Platform Design of Project Bidding System on the Basic of TCM-KNN and Genetic Algorithm

Lei Sun

{13351675317@163.com}

School of Environmental Arts and Architectural Engineering, Heilongjiang University of Technology, Jixi, Heilongjiang, 158100, China

Abstract. Project bidding is a widely used method of project transactions in the world. It promotes fairness and justice in the field of project construction, and also prevents and overcomes corruption in the field of project construction. Nowadays, electronic bidding has gradually replaced traditional bidding methods, and the big date (BD) brought by electronic bidding has also provided sufficient convenience for the establishment and improvement of engineering enterprise credit evaluation systems. Therefore, this paper is to solve the problem of insufficient application of BD in the current engineering electronic bidding. This paper combines the TCM-KNN classification algorithm with the genetic algorithm(GA), collects data information related to bidding transactions from multiple channels, and conducts research on the internal correlation and mutual influence between these data information, so as to determine the evaluation indicators suitable for engineering enterprise credit evaluation. And on the basic of the BD analysis of engineering electronic bidding, this paper establishes the corresponding engineering enterprise credit evaluation(ECE) index system and ECE model. Finally, the experimental results show that the GA is used to optimize the TCM-KNN model, and the parameters that are more in line with the actual situation are obtained.

Keywords: Engineering Bidding System, TCM-KNN, Genetic Algorithm, Enterprise Credit Evaluation

1. Introduction

1.1 Subject Research Background

With the increasing number of project bidding methods, the scope and scale of bidding are also gradually increasing, and some problems brought about by traditional bidding methods have become more and more prominent, such as time-consuming, laborious, low efficiency, etc., this issue has received widespread attention in various industries. In the age of information and big data, information management of engineering project bidding is the best choice for bidding in this industry. However, due to the lack of unified technology and standards, although many regions have built bidding network platforms, information sharing and intercommunication between regions cannot be realized [1-2]. Under such conditions, there will be many risk factors in the

ICSETPSD 2023, November 17-18, Coimbatore, India

DOI 10.4108/eai.17-11-2023.2342782

implementation process of electronic bidding, and these existing or potential risks will have an impact on the smooth implementation of electronic bidding for projects. Therefore, we have to look for and take corresponding measures to avoid it.

1.2 Significance of the Subject Research

At present, a large amount of transaction data generated by engineering units during electronic bidding has not been fully mined and utilized [3]. Therefore, on the basic of TCM-KNN (Transductive Confidence Machines for K Nearest Neighbors) and GA, this paper analyzes the data of engineering bidding. It is to collect, organize and analyze all kinds of information generated in the process of electronic bidding in the engineering industry, and present these information to users in a visual form. In this way, the utilization of these data can be improved, and it also helps to build a more perfect credit evaluation system.

In addition, in terms of practical application, compared with the traditional project bidding method, the method of using data mining technology for project bidding has obvious advantages such as high quality and rapidity. Project bidding personnel use data mining methods to analyze project bidding companies, which not only saves the cost of project bidding, but also realizes the specific operability of concealed information collection and data analysis through BD technology, thereby improving the quality and efficiency of project bidding.

2. Relevant Theoretical Concepts and Literature Review

2.1 Concept of Project Bidding and Electronic Bidding

Engineering project bidding refers to a kind of economic competition behavior in which the tenderer uses bidding methods to attract other units to participate in the bidding for the project to be constructed, and in accordance with legal procedures and established requirements, and selects the best unit among them to start the construction and implementation of the project [4-5]. At present, there are three main methods of bidding activities for engineering projects, one is the way of public bidding, the other is the way of invitation to bid, and the third is the way of project negotiation.

Electronic bidding refers to the process of using data messages as the main carrier, using electronic information technology, and using electronic bidding systems to complete all or part of the bidding procedures in accordance with laws, regulations and relevant regulations [6]. In the electronic platform, there are various intertwined bidding traders, and each person participating in the electronic bidding plays a different role in the electronic data platform. In the entire transaction process of electronic bidding, it is necessary to meet the requirements of resource sharing, resource handover, information acquisition, and information integration.

2.2 Literature Review

2.2.1. TCM-KNN algorithm

TCM-KNN is an algorithm that combines the classic classification algorithm K nearest neighbor method with TCM, uses the method of distance calculation, and classifies the confidence of observation points according to the classified data set. At present, there are many studies on this classic classification algorithm at home and abroad. Zhang Y R proposed an improved binary K-means algorithm for constructing a normal behavior feature training set model, and combined the transduction reliability mechanism and the K-nearest neighbor algorithm to design the ITCM-KNN algorithm. This algorithm is suitable for network anomaly detection in a big data environment and has good adaptability [7]. Dai L proposed a new method for predicting disease syndrome, location and nature on the basic of TCM diagnostic information. The algorithm can be combined into a new unified feature space according to the correlation between features and the correlation between class labels. Then, he used the MDMR (Multidimensional Mixed Reality) algorithm to select the most discriminative features from the new unified feature space, which helps to reduce the data dimensionality [8]. However, the TCM-KNN algorithm has problems such as a large amount of training and a large number of features, so it needs to be optimized and improved.

2.2.2. Application of big data technology in project bidding

In the field of engineering bidding, there are many studies using BD to evaluate bidding projects. Araújo provided a model for assessing the attractiveness level of a project and deciding whether to submit a proposal. This model introduced a classification model containing fuzzy preference information to evaluate item attractiveness in multiple dimensions, enabling decision makers to make more effective decisions. The main advantages of the model are that the level of attractiveness can be used as an input to a bidding strategy, attractiveness assessment can support contractors in prioritizing bids for projects [9]. Li H proposed a discriminating method for unbalanced bidding, which used the combination of entropy weight theory and hierarchical analysis to discriminate the unbalanced bidding, combined it with subjective and objective factors, and weighted it by AHP model. After research, he found that this model can be used used the fuzzy comprehensive evaluation method to test unbalanced bids, so that owners can identify unbalanced bids in the procurement stage [10]. Peng W proposed an optimal bidding scheme selection algorithm on the basic of big data, and all bidding schemes are gathered together. Through big data sampling, the big data processing process of the bidding scheme was accelerated. He chose the Single method to cluster the big data sampling results, so as to determine the centroid direction of the natural clustering of big data. He used the mean update method to modify the centroid direction of the natural clustering of BD, so as to determine the actual centroid direction of the natural clustering of BD [11].

On the basic of the research of the above-mentioned scholars, most of them focus on the decision-making field of project bidding, and there is little content on the credit evaluation of enterprises. Therefore, this paper will use GA and TCM-KNN algorithm to conduct research on enterprise credit evaluation on the basic of big data analysis of engineering electronic bidding, and construct a set of enterprise credit evaluation system suitable for engineering industry.

3. Enterprise Credit Evaluation Model on the Basic of TCM-KNN

3.1 TCM-KNN Algorithm

In the TCM algorithm, the object it studies is a sample sequence $A = \{(m_1, n_1), \ldots, (m_x, n_x), (m_{x+1}, n_{x+1})\} = \{a_1, a_2, \ldots, a_{x+1}\}$ of length x+1. In this sequence, $m_i \in M$ is used as the input attribute vector of sample a_i, and M represents the input vector space; $n_i \in N$ represents the class or label of the sample. The first x samples in the sample sequence A are n-type known training samples, and the x+1th sample is an unknown type of test sample. TCM is a predictive method given a p-value for an unknown point x in a given class n. The p-value is used to measure the extent to which the data in class n support an unknown point x. The smaller the value of p is, the less likely it is to be in class n.

However, in the TCM-KNN algorithm, it is first necessary to define a singular description function that can output singular values, and then to calculate the *p*-value of the tested samples.

I. The sample a_i to be tested, with respect to the singular value b_{in} of the category *n*, is defined as:

$$b_{in} = \frac{\sum_{e=1}^{k} F_{ie}^{n}}{\sum_{e=1}^{k} F_{ie}^{-n}} \tag{1}$$

In formula (1), F_i^nrepresents the distance sequence between sample a_i and all samples in category *n*, and F_{ie}^n represents the *e*-th shortest distance in the sequence; F_i^{-n} represents the distance sequence between sample a_i and all samples in other categories (except n), and F_{ie}^{-n} also represents the *e*-th shortest distance in the sequence. The parameter k represents the number of nearest neighbors we need to consider. Through this definition, it is not difficult to see that the singular value is designed on the basic of the distance of the sample feature vector in the feature space.

II. The distance between samples is calculated using the Euclidean distance, and its formula is as follows:

$$dist(M_1, M_2) = \sqrt{\sum_{e=1}^{|M_i|} (M_{1e} - M_2 e)^2} \tag{2}$$

In formula (2), M_1 and M_2 represent two samples (represented by the input attribute vector of the sample), M_{ie}represents the *e*-th dimension feature of vector M_i, and $|M_i|$ represents the number of attributes of vector M_i.

III. The definition of the *p* value of the sample a_i to be tested relative to a certain category is:

$$p(a_i) = \frac{\#\{e : \beta_e \geq \beta_i\}}{x+1} \tag{3}$$

Among them, # represents the "potential" of the set, which is usually calculated as the number of elements in a finite set; β_i is the singular value of the sample to be tested; *x* is the number of sets; β_e represents the singular value of any sample in the set.

3.2 Model Establishment

a. Data processing

The data used for modeling and analysis in this paper come from the official websites of various enterprises, public platforms for electronic bidding, etc. A total of 100 samples of project bidding data were received, including the credit evaluation research of 100 enterprises. 50 pieces of data are used as training samples for network training, and the remaining 50 pieces of data are used as test samples to verify the validity and reliability of the model. The simulation tool uses MATLAB, and normalizes the data through the mapminmax(.) function.

b. Enterprise credit evaluation index

At present, the commonly used rating symbols in the world, the third grade and nine grades of debt have been used for a long time, but due to actual work, enterprises with poor credit ratings rarely apply for credit ratings [12-13]. Therefore, the grades after the fifth grade in the third grade and nine grades lack practical significance. The credit ratings adopted by most evaluation agencies are shown in Table 1.

The engineering enterprise credit evaluation index system can be divided into the following indicators: First, the six first-level indicators include enterprise quality, enterprise operation ability, enterprise profitability, enterprise social credit status, enterprise solvency, and enterprise growth ability. Second, the 19 secondary indicators include the quality of the enterprise itself, the quality of the leadership of the enterprise, the inventory turnover rate, the accounts receivable turnover rate, the operating net cash flow interest coverage ratio, the operating profit margin, the return on total assets, and the net assets Yield, debt performance, contract performance, other credit conditions of the enterprise, and so on [14].

Table 1. Commonly used credit levels and their meanings

Credit rating symbol	Credit rating	Meaning of credit rating
AAA	First class	Good credit rating
AA	Second class	Good credit
A	Third class	Good credit level
B	Fourth class	Credit level - average
C	Fifth class	Bad credit

3.3 Experimental Results and Analysis

There are 100 experimental data in this paper, of which 45 are AAA-level enterprises; 25 are AA-level enterprises; 10 are A-level enterprises; 15 are B-level enterprises; 5 are C-level enterprises. Using all sample data to participate in the experiment, and evaluate the experimental results of different neighbor numbers k, as shown in Figure 1:

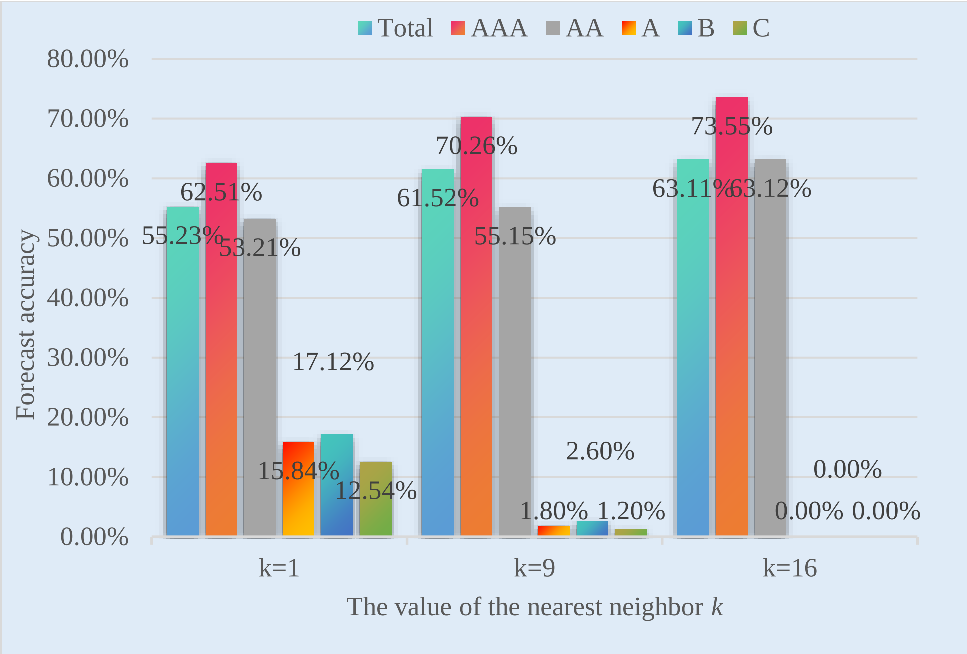

Fig.1 Dataset prediction accuracy on the basic of TCM-KNN algorithm

The figure 1 shows the prediction accuracy of the experimental data increases with the increase of k value. Since there are few experimental data of A, B, and C levels, they need to be excluded. Overall, the prediction accuracy is not high, the highest is only 63.11%. This shows that it is difficult for the singularity description function in the TCM-KNN method to completely describe the singularity of this type of data, so it is difficult to distinguish this type of data. However, there is not much data available in this paper, so we only need to perform confidence evaluation after selecting important features for this method.

4. Optimizing TCM-KNN Enterprise Credit Evaluation Model on the Basic of Genetic Algorithm

4.1 Feature Extraction Technology on the Basic of Genetic Algorithm

The feature extraction on the basic of the GA needs to be on the basic of the text, not the feature extraction algorithm on the basic of the entire text set. Therefore, when extracting text features, the feature words in the same text can be put into a feature vector representing the text, so as to avoid ignoring the connection between feature items [15-16]. On this basis, this paper proposes a text feature vector on the basic of X^2 statistics, which can not only preserve the correlation between text features, but also distinguish the correlation between features and classes; and uses this vector as the initial population, through multiple rounds of genetic screening, high-quality feature vectors are obtained to improve classification accuracy; through the coordination of crossover operation and mutation operation, global search can be realized and local minima can be avoided[17-18]; according to the characteristics of feature extraction, the fitness function and intersection rules are designed to solve the problem of inappropriate processing of low-frequency words in statistical analysis [19-20]. The flow chart of feature extraction on the basic of GA is shown in Figure 2:

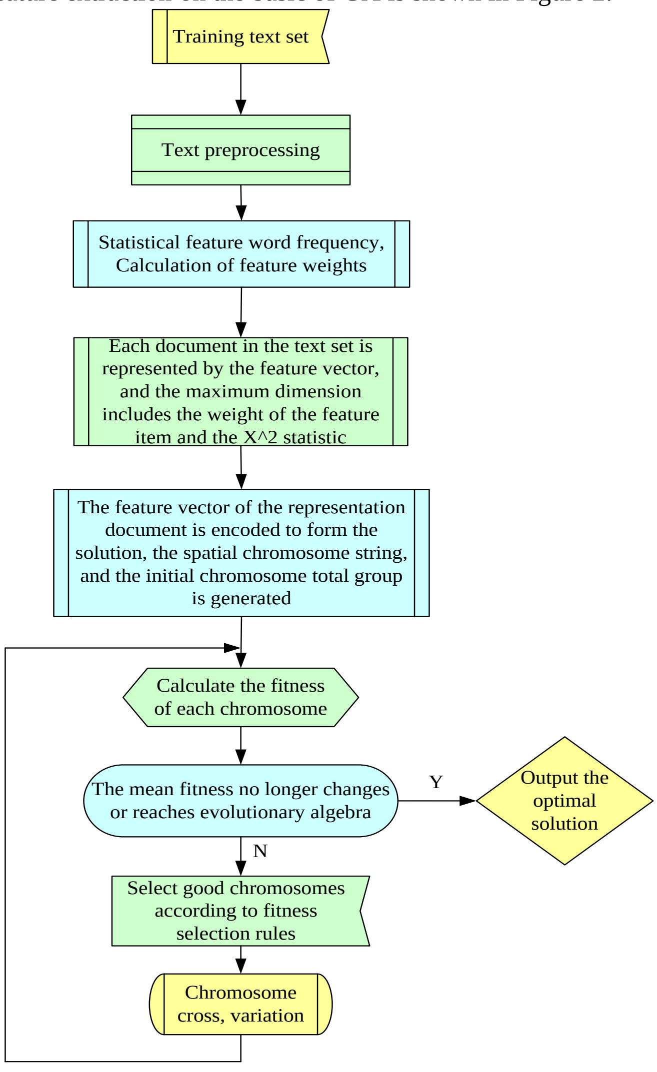

Fig.2 Flow chart of feature extraction on the basic of genetic algorithm

4.2 Experimental Results and Analysis

Calling the main.m function in the Matlab software, the main function of this function is to carry out large-scale training to the GA+TCM-KNN model, the average fitness curve and optimal fitness curve of the model after training are as shown in Figure 3:

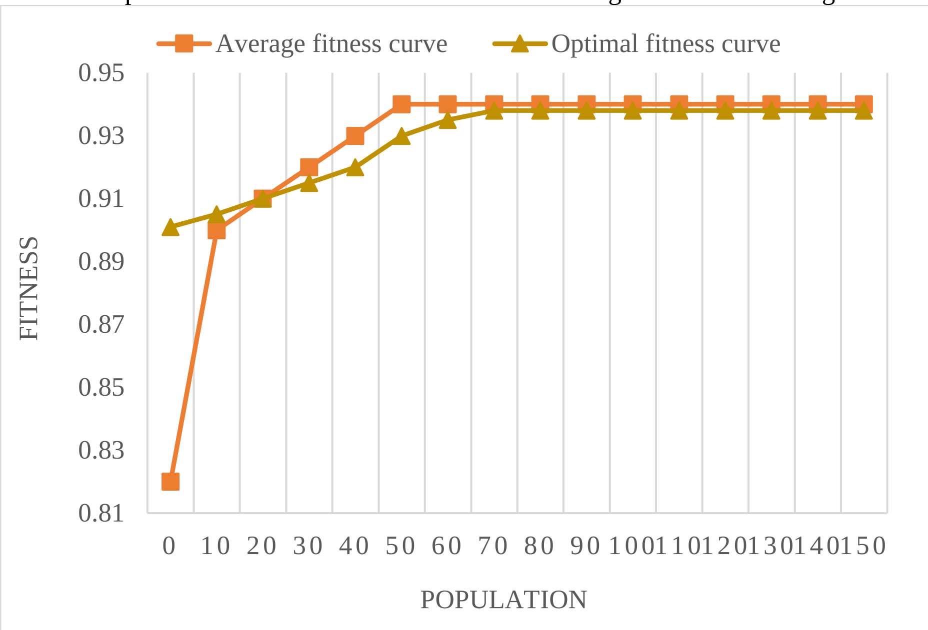

Fig.3 Average fitness curve and optimal fitness curve

The figure 3 shows the average fitness of the population optimized by the GA has reached above 0.935, and these data show that the fitness of individuals in the population is better, and the effect of evolution is better.

So as to further test the effectiveness of the optimized GA+TCM-KNN model, this experiment uses training sample data and test sample data to compare the GA+TCM-KNN model with the TCM-KNN model. The comparison results are shown in Figure 4.

In Figure 4, the perfect match rate of the training samples in the TCM-KNN model is 60%, and the error in the operating range is 40%; the perfect fit rate in the GA+TCM-KNN model is 50%, and the error within the operating range is 50%. In the training samples, the fitting effect of TCM-KNN is better than that of GA+TCM-KNN. But for the test sample, the prediction accuracy of the TCM-KNN model is only 55%, indicating that its prediction accuracy is not high enough. However, the prediction accuracy of the GA+TCM-KNN model for the sample can reach 81%, which has a good test effect. These data show that the TCM-KNN model is optimized by GA, and the parameters that are more in line with the actual situation are obtained.

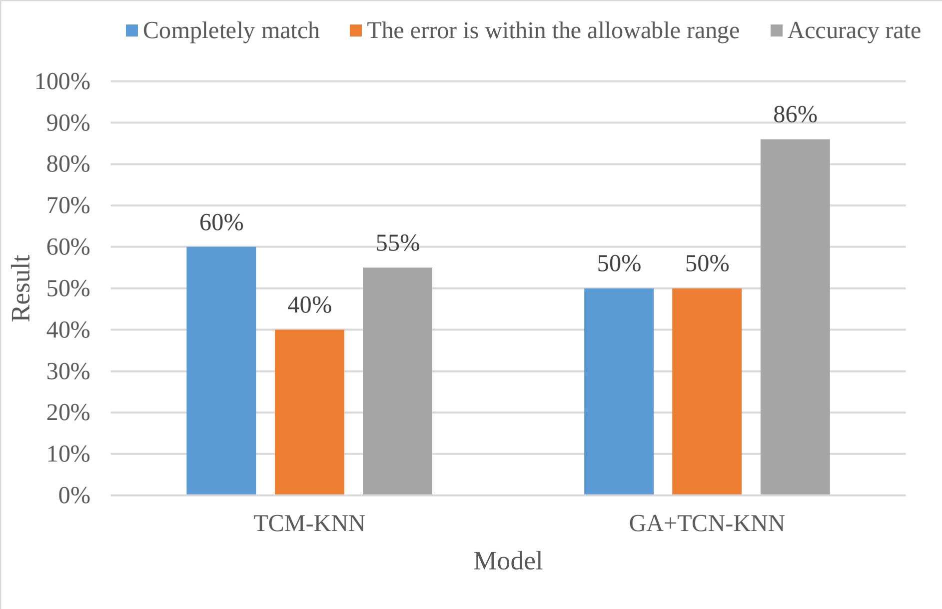

Fig.4 The results of model comparison and analysis before and after optimization

5. Conclusion

In this paper, combined with the project bidding example, the construction of TCM-KNN and GA engineering bidding evaluation model is verified, analyzed and applied.Through the comparison and analysis of the model reasoning and prediction results with the actual project, it is found that the model constructed in this text is feasible and applicable, and it can assist the project to actually evaluate and manage project bidding enterprises, and provide scientific basis and risk data for risk control.

Acknowledgement

This work was supported by the Fundamental Research Funds for the Provincial Universities of Heilongjiang Province (2022-KYYWF-0491)

References

[1] You-Ping F U , Yang-Bo Y U .Discussion on the Compilation of International EPC Engineering Bidding Quotation[J].Value Engineering, 2019,38(20):215-217.
[2] Wu M .Analysis of Existing Problems with Management of Construction Projects Bidding and Solutions to the Problems[J].Architectural development study, 2021, 005(002):P.30-33.
[3] Choi S J , Choi S W , Kim J H ,et al.AI and Text-Mining Applications for Analyzing

Contractor's Risk in Invitation to Bid (ITB) and Contracts for Engineering Procurement and Construction (EPC) Projects[J].Energies, 2021,14(15):1-28.
[4] Tso R, Liu Z Y, Hsiao J H. Distributed E-voting and E-bidding systems based on smart contract[J]. Electronics, 2019, 8(4): 1-22.
[5] Ahmed M O , El-Adaway I H , Coatney K T .Solving the Negative Earnings Dilemma of Multistage Bidding in Public Construction and Infrastructure Projects: A Game Theory-Based Approach[J].Journal of management in engineering, 2022,38(2):4021087.1-4021087.19.
[6] Bo X , Bing C .Application analysis of electronic bidding system in highway engineering project[J].Shanxi Architecture, 2019,45(14):179-180.
[7] Zhang Y R. Research on Network anomaly detection Technology based on improved dichotomous K-means algorithm [J]. Journal of Ezhou University,202,29(6):97-99.
[8] Dai L , Zhang J , Li C ,et al.Multi-label feature selection with application to TCMstate identification[J].CONCURRENCY PRACTICE & EXPERIENCE, 2019, 31(23):e4634.1-e4634.13.
[9] Araújo, Maria Creuza Borges, Alencar L H , Miranda Mota C M .Classification model for bid/no-bid decision in construction projects[J].International Transactions in Operational Research, 2022,29(2):1025-1047.
[10] Li H , Su L , Xia Q .Detecting unbalanced bidding to achieve economic sustainability using fuzzy logic approach[J].Construction Innovation, 2020,21(2):164-181.
[11] Peng W, Shuijiang P I .Optimal bidding scheme selection algorithm based on big data[J].Modern Electronics Technique,2019,42(04):105-108.
[12] Wang F , Ding L , Yu H ,et al.Big data analytics on enterprise credit risk evaluation of e-Business platform[J].Information systems and e-business management: special issue on emerging technologies for e-business engineering, 2020,18(3):311-350.
[13] Zhang F , Xu Z .Research on the Evaluation Index System of Enterprise's Online Credit in B2C E-commerce[J].International Journal of Sciences, 2019, 8(06):117-121.
[14] Gaosheng Y , Xiaojing Z , Business S O ,et al.Time Sensitivity and Evaluation Cycle of Long Term Engineering Credit Evaluation Index[J].Science and Technology Management Research, 2019,39(03):63-70.
[15] Eliguzel N , Cetinkaya C , Dereli T .A novel approach for text categorization by applying hybrid genetic bat algorithm through feature extraction and feature selection methods[J].Expert Systems with Application, 2022,202(Sep.):117433.1-117433.12.
[16] Priya R D , Sivaraj R , Anitha N ,et al.Forward feature extraction from imbalanced microarray datasets using wrapper based incremental genetic algorithm[J].International Journal of Bio-Inspired Computation, 2020, 16(3):1711-180.
[17] Alagumariappan P , Krishnamurthy K , Kandiah S ,et al.Feature Extraction and Genetic Algorithm based Feature Selection for Diagnosis of Type-2 Diabetes using Electrogastrograms (Preprint)[J].JMIR Biomedical Engineering, 2020, 5(1):1-11.
[18] Lee C Y , Le T A .Optimised approach of feature selection based on genetic and binary state transition algorithm in the classification of bearing fault in BLDC motor[J].IET Electric Power Applications, 2020, 14(3):2598-2608.
[19] Li Y , Wu L , Wang T ,et al.EEG Signal Processing based on Genetic Algorithm for Extracting Mixed Features[J].International Journal of Pattern Recognition and Artificial Intelligence, 2019, 33(6):1958008.1-1958008.16.
[20] Alharbi A , Alghahtani M .Using Genetic Algorithm and ELM Neural Networks for Feature Extraction and Classification of Type 2-Diabetes MelIitus[J].Applied artificial intelligence, 2019,33(1/4):311-328.