

Research Article

A Method of Enterprise Financial Risk Analysis and Early Warning Based on Decision Tree Model

Xianfu Wei 

Faculty of Management of Chuzhou Polytechnic Chuzhou, Chuzhou, Anhui 239000, China

Correspondence should be addressed to Xianfu Wei; weixianfu@chzc.edu.cn

Received 6 August 2021; Revised 29 August 2021; Accepted 2 September 2021; Published 26 September 2021

Academic Editor: Jian Su

Copyright © 2021 Xianfu Wei. 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.

At present, the domestic and foreign financial crisis early-warning model research will provide only prediction accuracy as the only standard of success for early-warning model, ignoring an important problem, namely, will the financial crisis early-warning model for normal business, compared with the normal enterprise, forecast the financial crisis? This paper reviews the research situation at home and abroad from the perspective of the definition of the enterprise financial crisis, the form of expression, and so on. From the theoretical level, the relationship between the cause of the financial crisis and the change of financial indicators is established by explaining the early-warning theory, early-warning theory of financial crisis, and cost-sensitive learning theory, and the framework of early warning modeling of financial crisis based on decision tree is put forward. The decision tree model is constructed on several training subsets as the base learner so that the decision tree base learner can learn the characteristics of the healthy sample and crisis sample roughly equally. Taking the bond issuing enterprises of manufacturing industry as samples, the empirical comparison shows that the financial warning model based on decision tree integration is more accurate, which indicates that the model can improve the correct identification rate of financial crisis enterprises under the premise of higher overall warning accuracy.

1. Introduction

Financial crisis refers to an economic phenomenon in which an enterprise loses repayment of maturing debts or expenses. At present, with the deepening of global economic integration and China's market economy, enterprises are facing increasingly fierce market competition, and more and more enterprises have financial crises and even bankruptcy liquidation phenomena. The occurrence of a financial crisis is not accidental, but a process of gradual evolution, mostly with aura; as a result, how to detect and predict the financial crisis, to reduce the risk of enterprise operation, to protect the interests of the investors, creditors, and government supervision of the listed companies, and prevent the financial crisis has very important practical significance.

There are many methods of financial early warning, including statistical method and neural network method. The statistical model has obvious explanatory properties, so its stability is not high. An artificial neural network

(ANN) is a system developed by imitating the processing characteristics of the human brain. It has great ability of parallel processing and quick repair of information, but its modeling method is complex, and its understandability is poor. A decision tree is a kind of efficient classifier; classification results are easy to understand, but the decision tree method possesses certain characteristics of condition attributes. Particularly sensitive to the presence of these attributes, YanYing influenced the decision tree classification accuracy and the evaluation rules of conciseness. Pruning will greatly increase the computational complexity and, at the same time, also can not fundamentally solve the problem of incorrect root knot points. This paper puts forward an improved decision tree algorithm, and it is used in the enterprise financial crisis early warning, and through the practical analysis, it has achieved good results.

Existing research on financial early-warning models can be roughly divided into 3 categories.

The first is the model based on statistical measurement methods; the representative methods include discrimination, clustering, logistic regression, and so on. Zhang et al. [1] added the Benford factor to the financial early-warning system and used the Lasso-logistic model to build a financial risk early-warning model. Cielen et al. [2] constructed the weighting decision matrix of dynamic credit evaluation by using the Topsis-GRA method and obtained the dynamic credit evaluation results. Xu and Wang [3] built a dynamic risk warning model for zombie enterprises based on the Kalman filter algorithm. Park and Han [4] used the sequential Probit regression model to predict the default risk of American bond issuers.

The second is the model based on machine learning methods, among which the representative methods include decision tree, support vector machine, neural network, and so on. Chen et al. [5] built a financial risk warning mechanism from the perspective of big data on the basis of analyzing big data technology and enterprises' financial risk warning needs. Min and Jeong [6] used three improved algorithms of the BP neural network to build a financial early-warning model. Li and Sun [7] established an early-warning system of currency crisis by using a decision tree, neural network, and logistic regression.

The third is the combination model based on multiple methods. Lin et al. [8] used the decision tree method to screen personal credit indicators and used a neural network to build a classification model. Chen and Hsiao [9] used logistic regression, decision tree, and support vector machine as the primary learner and support vector machine as the secondary learner to predict the default risk of P2P network loan.

The concept of the decision tree model was first proposed by Hunt et al. in 1966. The most influential model is the ID3 algorithm-based model, which is based on information gain selection of node splitting attributes. Later, an improved C4.5 algorithm based on the information gain ratio selection attribute C5.0 algorithm further improves the recognition rate on the basis of C4.5 algorithm. In recent years, the decision tree C5.0 algorithm has been widely used in risk warning and credit rating. Tsai and Wu [10] used the C5.0 algorithm of decision tree to construct the personal credit rating model of banks. Wang Maoguang et al. [11] established the risk monitoring model of the small-amount online loan platform through the C5.0 algorithm of the decision tree. The above decision tree [12, 13] financial early-warning model ignores the problem of unbalanced proportion between financial normal samples and crisis samples. In the current capital market of China, the financing enterprises (bond issuers, borrowers, etc.) that have financial crisis and insolvency are still a minority, and most of the financing enterprises are in a normal financial state [14]. Moreover, in real life, data are mostly continuous, and it is difficult to directly apply to machine classification control. At the same time, enterprise financial crisis early warning has certain characteristics and complexity, which is also difficult to use traditional statistical analysis to solve [15, 16]. This imbalance between the number of crisis samples and normal samples will lead to the classification model learning more data rules of normal samples during training while ignoring the

regulation mining of crisis samples. Thus the prediction accuracy of crisis samples is too low. Therefore, considering the unbalanced data feature that the number of financial crisis enterprises among bond issuers is much smaller than that of financial health enterprises; this paper builds a decision tree integrated model to solve the credit crisis early-warning problem under the unbalanced data feature and improve the accuracy of early warning [17–19].

1.1. Our Contribution Is Threefold

- (1) This paper presents a financial crisis early-warning modeling framework based on the decision tree
- (2) The decision tree model is constructed on several training subsets as the base learner so that the decision tree base learner can learn the characteristics of the healthy sample and crisis sample roughly equally
- (3) The empirical comparison shows that the financial warning model based on decision tree integration is more accurate, which indicates that the model can improve the correct identification rate of financial crisis enterprises under the premise of higher overall warning accuracy

2. Financial Early-Warning Model Based on Decision Tree Integration

Ensemble learning is the integration of multiple machine learning models (called “individual learners”) in a certain way. Classical integration methods include AdaBoost, Bagging, and random forest. The characteristic of these classical methods is that they can keep individual learners differentiated so as to ensure that each individual learner can reflect different information, and the integrated results can be more comprehensive so as to improve the prediction accuracy. In this paper, homogeneous integration is adopted. That is, the integration only contains the same type of individual learners, and the individual learners are called “basic learners.” Decision tree algorithm is used to build decision tree base learners, and multiple decision tree base learners are integrated by the “clustering Bagging” method so as to solve the problem of financial early-warning prediction under the characteristics of unbalanced data.

2.1. Construction of the Base Learner

2.1.1. Decision Tree Algorithm. The calculation process of information gain ratio is as follows: calculation of information entropy $E(S)$: let S be the sample set; m is the number of categories. In this study, there are two categories: financial normalness and financial crisis, so $m = 2$. p_i is the proportion of each type of sample to the total sample S . Then,

$$E(S) = - \sum_{i=1}^m p_i \log_2(p_i). \quad (1)$$

Indicators of conditional information entropy $E(S|X)$ calculation: suppose an index X has the value v kind of

$\{x_1, x_2, \dots, x_v\}$, the sample can be divided into v subsets $\{S_1, S_2, \dots, S_v\}$. According to equation (1), the information entropy of V subsets $\{E(S_1), E(S_2), \dots, E(S_v)\}$, weighted average the information entropy on each subset, and obtain the conditional information entropy:

$$E(S|X) = - \sum_{j=1}^v \frac{n(S_j)}{n} \times E(S_j), \quad (2)$$

where $n(S_j)$ is the number of samples of sample subset S_j and n is the total number of samples. Conditional information entropy $E(S|X)$ reflects the sample collection classified according to the index value of X , the average resolution of financial crises. Conditional information entropy $E(S|X)$ is smaller, index X for the resolution of the financial crisis.

When calculating the information gain of $G(X)$, you need to calculate the entropy $E(S)$ and the conditional information entropy $E(S|X)$ before you can get the information gain $G(X)$:

$$G(X) = E(S) - E(S|X). \quad (3)$$

The information gain $G(X)$ reflects the ability of index X to distinguish “whether financial crisis occurs.” The larger the information gain $G(X)$ is, the stronger the ability of index X to distinguish “whether financial crisis occurs” is so that the financial crisis samples can be identified more accurately. In order to eliminate the impact of the number of indicator values, the information gain ratio $R(X)$ is further calculated:

$$R(X) = \frac{G(X)}{I(X)} = \frac{G(X)}{- \sum_{j=1}^v (n(S_j)/n) \times \log_2(n(S_j)/n)}, \quad (4)$$

where $n(S_j)$ is the number of samples of sample subset S_j after the sample set is divided according to the value of index X and n is the total number of samples.

Above is the information gain ratio calculation process. A decision tree model is constructed with the information gain ratio as the key parameter, and the steps are as follows:

Step 1: take the complete set of samples as the root node of the decision tree and calculate the information gain ratio $R(X_i)$ of all evaluation indexes. The sample is divided into several subsets according to the value of the split fission quantity, and each subset serves as a node of the next layer. It is assumed that “education background” is the index with the largest information gain ratio among all indicators, and “education background” is selected as the split variable at the root node. The samples are divided into four categories according to the four values (high school, undergraduate, master and above, others) under the index of “education background” to form four nodes in the second layer [20].

Step 2: in the second layer of the decision tree, for the sample set at each node, calculate the information gain ratio of each index on the sample set, and select the index with the largest information gain ratio as the split variable at the current node. Similarly, it continues to split into nodes at the third level according to the value of the split variable [21].

Step 3: in this way, the nodes are generated layer by layer until one of the following three situations is met: (1) All the samples in the sample set of the current node belong to the same category (in this study, all belong to financial crisis enterprises or financial normal enterprises), and the current node is the leaf node. (2) When the sample set of the front node has the same value on all indicators, there is no way to further divide the samples. In this case, the current node is marked with the category to which most of the samples on the current node belong, and the current node is a leaf node. (3) The current node is marked as a leaf node by the class to which most of the samples of the current node’s parent node (nodes directly related to the node at the next level above the node) belong. The flowchart of the decision tree model is shown in Figure 1.

2.1.2. Pruning. In the generation of the decision tree, in order to identify enterprises in financial crisis as much as possible, the samples are divided constantly, and the decision tree is too large, and the training samples are fitted too well, so the prediction ability of new samples out of the training samples is lost. In order to avoid the overfitting problem, EBP (pruning based on error) method is used in this paper to pruning each node of the decision tree from bottom to top. The basic idea is to calculate the prediction error rate before and after pruning. If the error rate after pruning does not increase significantly compared with before pruning, it indicates that this subtree has little influence on the prediction effect and belongs to the redundant branch, which should be cut off.

Suppose T_j is a subtree with node j as the root, the leaf node before pruning is the leaf node of the subtree T_j , and node j is the leaf node after pruning. The prediction error rates e_1 and e_2 of the subtree samples before and after pruning were calculated by pessimistic error rate calculation method. Suppose the sample prediction error rate is a random variable that follows a binomial distribution $U(e, n)$. Given a confidence CF, a confidence interval $[L_{CF}, U_{CF}]$ on the error rate can be obtained. If the expected error rate $n \times e_2$ after pruning is less than U_{CF} , it indicates that the error rate after pruning does not increase significantly compared with before pruning, then pruning. Otherwise, do not prune. The greater the confidence degree CF is, the more serious the pruning is. Generally, the value of CF is 0.75.

2.2. Decision Tree Integration. Most of the bond issuers in the market are financially healthy enterprises, while less than 5% of the bad issuers have a financial crisis. The number of the two types of samples is extremely unbalanced. This situation will lead to the training of the decision tree model to learn more data characteristics of financial healthy enterprises but ignore the feature mining of financial crisis enterprises. This phenomenon is known as the unbalanced sample problem.

Based on the “clustering Bagging” integration method, this paper divides a large number of financial health

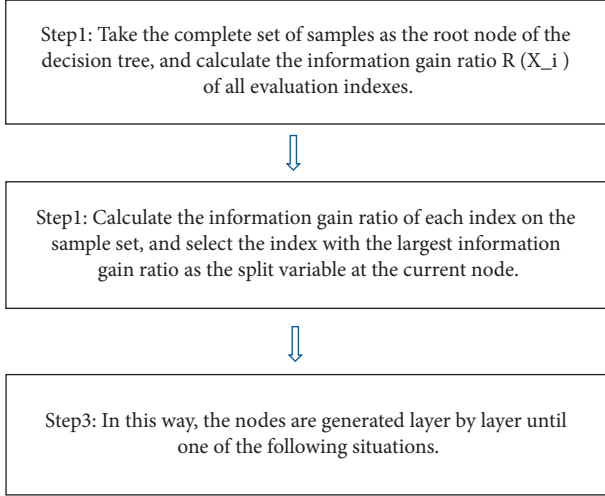


FIGURE 1: Flowchart of decision tree model.

enterprise samples into K groups by k -means clustering method and pairs the financial health samples of K groups with the financial crisis samples to form K roughly balanced training subsets. The decision tree is constructed on K training subsets as the base learner and then integrated to form the final early-warning model so as to solve the problem of nonequilibrium samples in the process of financial early-warning model construction. The specific process of model construction is as follows:

Step 1: clustering. The samples are divided into a training set and a test set. The training set is to train the sample set of the model, and the test set is to verify the prediction accuracy of the trained model. In the training set, it is assumed that D is the sample set of financial health and F is the sample set of the financial crisis. The K mean value clustering method was used to divide healthy enterprise sample D into K parts $\{D_1, D_2, \dots, D_k\}$. Due to the characteristics of the clustering method, it can ensure the maximum difference among all kinds of samples so as to ensure the difference of decision tree-based learning instruments trained by different sample subsets.

Step 2: generate multiple training samples. Will $\{D_1, D_2, \dots, D_k\}$, each set is paired with the crisis sample set F to form k training subsets $\{D_1 \cup F, D_2 \cup F, \dots, D_k \cup F\}$. Because D , which was too large, was divided into k , the number of health samples in each health subset D_i was significantly reduced. Therefore, in the new training subset $D_i \cup F$, the number of health samples and the crisis samples became relatively balanced, thereby reducing the imbalance problem in the overall sample.

Step 3: decision tree-based learner. By using the abovementioned method, decision trees are constructed on the above k training subsets and k base learners $\{M_1, M_2, \dots, M_k\}$. The characteristic of the clustering method makes the difference between different training subsets and ensures the difference of

decision tree based learning instruments trained by different subsets.

Step 4: decision tree integration. Based on the prediction accuracy of the decision tree base learner, the base learner is weighted. The higher the prediction accuracy, the higher the weight so as to form the decision tree integrated learner. The specific method is using k base learners $\{M_1, M_2, \dots, M_k\}$ made a prediction on the test set, compared the forecast results with the actual financial status, and drew the ROC curve.

The abscise of the ROC curve is the pseudopositive rate, that is, the proportion of the predicted positive sample but actually negative sample to all negative sample (in this paper, “the occurrence of the financial crisis” is the research object and is the positive example); the ordinate is the true positive rate, that is, the proportion of all positive samples that are predicted to be positive and actually positive. AUC value is the area surrounded by the ROC curve and absciss, and the AUC value comprehensively reflects the accuracy and sensitivity of the prediction model. The AUC value is used as the weight to add the weight to the decision tree base learner, and then the decision tree integrated learner is obtained as the financial early-warning model. After the above process, the decision tree based learner is integrated, and the financial early-warning model is finally obtained. The above process is shown in Figure 2.

2.3. Technical Route. The specific research technical route in this paper is shown in Figure 3. First of all, the research problem is enterprise financial crisis early warning and then refer to relevant literature to collect relevant data. On this basis, the relevant research on this issue is carried out, and the shortcomings of the existing research are analyzed: the existing financial crisis early-warning model lacks relevant systematic research on the different error costs of the model, and the research perspective of the financial crisis early-warning model based on cost-sensitive learning is determined. In the whole research process, related research results of financial management, pattern recognition, statistics, data mining, and other fields were used for reference. On the basis of proposing the theoretical basis of financial crisis early warning based on cost-sensitive learning and building a financial indicator system combining quantitative and qualitative analysis. The cost-sensitive financial crisis early warning based on cross-section data and cost-sensitive financial crisis early warning based on integrated learning are creatively researched. Comprehensive reference statistical analysis and computer programming and other related technologies, on this basis, through data experiments and data analysis, verify the usefulness and effectiveness of the financial crisis quantitative prediction method.

3. Improved Decision Tree Algorithm

3.1. Decision Tree Attribute Selection. The core problem of the decision tree algorithm lies in the selection criteria of test attributes of each node, which not only affects the size and

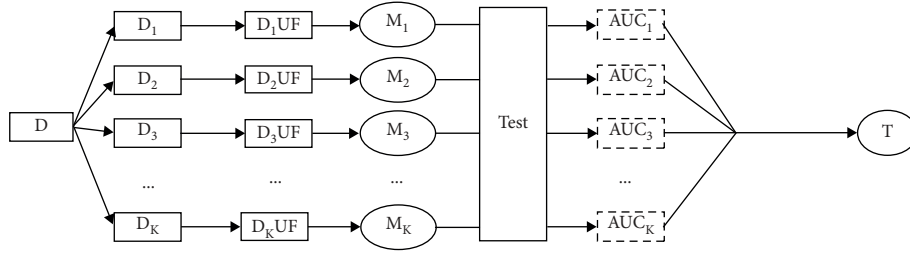


FIGURE 2: Construction process of decision tree integrated learner.

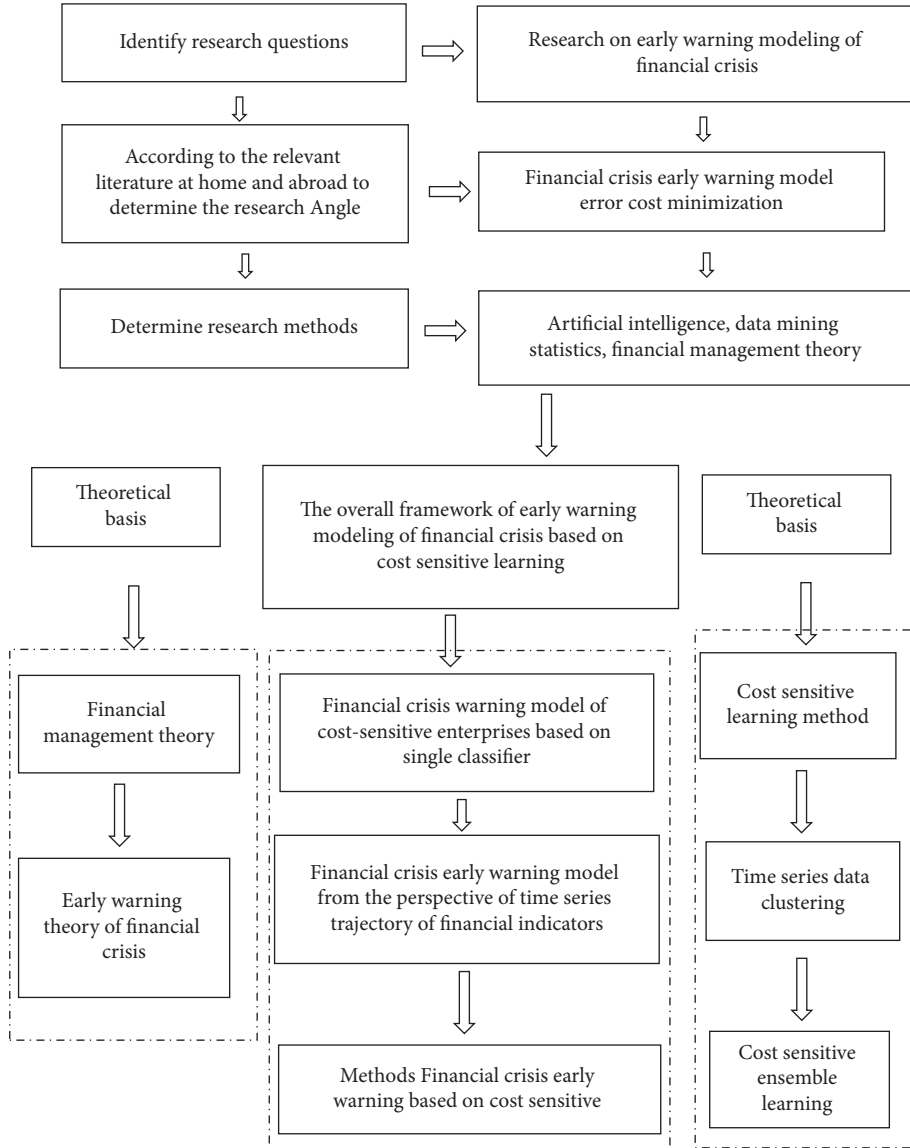


FIGURE 3: Technique route of the research.

prediction accuracy of the decision tree but also is the main part of the calculation of the decision tree. At present, the most widely used are various attribute selection criteria based on information entropy, such as the information benefit standard in the ID3 algorithm and the information gain rate standard in the C4.5 algorithm.

3.1.1. The Information Entropy

$$\text{Entropy}(s) = - \sum_{i=1}^n p_i \log_2 p_i. \quad (5)$$

Among them, S for the sample set, n is the number of target attributes, p_i is the proportion of S belonging to the category of the i .

3.1.2. Information Gain

$$\text{Gain}(S, A) = \text{Entropy}(s) - \sum_{v \in \text{Value}(A)} \frac{|S_v|}{|S|} \text{Entropy}(s), \quad (6)$$

where $\text{Value}(A)$ is the set of values of attribute A , v is A Value in $\text{Value}(A)$, $|S|$ is the total number of samples, and $|S_v|$ is the number of samples of attribute A with the Value of v .

3.1.3. Information Gain Rate

$$\text{GainRatio}(S, A) = \frac{\text{Gain}(S, A)}{SI}, \quad (7)$$

$$SI = - \sum_{i=1}^n \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{|S|},$$

where S_1 to S_n is the set of n samples formed by the attributes of n values dividing S .

3.1.4. Formal Gain

$$\text{NG} = \frac{\text{Gain}(S, A)}{\log_2 n}. \quad (8)$$

The display information gain standard has a preference for more finely divided tilts, while the information gain rate has a preference for nonuniform partitioning, and the normal gain is more efficient than the first two. Therefore, Normalized Gain (NG) is adopted as the attribute selection standard in this paper.

3.2. Dimension Reduction of Decision Tree. Due to the influence of noise or disturbance attributes in the training example set, the decision tree generated from the training set often contains wrong information. This phenomenon is known as overfitting. Pruning techniques are usually used to deal with the overfitting problem after the decision tree is generated. Pruning is another important part of the decision tree size, prediction accuracy, and computation. Since the decision tree is allowed to be overfitted and then cut back repeatedly, the computation is greatly increased, and the

problem of improper root node selection cannot be solved fundamentally. Therefore, it is very important to reduce the dimension before building the decision tree.

Dimension reduction can not only reduce the amount of computation in the subsequent links but also filter out the impact of redundant and harmful attributes on the decision tree. In this paper, the most important attributes are selected to build the decision tree, and the rest are taken as the alternative attribute set. When the decision tree built according to the basic attribute set cannot meet the accuracy requirements, the alternative attribute can be added to the corresponding branch to build the decision tree. This method does not need to use all attributes to build a decision tree like the traditional decision tree algorithm but only calculates some attributes on the basis of the rank of generic importance, which greatly reduces the calculation amount. At the same time, because only the attributes that have the most significant effect on classification are used to establish the decision tree, the minimum decision tree can be generated directly, avoiding pruning calculation and further improving the tree construction efficiency.

3.3. Algorithm Optimization Procedure

- (1) Calculate the NG value of each attribute by formula (5).
- (2) The attributes are sorted according to the NG value from large to small, and the attributes with the largest NG value (such as half) are selected as the basic attribute set, and the rest are the alternative attribute set.
- (3) The decision tree is built on the basic e-set, and attribute importance is used as the criterion for node generic selection.
- (4) In the branch with a high error rate, the attribute with the largest NG value at the node is selected from the set of alternative attributes to build the decision tree until it meets the accuracy requirement. It is worth pointing out that this method takes the NG value as the criterion of node attribute selection, organically combines the establishment of decision tree and attribute dimension reduction, and makes the steps of decision tree construction more compact. In addition, the minimum decision tree can be produced directly by the combination of the two, which avoids the complicated pruning calculation and greatly improves the tree construction efficiency.

4. Experimental Verification and Simulation Results

4.1. Experimental Scheme Design and Parameter Setting

4.1.1. Selection of Experimental Datasets. The experimental dataset in this section is based on the research content. The financial index datasets $U(t-2)$ and $U(t-3)$ of $t-2$ and $t-3$, and the dataset $U(t-2)$ composed of the financial index and the time series of the financial index in $t-2$ years are selected as the research objects.

4.1.2. Determination of the Cost Parameters of Enterprise Error Classification. This paper still assumes that the user of the early-warning model is the granting bank, and the granting bank decides whether to extend loans to the enterprise through the early warning of the financial crisis so as to effectively prevent the emergence of bad debts. When the early-warning model makes a class II error, the bank's loss is the loss of loan interest due to the failure to lend to enterprises. When the early-warning model makes a class I error, the cost of class I error of the early-warning model is assumed to be the full amount of the bank's loan. The cost of type II error is calculated as the lending bank's 3- to 5-year loan interest rate of 6.9, rising by 30%.

4.1.3. Determination of the Number of Rounds of Training Cost-Sensitive Single Classifier. To determine the number of rounds of training cost-sensitive single classifier is also to clarify the number of single classifiers contained in cost-sensitive ensemble classifier. According to the experience of domestic and foreign scholars, the number of single classifiers should not be too small. Otherwise, the classification performance of the ensemble classifier cannot be effectively improved. However, the number of single classifiers should not be increased without limit. Otherwise, the classification ability of the ensemble classifier will not be improved again, and the time complexity and space complexity required for training the ensemble classifier will also be increased. Therefore, the number of training rounds is set as $T = 200$ in this experiment.

4.1.4. Setting of the Type of Single Classifier. With the continuous development of artificial intelligence technology, lots of excellent classification models were found by scholars at home and abroad and are widely used in the real world in many fields, such as neural network, spending vector machine (SVM), Bayes network, and decision tree; however, most of them are only new data points class model, unable to effectively deal with discrete data. Therefore, this section sets the single classifier type as C4.5 decision tree, which can effectively process continuous data and discrete data.

4.1.5. Genetic Algorithm Parameter Setting. According to the experience range of the genetic algorithm, the operation parameters are set as follows: population size is 100, chromosome length is 100, crossover probability and mutation probability are 0.3 and 0.1, respectively, and termination algebra is 200.

4.1.6. Construction of Cost-Sensitive Selective Integrated Early-Warning Model. Datasets $U(t-2)$, $U(t-3)$, and $U(2)$ were randomly divided into two disjointed subsets, respectively, one of which contained 60 pairs of companies in normal financial situation and companies in financial crisis as the training dataset. The evolution curves of maximum fitness and average fitness with $1 - e$ as the fitness standard in each generation of the training dataset are shown in Figures 4–6.

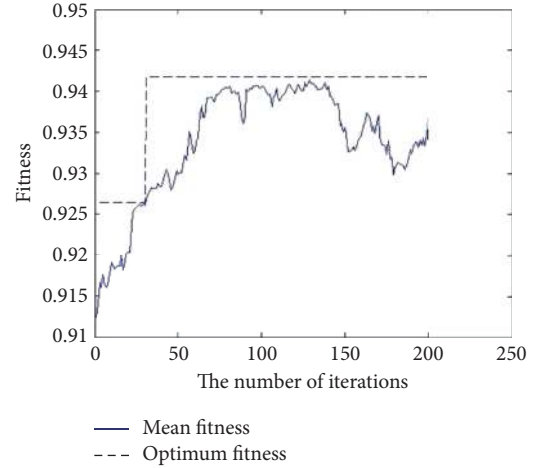


FIGURE 4: Average fitness of $U(t-3)$ dataset and evolution curve of the optimal fitness.

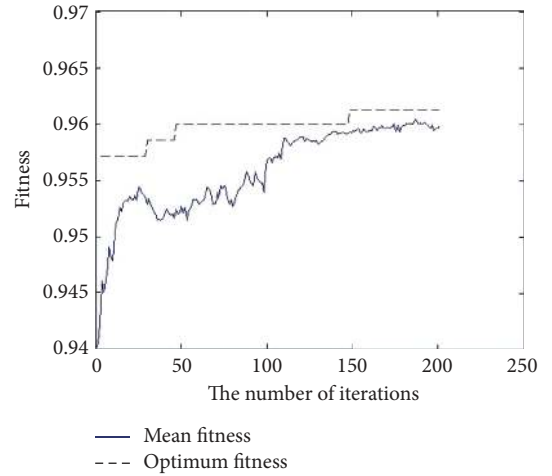


FIGURE 5: Average fitness of $U(t-2)$ dataset and evolution curve of the optimal fitness.

4.2. Experimental Results and Analysis. This section uses test dataset validation and comparative analysis methods to test the effectiveness of enterprise financial crisis early-warning model based on cost-sensitive selective ensemble learning. Datasets $U(t-2)$, $U(t-3)$, and $U(t)$ were randomly divided into two disjointed subsets, respectively. One of the subsets contained 60 pairs of normal companies and companies in financial crisis as the training dataset, and the other contained 25 pairs of normal companies and companies in financial crisis as the test dataset. The enterprise financial crisis early-warning model based on cost-sensitive selective ensemble learning is compared with the experimental results of the enterprise financial crisis early-warning model based on a single classifier and the enterprise financial crisis early-warning model based on a cost-sensitive single classifier. The comparative results are shown in Table 1:

The results of the experiment on $U(t-2)$ show that the average classification error rate of $U(t-2)$ is significantly lower than that of $U(t-3)$. In the three test datasets, the error cost of the cost-sensitive selective integration of the enterprise financial crisis early-warning model is the smallest, which is

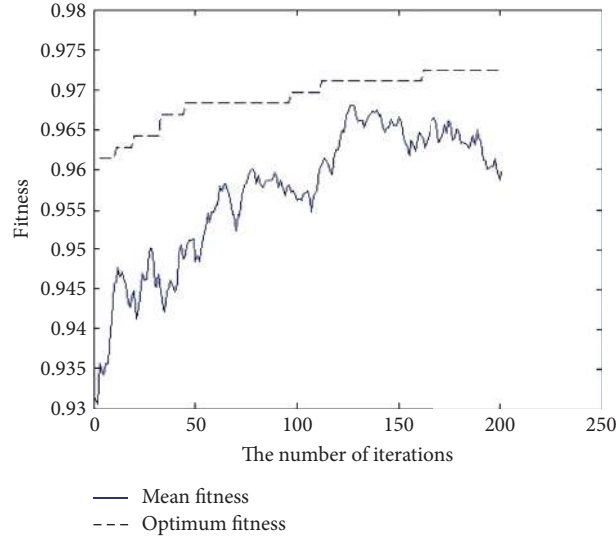


FIGURE 6: Average fitness of $U(t)$ dataset and evolution curve of the optimal fitness.

TABLE 1: Experimental result comparison of business financial distress prediction models.

Data	Decision tree			Single classifier model			Single classifier model based on cost sensitivity		
	I (%)	II (%)	Error percentage	I (%)	II (%)	Error percentage	I (%)	II (%)	Error percentage
$U(t-2)$	8	28	9.7	28	8	26.3	8	48	11.3
$U(t-3)$	12	28	13.3	32	24	31.3	12	52	15.3
$U(t)$	4	36	6.6	12	12	12	4	44	7.3

obviously better than that of the enterprise financial crisis early-warning model based on a single classifier, thus verifying the effectiveness of the method in this section. At the same time, based on the price-sensitive, selective integration enterprise financial crisis warning model fault points price are less price-sensitive single classifier-based enterprise financial crisis warning model, the main reason is that both can significantly reduce the case of type I error, so as to make the early-warning model of fault points cost decrease significantly, but the former which is the first type I error is smaller than the latter. So we get a smaller sum of the cost of the error.

Longitudinal analysis: because all test datasets contain the same number of companies in financial normal and financial crisis, the average classification error rate of the enterprise financial crisis early-warning model is equal to the sum of half the type I error rate and type II error rate. According to the experimental results of datasets $U(t-2)$ and $U(t-3)$, it can be seen that the average classification error rate of the three early-warning models of the corporate financial crisis in $t-3$ years is greater than that in $t-2$ years, indicating that the longer the time is from the outbreak of the financial crisis, the more difficult it is to use financial indicators to predict the financial crisis.

5. Conclusion

With the continuous development of world economic integration, the financial risks faced by enterprises are increasing day by day. How to predict the financial crisis of enterprises in time and accurately is the objective requirement of market competition and the necessary condition for the survival and development of enterprises. In this paper, a new algorithm based on a decision tree is proposed according to the development status of listed companies in China. The algorithm has no requirement on data distribution and is very suitable for financial crisis prediction. At the same time, this method only uses several attributes which have the most significant effect on classification as the reference set to build the decision tree, filters out the interference of harmful redundant attributes, and directly generates the minimum decision tree, which avoids the tedious step of postpruning and greatly reduces the calculation time. In addition, this method takes the NG value as the node attribute selection standard organically combines the establishment of decision tree with attribute dimension reduction and further improves the operation efficiency. The simulation results show that the prediction accuracy of this algorithm is better than that of an ordinary decision tree algorithm.

However, the simple decision tree model is almost invalid for the financial crisis prediction, and nearly 80% of the enterprises in crisis are not identified, which indicates that the model can greatly improve the correct recognition rate of financial crisis under the premise of high overall warning accuracy.

Data Availability

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

Conflicts of Interest

The author declares no conflicts of interest or personal relationships that could have appeared to influence the work reported in this study.

References

- [1] G. Zhang, M. Hu, B. Eddy Patuwo, and D. Indro, "Artificial neural networks in bankruptcy prediction: general framework and cross-validation analysis," *European Journal of Operational Research*, vol. 116, no. 1, pp. 16–32, 1999.
- [2] A. Cielen, L. Peeters, and K. Vanhoof, "Bankruptcy prediction using a data envelopment analysis," *European Journal of Operational Research*, vol. 154, no. 2, pp. 526–532, 2004.
- [3] X. Xu and Y. Wang, "Financial failure prediction using efficiency as a predictor," *Expert Systems with Applications*, vol. 36, no. 1, pp. 366–373, 2009.
- [4] C.-S. Park and I. Han, "A case-based reasoning with the feature weights derived by analytic hierarchy process for bankruptcy prediction," *Expert Systems with Applications*, vol. 23, no. 3, pp. 255–264, 2002.
- [5] H.-J. Chen, S. Y. Huang, and C.-S. Lin, "Alternative diagnosis of corporate bankruptcy: a neuro fuzzy approach," *Expert Systems with Applications*, vol. 36, no. 4, pp. 7710–7720, 2009.
- [6] J. H. Min and C. Jeong, "A binary classification method for bankruptcy prediction," *Expert Systems with Applications*, vol. 36, no. 3, pp. 5256–5263, 2009.
- [7] H. Li and J. Sun, "Majority voting combination of multiple case-based reasoning for financial distress prediction," *Expert Systems with Applications*, vol. 36, no. 3, pp. 4363–4373, 2009.
- [8] F. Lin, D. Liang, and E. Chen, "Financial ratio selection for business crisis prediction," *Expert Systems with Applications*, vol. 38, no. 12, pp. 15094–15102, 2011.
- [9] L.-H. Chen and H.-D. Hsiao, "Feature selection to diagnose a business crisis by using a real GA-based support vector machine: an empirical study," *Expert Systems with Applications*, vol. 35, no. 3, pp. 1145–1155, 2008.
- [10] C. Tsai and J. Wu, "Using neural network ensembles for bankruptcy prediction and credit scoring," *Expert Systems with Applications*, vol. 34, no. 4, pp. 2639–2649, 2008.
- [11] H. Entorf and P. Winker, "Investigating the drugs–crime channel in economics of crime models: empirical evidence from panel data of the German states," *International Review of Law and Economics*, no. 3, pp. 8–22, 2008.
- [12] S. Y. Ho and A. Saunders, "The determinants of bank interest margins," *Journal of Financial and Quantitative Analysis*, vol. 15, no. 1, pp. 191–200, 1980.
- [13] J. Scott, "The probability of bankruptcy," *Journal of Banking & Finance*, vol. 5, no. 3, pp. 317–344, 1981.
- [14] R. E. Schapire, "The strength of weak learnability," *Machine Learning*, vol. 5, no. 2, pp. 197–227, 1990.
- [15] X. Guo, "Research on enterprise financial risk control based on intelligent system," in *Proceedings of the 2019 12th International Conference on Intelligent Computation Technology and Automation (ICICTA)*, pp. 529–534, IEEE, Xiangtan, China, October 2019.
- [16] E. Ayvaz, K. Kaplan, and M. Kuncan, "An integrated LSTM neural networks approach to sustainable balanced scorecard-based early warning system," *IEEE Access*, vol. 8, pp. 37958–37966, 2020.
- [17] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," in *Proceedings of the International Conference on Machine Learning (13th)*, no. 1, pp. 148–156, San Francisco, CA, USA, July 1996.
- [18] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 1–13, 1997.
- [19] Z. H. Zhou, J. X. Wu, and Y. Jiang, "Genetic algorithm based selective neural network ensemble," in *Proceedings of the International Joint Conference on Artificial Intelligence. (17th)*, pp. 797–802, San Francisco, CA, USA, August 2001.
- [20] L. Sun and P. P. Shenoy, "Using bayesian networks for bankruptcy prediction: some methodological issues," *European Journal of Operational Research*, vol. 180, no. 2, pp. 738–753, 2007.
- [21] W. R. Leonard and R. Godoy, "Tsimane' Amazonian panel study, (TAPS): the first 5 years, (2002–2006) of socioeconomic, demographic, and anthropometric data available to the public," *Economics and Human Biology*, vol. 6, no. 7, pp. 299–301, 2008.