of Electronic Commerce*, vol. 17, no. 2, pp. 99–126, 2012.
- [7] Y. Chen and J. Xie, "Online consumer review: word-of-mouth as a new element of marketing communication mix," *Management Science*, vol. 54, no. 3, pp. 477–491, 2008.
- [8] L. Wei and X. Pengtao, "Research on the identification of factors affecting the usefulness of online reviews of O2O e-commerce platforms: taking the O2O model of the catering industry as an example," *China Management Science*, vol. 24, no. 5, pp. 168–176, 2016.
- [9] Z. Yanhui, L. Zongwei, and Z. Yicheng, "The impact of information quality based on Taobao comment data on the usefulness of online comments," *Chinese Journal of Management*, vol. 14, no. 1, pp. 77–85, 2017.
- [10] M. Rui and X. Jian, "The impact of scoring inconsistency on the usefulness of online reviews—from the perspective of attribution theory," *China Management Science*, vol. 26, no. 5, pp. 178–186, 2018.

- [11] Z. W. Liu and S. Park, "What makes a useful online review? Implication for travel product websites," *Tourism Management*, vol. 47, pp. 140–151, 2015.
- [12] A. H. Huang, K. Chen, D. C. Yen, and T. P. Tran, "A study of factors that contribute to online review helpfulness," *Computers in Human Behavior*, vol. 48, pp. 17–27, 2015.
- [13] Y. Guopeng, "What kind of online reviews do consumers think is more useful?—the influence of social factors," *Management World*, vol. 12, pp. 115–124, 2012.
- [14] Z. Liye, Y. Denghua, and Z. Jingyi, "The influence of online user review quality and reviewer rating on consumer purchase intention: the moderating effect of product involvement," *Management Review*, vol. 29, no. 2, pp. 87–96, 2017.
- [15] A. Ghose and P. G. Ipeirotis, "Estimating the helpfulness and economic impact of product reviews: mining text and reviewer characteristics," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 10, pp. 1498–1512, 2011.
- [16] C. Shuqin, Q. Zhiyong, L. Cuiping, Y. Qian, and Z. Bipan, "Research on the influence of emotional intensity of negative online reviews on usefulness," *Management Review*, vol. 29, no. 2, pp. 79–86, 2017.
- [17] B. Fang, Q. Ye, D. Kucukusta, and R. Law, "Analysis of the perceived value of online tourism reviews: Influence of readability and reviewer characteristics," *Tourism Management*, vol. 52, pp. 498–506, 2016.
- [18] Y. Jianyuan, L. Yang, F. Miao, and L. Kai, "The interactive influence of user Q & A and online comments on consumer product attitudes," *Management Science*, vol. 33, no. 2, pp. 102–113, 2020.
- [19] W. Chen, J. Zhu, M. Zhu, and T. Yao, "Text representation using domain dictionary," *Computer Research and Development*, vol. 42, no. 12, pp. 2155–2160, 2005.
- [20] S. W. Sussman and W. S. Siegal, "Informational influence in organizations: an integrated approach to knowledge adoption," *Information Systems Research*, vol. 14, no. 1, pp. 47–65, 2003.
- [21] C. Cheung, M. Lee, and N. Rabjohn, "The impact of electronic word-of-mouth," *Internet Research*, vol. 18, no. 3, pp. 229–247, 2008.
- [22] Z. S. Chen, X. L. Liu, K. S. Chin, W. Pedrycz, K. L. Tsui, and M. J. Skibniewski, "Online-review analysis based large-scale group decision-making for determining passenger demands and evaluating passenger satisfaction: Case study of high-speed rail system in China," *Information Fusion*, vol. 69, pp. 22–39, 2021.
- [23] I. Erkan and C. Evans, "The influence of eWOM in social media on consumers' purchase intentions: an extended approach to information adoption," *Computers in Human Behavior*, vol. 61, no. AUG., pp. 47–55, 2016.
- [24] J. Chunhua and G. Feipeng, "Distributed data mining model DSVM based on support vector machine," *Systems Engineering - Theory & Practice*, vol. 30, no. 10, pp. 1855–1863, 2010.
- [25] X. Binqing, Y. Yang, L. Xindan, and L. Haohua, "Research on the credit rating of small and micro enterprises based on fuzzy neural network," *Journal of Management Science*, vol. 19, no. 11, pp. 114–126, 2016.