ISSN 2616-5902 Vol. 5, Issue 26: 9-14, DOI: 10.25236/AJBM.2023.052602

Bond Credit Rating Based on Machine Learning Model

Tian Zhou[1,*], Weiming Lu[2]

[1]*School of Business, Nanjing University, Nanjing, China*
[2]*School of Accounting, Nanjing University of Finance and Economics, Nanjing, China*
[*]*Corresponding author*

Abstract: *The bond credit rating model in the world is still dominated by the traditional credit rating model of European and American countries, and investors' judgment of bond credit risk almost comes from the credit rating report given by credit rating agencies. However, such a credit rating model may lead to problems such as different rating methods, rating pandering, adverse selection, and the lag of rating results. With the development of artificial intelligence, the bond credit rating model based on machine learning has a good development prospect, which can make up for the shortcomings of the traditional credit rating with strong subjectivity and the uneven level of the corresponding assessors. It can also constantly learn to optimize and automatically adjust the weight to adapt to the new environment, so as to adjust the bond credit rating timely and accurately. The accuracy is much higher than that of the classical statistical regression analysis model.*

Keywords: *Machine Learning; Bond Credit Rating; BP Neural Network*

1. Introduction

In contemporary world, the bond credit rating model is still dominated by the traditional credit rating pattern of European and American countries. Almost all investors' judgment on the bond credit risk is derived from the credit rating report given by the credit rating agency. In our country, credit rating agency is established relatively late. Methods for rating credit are incomplete and different credit rating agencies have different rating methods. In addition, the rating agencies may cater to the issuers, even the bond issuers may pay bribes to the rating agencies to obtain high ratings. Therefore, the rating results given by rating agencies usually lack credibility, so that the rating results are of little significance to investors. To a certain extent, it will also affect investors to make reasonable investment decisions.

In recent years, with the development of Artificial Intelligence, machine learning method has been widely used in Face Recognition, Automatic Control, Medical Diagnosis, Signal Processing and Auxiliary Decision-Making. In the aspect of credit rating, machine learning theory can also provide certain methods and techniques, which provides a whole new perspective for bond credit rating. The bond credit rating based on machine learning model has a promising future, which can make up for the shortcomings where the traditional credit rating is more subjective, and the quality of the evaluators is inconsistent. It can also continuously learn and automatically adjust the weight to adapt to the new environment so that the bond credit rating can be adjusted timely and accurately, whose accuracy is much higher than the classical statistical regression analysis model. Therefore, to establish a bond credit rating based on machine learning model has strong theoretical significance and practical application value.

Based on the deep understanding on connotations and characteristics of bond credit rating, combing the domestic and foreign theory of the bond credit rating and machine learning theory, through in-depth study of the main factors influencing the bond credit rating, this paper will establish a scientific and reasonable bond credit rating index system, and then guided by practical application, this paper will use machine learning method and computer technology to build a bond with self-learning and self-adaptive the credit rating model. Then the adjusted bond credit rating based on machine learning model is used to evaluate the credit of the bonds in the market, and part of the credit results evaluated by the model are displayed. Finally, this paper tests the bond credit rating models based on machine learning model, making a comparative analysis of the advantages and disadvantages of each model.

ISSN 2616-5902 Vol. 5, Issue 26: 9-14, DOI: 10.25236/AJBM.2023.052602

2. Literature Review

Machine learning theory integrates the contents of statistics, neuroscience, cybernetics, information theory, computing science and other disciplines, and belongs to a multi-disciplinary interdisciplinary field. With the development of artificial intelligence, machine learning methods have been widely used in face recognition, automatic control, medical diagnosis, signal processing, auxiliary decision-making and other fields. In terms of credit rating, machine learning theory can also provide certain methods and technical support, which provides a new perspective for bond credit rating.

The research on credit rating based on machine learning model mainly focuses on using neural network and support vector machine to perform credit rating on evaluation objects. Coats P.K. and Fant L.F. (1993)[1] used neural network tools for the first time to identify businesses with credit crises, separating generally healthy companies from those in distress. Lee Y.C. (2007) [2] applied support vector machine to enterprise credit rating, proposed a credit rating model with strong explanatory power and good stability, and used 5-fold cross-validation to find out the optimal parameter values of the RBF kernel function of support vector machine. Shoumo S.Z.H. et al. (2019) [3] proposed a machine learn-based assessment of personal credit risk in commercial banks from the perspective of building intelligent banks. Kim K.J. and Ahn H. (2012) [4] compared the results of multiple machine models, and the results showed that support vector machines can improve the performance of the classifier while occupying less computing space. Zhong H.M. et al. (2014) [5] believe that the method based on extreme learning machine is superior to support vector machines in reliability and outperforms other support vector machines in output distribution. Dastile X. Et al. (2020) [6] put forward two major problems: some machine learning models cannot explain the problems of prediction and data set imbalance, which is an important factor restricting the widespread use of machine learning in the field of bond credit rating.

In view of these deficiencies, Luo C.C. (2020) [7] proposed to dig deep into data from text information and use artificial intelligence to provide a comprehensive decision support method for credit risk assessment based on digital and text information. Choi J. et al. (2020) [8] used text vectorization technology to predict qualitative information of MD&A and converted unstructured text data into numerical vectors so that machine learning algorithms could accept it as an input set.

3. Model Building and Analysis

Table 1: Bond Credit Rating Index System

Primary Index	Secondary Index	Symbol	Type	Description
Basic attribute of bond	Issue scale	SCALE	Continuous	The size of the bond issue
	Issue price	PRICE	Continuous	The price at which the bond is issued
	Rating interval	INT	Scatter	The number of days between a bond's credit rating and its issue date
	Degree of premium	PREM	Continuous	Ratio of issue price to book value
	Maturity of bond	TIME	Scatter	The term of the bond issue
	Linked bond	ASSO	Scatter	The number of associated bonds
	Bond type	TYPE	Scatter	Convertible bonds are 1 and ordinary bonds are 0
Basic attributes of bond issuing enterprises	Listed enterprise	LIST	Scatter	Listed companies are 1, and unlisted companies are 0
	State-owned enterprise	SOE	Scatter	1 for state-owned enterprises and 0 for non-state-owned enterprises
	Corporate attributes	ATTR	Scatter	The number of local enterprises is 0, the number of central enterprises is 1 and the number of foreign enterprises is 2
	Industry type	FORM	Scatter	It is numbered according to the industry
	Number of bonds issued	NUM	Scatter	Whether the same issuer has issued bonds multiple times
Operating conditions of debt-issuing enterprises	Enterprise scale	SIZE	Continuous	The natural logarithm of an enterprise's assets
	growth	GROW	Continuous	Growth rate of business income
	Asset turnover rate	TURN	Continuous	Ratio of total turnover to total assets
	Industry risk	RISK	Scatter	Whether there is a significant risk
Financial status of bond issuing enterprises	Leverage ratio	LEVE	Continuous	The ratio of enterprise liabilities to total assets at the end of the period
	Earnings per share	EPS	Continuous	Ratio of net profit to number of shares outstanding
	Interest cover	EBIT	Continuous	Ratio of EBIT to interest expense
	Enterprise profitability	ROA	Continuous	Ratio of net profit to total assets
	Current assets ratio	CASSET	Continuous	Ratio of cash to ending assets

ISSN 2616-5902 Vol. 5, Issue 26: 9-14, DOI: 10.25236/AJBM.2023.052602

The construction of bond credit rating index system and data preprocessing is the first step in the process of big data analysis, and the next step is to build a bond credit rating model based on machine learning. Machine learning methods can be divided into unsupervised learning and supervised learning according to whether the training data needs labels. The following will use these two methods in order to build a machine learning rating model.

In many theoretical researches on classification decision-making, large data sets with multiple variables will indeed provide rich data information for corresponding research, but also increase the difficulty of data analysis, so it is necessary to adopt necessary technical means to screen the characteristic variables. Dimensionality reduction algorithm is to use fewer indicators to represent more indicators under the condition of minimizing the information loss of original indicators. According to the contribution value, this paper selects 21 main indicators to build a bond credit rating index system, as shown in Table 1.

Since each class is unmarked, we need to define the class ourselves. The definition of category is to dig out the inherent attributes and rules between samples according to the clustering results, and define the samples belonging to the same class in name and attribute. In the bond credit rating, the direct definition of credit rating will be affected by subjective bias, so this paper also refers to the historical rating data of rating agencies, including China Credit, United Credit, Dagong International, Shanghai New Century, China Debt Credit and other historical rating data. Similar to credit ratings from traditional rating agencies, quantified values are set for each credit rating. The credit rating ranges from 1 to 10, corresponding to D (1), C (2), CC (3), CCC (4), B (5), BB (6), BBB (7), A (8), AA (9), and AAA (10).

Neural network is a typical supervised machine learning model. The neural network does not need to determine the function expression of the mapping relationship between input and output in advance, but constructs a network structure that simulates human brain, and through its own learning and training, constantly optimizes and adjusts, it can get a result close to the expected output when the input value is given.

Different from Logistic regression, which can obtain regression expression, neural network is essentially a black box operation. Neural network can construct a mapping relationship between bond credit rating and all influencing factors, but there is no clear expression for this mapping relationship, so we cannot make clear which variable plays a major role, or how much of a role. However, neural network model has more advantages, neural network has strong coupling, can carry out non-linear operations, and the prediction accuracy of the regression function model is often higher. In addition, neural network can simulate the human brain and has the characteristics of intelligence, which is more suitable for the processing of big data.

This part mainly builds a bond credit rating model based on BP neural network. Because BP neural network is a feedforward network with backpropagation, it can constantly modify the weights of connections between nodes by training sample data, so that the error function decreases along the negative gradient direction and constantly approaches the expected output result, which has good robustness. So we choose to use BP neural network for bond credit rating. The specific steps of building a bond credit rating model based on BP neural network are as follows:

Step (1): Initialize network and learning parameters: set the number of nodes in the input layer as, the number of nodes in the hidden layer as, and the number of nodes in the output layer as. The weight from input layer to hidden layer is, the weight from hidden layer to output layer is, the bias from input layer to hidden layer is, the learning efficiency is, and the excitation function is. The excitation function in this paper uses Sigmiod function, and the Sigmiod function is shown in equation (1).

$$\varphi(v) = \frac{1}{1 + exp(-\alpha v)} \tag{1}$$

Step (2) :Provide training mode to train the network.

Step (3) :Forward propagation: In the case of given training mode, input the activation value from the input layer, calculate the actual output of the network, and compare with the expected output, if the error between the actual output and the expected output is controlled within the allowed range, then perform step (4), otherwise go back to step (2). Where, the output of the hidden layer is

$$H_j = g\left(\sum_{i=1}^{n} \omega_{ij} x_i + a_j\right) \tag{2}$$

ISSN 2616-5902 Vol. 5, Issue 26: 9-14, DOI: 10.25236/AJBM.2023.052602

Where $i=1,2,\cdots,n$, $j=1,2,\cdots,l$

The output O_k of the output layer is

$$O_k=\sum_{j=1}^{l}H_j\omega_{jk}+b_k \tag{3}$$

Where $j=1,2,\cdots,l$, $k=1,2,\cdots,m$.

The formula for calculating the error E is

$$E=\frac{\sum_{k=1}^{m}\left(Y_k-O_k\right)^2}{2} \tag{4}$$

Where Y_k is the expected output. $e_k=Y_k-O_k$ denoted as the difference between the expected output and the actual output, then

$$E=\frac{\sum_{k=1}^{m}{e_k}^2}{2} \tag{5}$$

Step (4): Reverse correction: Calculate the error between the same level, correct the weight and threshold, and return to step (2). Wherein, the updating formula of the weight value is:

$$\begin{cases}\omega_{ij}=\omega_{ij}+\eta H_j(1-H_j)x_i\sum_{k=1}^{m}\omega_{jk}e_k\\ \omega_{jk}=\omega_{jk}+\eta H_j e_k\end{cases} \tag{6}$$

The updated formula for the bias is

$$\begin{cases}a_j=a_j+\eta H_j(1-H_j)\sum_{k=1}^{m}\omega_{jk}e_k\\ b_k=b_k+\eta e_k\end{cases} \tag{7}$$

Through the previous dimensionality reduction algorithm, we have obtained 21 main indicators for evaluating bond credit rating, so the number of nodes in the Input layer of the BP neural network is set to 21. Since there is only one variable expected to be Output, that is, the bond credit rating, the number of nodes in the output layer of this paper is 1. In addition, the number of nodes in the Hidden layer is obtained from many tests. After many tests, the number of nodes in the hidden layer is finally set to 33. The detailed BP neural network structure is shown in Figure 1.

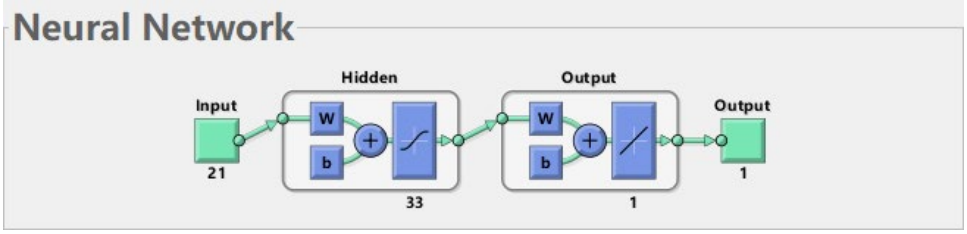

Figure 1: Structure of BP Neural Network

Next, the data Set was divided into Training Set, Validation Set and Test Set according to random partitioning method, and the Levenberg-Marquardt algorithm was used for training, and the mean square error (MSE) was used to calculate the error. The maximum number of iterations allowed in this network is 1000, but in practice only 5 iterations can satisfy the error accuracy requirement. During the training, the maximum error is 1.52, and the actual output error is 0.00201. In addition, the maximum gradient of the network is 5.11, the threshold gradient is $1\times e^{-07}$, and the actual gradient is 0.0375. The minimum damping factor in the Levenberg-Marquardt algorithm used in this network is 0.001, the threshold is 1e-

ISSN 2616-5902 Vol. 5, Issue 26: 9-14, DOI: 10.25236/AJBM.2023.052602

10, and the actual value is $1\times e^{-06}$. In this paper, the network generalization ability check standard is set to 6, which means that if the training set error cannot be reduced for 6 consecutive times during the training process, the training task will be ended.

4. Model Results and Tests

Through the analysis of the results of the bond credit rating model based on machine learning, there may be a situation where the number of samples of two or more categories is very different, resulting in the accuracy of the classifier, which is often called the class imbalance problem. In this paper, we choose to continuously expand the samples of low credit ratings until the number of samples is roughly equal.

The optimized and adjusted machine learning model has better classification performance, and its results on bond credit rating are more reasonable. Next, this paper presents the rating results of the bond credit rating model based on BP neural network. The sample of credit ratings is from 2011 to 2019, with A total of 10,667, and the credit ratings of a large number of bonds are mainly concentrated in AAA, AA, A and BBB grades. As shown in Figure 2, there are 79 samples with bond credit rating D, accounting for 0.7% of the total sample, while there are 4,820 samples with bond credit rating BBB, accounting for 45.2% of the total sample.

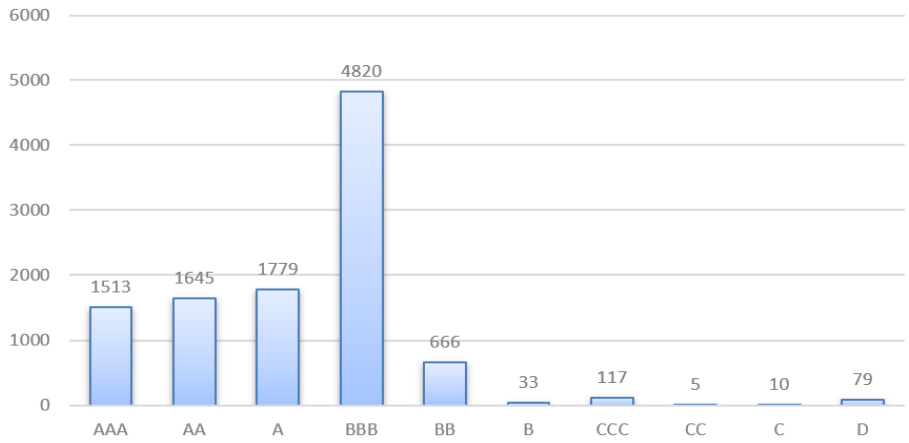

Figure 2: Bond Credit Rating Distribution Chart

Table 2: State-Decision Table of The Bond Credit Rating Model Based on BP Neural Network

Decision \ Status		Quantitative Credit Score									
		1	2	3	4	5	6	7	8	9	10
Quantitative Credit Score	1	0	0	1	0	0	0	0	0	0	0
	2	0	3	0	0	0	0	0	0	0	0
	3	1	0	0	0	0	0	0	0	0	0
	4	1	1	0	7	0	0	0	0	0	0
	5	0	1	0	4	6	0	0	0	0	0
	6	1	0	0	1	0	87	2	1	0	0
	7	0	0	0	0	0	1	377	8	0	3
	8	0	0	0	1	0	3	9	409	2	5
	9	0	0	0	2	1	9	17	12	414	41
	10	0	0	0	0	0	0	1	0	0	168

Next, the bond credit rating based on the machine learning model is tested. Table 2 shows the status-decision table of the bond credit rating model based on BP neural network, specifically which shows the known credit quantization score (Status) and the credit quantization score (Decision) obtained by using the bond credit rating model based on BP neural network. Thus, the accuracy and sensitivity of the model can be obtained.

Kappa coefficient is an index to evaluate the performance of a classifier, which is mainly aimed at multiple classification problems and can comprehensively reflect the classification performance of a classifier.

$$Kappa = \frac{P_0 - P_e}{1 - P_e} \tag{8}$$

The Kappa coefficient is calculated based on the confusion matrix, and its value range is, which is usually greater than 0 when the classifier performance is actually tested. In the state-decision table, the

ISSN 2616-5902 Vol. 5, Issue 26: 9-14, DOI: 10.25236/AJBM.2023.052602

more consistent the state and decision are, the higher the Kappa coefficient is, which means the better the performance of the classifier. Table 3 shows the consistency criteria of Kappa coefficient to test the performance of the classifier.

Table 3: Consistency Criteria of Kappa Coefficient Test Classifier

Kappa Coefficient	0~0.2	0.2~0.4	0.4~0.6	0.6~0.8	0.8~1.0
Consistency Level	Low	General	Medium	High	Almost

The value of the bond credit rating model based on BP neural network is 0.919 and 0.232, and its Kappa coefficient is calculated as 0.895 by equation (8). According to the consistency criterion of the Kappa coefficient test classifier, it can be seen that the classification performance of the bond credit rating model based on BP neural network is superior. The model's decision on credit quantization score is almost exactly consistent with the real state of credit quantization score.

5. Conclusions

This paper uses machine learning method and computer technology to construct a bond credit rating model with self-learning and adaptive ability - bond credit rating based on BP neural network model. Then an optimized and adjusted machine learning-based bond credit rating model is used to evaluate the credit of the bonds in the market.

First, the model can process a large number of bond samples in a short period of time and directly give the result of bond credit rating, which is time-saving, high efficiency and low cost. It is a bond credit rating theory in which "artificial intelligence" replaces "human labor".

Second, the bond credit rating model based on machine learning constructed in this paper is completely objective and free from human subjectivity. This is because the 21 main indicators that play a major role in bond credit rating are obtained through principal component analysis, which is not subject to human interference, and avoids the disadvantages of strong subjectivity caused by the traditional bond credit rating mode by experts.

Third, the bond credit rating model based on BP neural network constructed in this paper is the best bond credit rating model. BP neural network is suitable for solving nonlinear problems with strong coupling. No matter how complex its internal mechanism is, BP neural network will learn slowly and approach the expected output indefinitely, which has strong robustness.

References

[1] Coats P.K. and Fant L.F. Recognizing Financial Distress Patterns Using a Neural Network Tool. Financial Management, 1993,22(3):142~155
[2] Lee Y.C. Application of Support Vector Machines to Corporate Credit Rating Prediction. Expert Systems with Applications, 2007, 33(1): 67~74.
[3] Shoumo S.Z.H., Dhruba M.I.M., Hossain S., Ghani N.H., Arif H. and Islam S. Application of Machine Learning in Credit Risk Assessment: A Prelude to Smart Banking. New Jersey: Proceedings of the 2019 Ieee Region 10 Conference, 2019. 2023~2028
[4] Kim K.J. and Ahn H. A Corporate Credit Rating Model Using Multi-Class Support Vector Machines with an Ordinal Pairwise Partitioning Approach. Computers & Operations Research, 2012, 39(8): 1800~1811
[5] Zhong H.M., Miao C.Y., Shen Z.Q. and Feng Y.H. Comparing the Learning Effectiveness of BP, ELM, IELM, and SVM for Corporate Credit Ratings. Neurocomputing, 2014(128):285~295
[6] Dastile X., Celik T. and Potsane M. Statistical and Machine Learning Models in Credit Scoring: A Systematic Literature Survey. Applied Soft Computing, 2020(91):1~21
[7] Luo C.C. A Comprehensive Decision Support Approach for Credit Scoring. Industrial Management & Data Systems, 2020, 120(2): 280~290
[8] Choi J., Suh Y. and Jung N. Predicting Corporate Credit Rating Based on Qualitative Information of MD&A Transformed Using Document Vectorization Techniques. Data Technologies and Applications, 2020, 54(2): 151~168