Received: 16 July 2022 | Revised: 30 July 2022 | Accepted: 19 August 2022 | Published online: 15 September 2022

RESEARCH ARTICLE

Journal of Global Humanities and Social Sciences
2022, Vol.3(3)57-60
DOI: 10.47852/bonviewGHSS2022030305

Financial Risks Classification Early Warning Analysis of Data Mining Technology

BON VIEW PUBLISHING

Habib-ur Rehman[1], Zhifeng Yao[1], Jia An[1,*] & Yujie Tang1
[1]*York University, Canada*

Abstract: To further improve the informatization level of financial risk early warning, a financial risk classification early warning method based on a neural network quantile regression algorithm is proposed. Among them, macro and micro indicators are selected as the index input of early warning, and then the neural network quantile regression algorithm is used to classify and warn of the financial risks, finally, the specific risk level is output. Simulation results show that the MAE and RMSE of the neural network quantile regression algorithm are 7.12e-09 and 1.301e-08, which are lower than those of the BP neural network and generalized neural network. Thus the superiority of the neural network quantile regression model is verified.

Keywords: evaluation index; neural network quantile regression; financial risks early warning sharing economy; alternative capitalist model; the form of capitalism

1.Introduction

The early warning of financial risk has always been the focus of scholars at home and abroad, so that the design of financial risk prediction model has been an endless stream. Liu Xixian et al. constructed a financial risk early warning system based on the credit data of credit investigation system, which plays a certain optimization role in the risk detection and overall detection effectiveness in the credit market (Liu & Zhang, 2022). Yang Kaisen et al. constructed a financial risk early warning and prediction system based on big data, which takes machine learning algorithm as the basis, effectively improving the accuracy and scientificity of warning system (Yang & Fu, 2021). The above design process of the financial risk early warning system lacks reasonable solutions to nonlinear problems, so that the final accuracy of prediction and early warning still has a large space to improve. In this study, the financial risk classification and early warning model will be constructed based on the neural network quantile regression algorithm, so as to improve the accuracy of prediction and early warning. Finally, the superiority of the model will be verified through experiments.

2.Construction of financial risk index system

The premise of financial risk early warning is to build a reasonable index system (Dang et al., 2022)(Zhao, 2022). Referring to the research results of scholar You Yue, the index system selected in this study is as shown in Table 1, which is evaluated from a macro and micro perspective, respectively.

Table 1 Financial risk classification early warning index system

Dimension	Early warning index
Macro	Inflation rate
	GDP growth rate
	Dependency of debt
	M2/GDP
Micro	Non-performing loan ratio
	Average P / E ratio
	Capital adequacy ratio
	Year-on-year growth rate of online credit / GDP
	Investment of real estate developers / total fixed investment of social assets
	Housing price growth rate

Corresponding Author: Jia An
York University, Canada. Email: anjia0816@icloud.com

3.Construction of Financial Risk early warning model

3.1 Neural network quantile regression algorithm (Li et al., 2022)(Kang & Liu, 2022)

The neural network quantile regression algorithm (Kang & Liu, 2022)(He et al., 2022) selected in this study is analyzed.

For feedforward neural network, its overall structure is composed of input layer, hidden layer and output layer. The structure of nonlinear relational neural network can be expressed as:

$$f(X_t,W,b)=g^{(0)}\left(\sum_{j=1}^{n}\omega_j^{(0)}\left(g_j^{(h)}\left(\sum_{i=1}^{m}\omega_{ij}^{(h)}X_{t,i}+b_j^{(h)}\right)\right)+b^{(0)}\right) \quad (1)$$

Where, $\omega_{ij}^{(h)}(j=1,2,\cdots,n)$ represents the weight of output layer to connect hidden layer; $\omega_j^{(0)}$ represents the weight of hidden layer to connect output layer; $g_j^{(0)}(t)=t$ represents the output layer activation function; $g_j^{(h)}(t)=\tanh(t)$ represents the activation function of hidden layer; $W=\left(\omega_{11}^{(h)},\omega_{12}^{(h)},\cdots,\omega_{mn}^{(h)},\omega_1^{(0)},\cdots,\omega_n^{(0)}\right)'$ represents the weight vector and $b=\left(b_1^{(h)},b_2^{(h)},\cdots,b_n^{(h)},b^{(0)}\right)'$ represents the threshold vector; $b_j^{(h)}$ and $b^{(0)}$ are the threshold of hidden layer and output layer.

The constructed neural network and quantile regression are combined to construct the neural network quantile regression model, which is shown as follows:

$$Q_{Y_t}(\tau|X_t)=f(X_t,W(\tau),b(\tau))$$
$$=g^{(0)}\left(\sum_{j=1}^{n}\omega_j^{(0)}(\tau)\left(g_j^{(h)}\left(\sum_{i=1}^{m}\omega_{ij}^{(h)}(\tau)X_{t,i}+b_j^{(h)}\right)\right)+b^{(0)}(\tau)\right) \quad (2)$$

Where, $W(\tau)$ stands for weight vector; $b(\tau)$ stands for threshold vector; τ stands for quantile; X_t stands for input variable.

To avoid overfitting, a penalty term is introduced to the objective function (He et al., 2022)(Zhuo & Lin, 2022). The parameter vector after introducing penalty term is:

$$\left(\hat{W}(\tau)',\hat{b}(\tau)'\right)=\arg\min_{W,b}\frac{1}{T}\sum_{t=1}^{T}\rho_\tau(Y_t-f(X_t,W,b))+\lambda\frac{1}{mn}\left\|\omega^{(h)}(\tau)\right\|_2 \quad (3)$$

Where, λ is the penalty parameter; $\omega^{(h)}(\tau)$ represents the weight vector of hidden layer; Y_t is the response variable; $\hat{W}(\tau)'$ and $\hat{b}(\tau)'$ are the estimate values of parameter vector.

Substituting the obtained estimate values $\hat{W}(\tau)'$ and $\hat{b}(\tau)'$ of parameter vector into (2), the conditional quantile estimation of Y at the quantile τ can be obtained:

$$Q_{Y_t}(\tau|X_t)=f_\tau\left(X_t,\hat{W}(\tau),\hat{b}(\tau)\right) \quad (4)$$

3.2 Early warning process

Combined with the above neural network quantile regression algorithm, the financial risk early warning process of this study is designed as shown in Fig. 1.

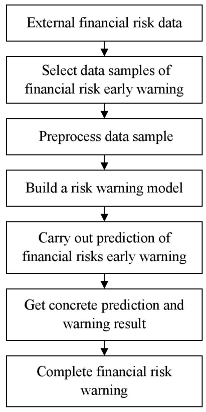


Fig.1 Early warning process

4.Experimental verification

4.1 Experimental environment

The main experimental platform for this study is MATLAB. The probability density curve simulation of response variable Yt, risk prediction experiment and model error comparison experiment are all implemented on MATLAB.

4.2 Data source and preprocessing

4.2.1 Data source

This study selects the financial risk data of a city from the first quarter of 2014 to the fourth quarter of 2021 as a sample and the data of each quarter as a sample.

4.2.2 Data preprocessing

4.2.2.1 Normalization processing

The proposed early warning index system differs greatly in data expression, so the data used in the early warning experiment needs to be normalized:

$$y_{ij}=\frac{x_{ij}-\min x_i}{\max x_i-\min x_i},i=1,2,\cdots,15 \quad (5)$$

Where, i represents the constructed financial risk warning index; j is the year; x_{ij} represents the value of each indicator in each year. y_{ij} represents the index after standardized treatment.

After normalization, the data used in the experiment is in the interval [0,1].

4.2.2.2 Missing value processing

The problem of missing values has a greater impact on the accuracy of data. In the financial risk early-warning indicators constructed in this study, unqualified indicators are eliminated, and average filling treatment is carried out for the remaining indicators to ensure the data quality.

4.3 Evaluation index of model

4.3.1MAE & RMSE

(1) Mean absolute error (MAE)

$$MAE=\frac{1}{T}\sum_{t=1}^{T}\left|\hat{y}_t-y_t\right| \quad (6)$$

(2) Root mean square error (RMSE)

$$RMSE=\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\hat{y}_t-y_t\right)^2} \quad (7)$$

4.3.2 Accuracy

The formula of accuracy is as follows(Xiang et al., 2022)(Cui & Ma, 2022):

$$Accurary = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

4.3.3AUC&KS

Both AUC and KS evaluation indicators are based on the area under the ROC curve to show evaluation effect, and the ROC curve reflects the classification ability of classifier. The difference is that the horizontal and vertical axis data for the coordinate system are different, and the KS calculation equation is:

$$KS = \max(TPR - FPR) \quad (9)$$

4.4 Parameter settings of the neural network quantile regression model

The upper limit of iterations of neural network quantile regression model is set as 6000. To facilitate data analysis, the range of quantile is set as (0,1), and the distance between each quantile is set as 0.01. Set the number of repeated tests to 2 to prevent the occurrence of local minimum; In the selection of transfer function, sigmoid is taken as the transfer function of hidden layer. The relationship equation can be obtained according to AIC criterion:

$$AIC(\lambda, n) = \min\left(\ln\left(E_{y_t}\left(\tau;\theta(\tau;\lambda,n)\right)\right)\right) + \frac{m(n+1)}{T} \quad (10)$$

Where, n represents the optimal value of hidden layer node; λ represents the penalty parameter.

According to the (10), it can be concluded that the number of output nodes, hidden layers and output layers of neural network used in the financial risk warning are 4,1 and 3, respectively.

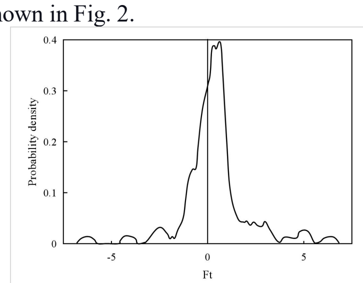
After the specific structure of neural network used by the early warning model is determined, the prediction of response variable Y_t can be made, and the probability density curve of Y_t can be further estimated by introducing Epanechnikov kernel density estimation function.

On the premise of kernel density estimation, the risk level of early warning should be divided. Referring to the results of some researchers, the financial risk level is divided into four levels: security (A), basic security (B), high vigilance (C) and danger (D), and the corresponding interval is divided into [-1.04, -0.55), [-0.55, 0.03), [0.03, 0.78), and (0.78, 1.05].

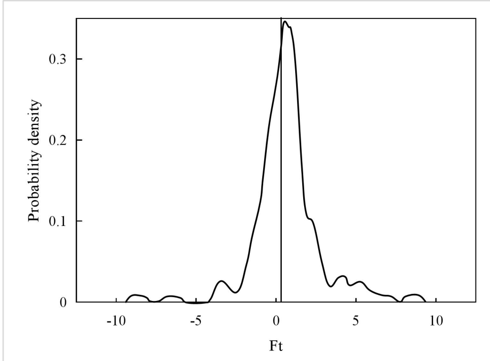
4.5 Financial risk early warning results

4.5.1 Financial risk early warning prediction experiment

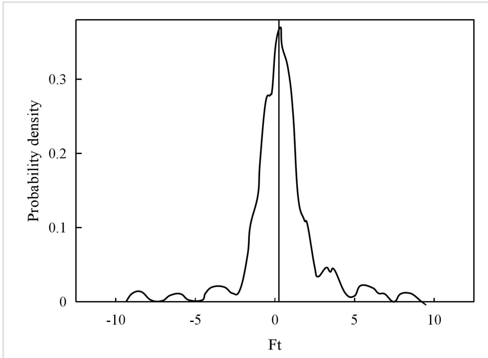
In this study, the index data of the first quarter of 2018 in the data sample is selected as input, and the factor data after two quarters lag is used as nonlinear mapping of response variables, thus the obtained probability density curve is shown in Fig. 2.



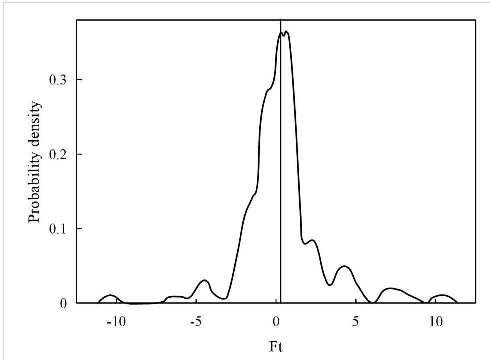
（a）3th month



（b）6th month



（c）9th month



（d）12th month

Fig.2 Probability density curve after lagging two quarters

As can be seen from the figure above, the predicted value is always within the reasonable range of the true value and meets the experimental requirements.

Ten groups of processed 22 data samples are selected as test sets, and the prediction results are shown in Fig. 3:

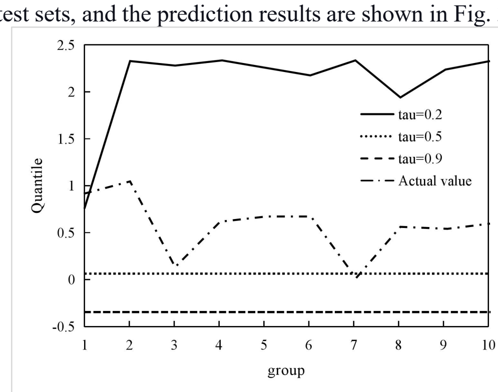


Fig.3 Prediction situation

The quantile numbers are chosen as the predicted value in this study. The interval estimation is determined by referring to the corresponding values of quantiles 0.5 and 0.2, and the prediction results are as follows:

Time	Interval estimation	Point estimation	Risk level
The Fourth quarter of 2019	（0.069,2.349）	0.069	C
The First quarter of 2020	（0.068.2.365）	0.070	C
The Second quarter of 2020	（1.119,3.335）	1.110	D
The Third quarter of 2020	（1.120,2.560）	0.970	D

Table 2 Financial Risk prediction Structure

As can be seen from the table above, the second two quarters of the 10th year are in a state of high vigilance and even danger, which deserves attention.

4.5.2 Model Comparison

To verify the rationality of index evaluation model selected in this study, it is compared with the BP neural network model and generalized neural network model.

Table 3 Comparison of model prediction errors

Model	GRNN	BP neural network	QRNN
RMSE	0.2950	0.1151	1.301e-08
MAE	0.1035	0.0449	7.12e-09

It can be seen from the above table that the neural network quantile regression model has higher prediction accuracy and more accurate prediction results.

To verify the rationality of selection algorithm in this study, the random forest algorithm and logistic regression algorithm are used for modeling, and the KS value, Accuracy and AUC value are compared and analyzed. The results are shown in Table 4:

Table 4 Model test effect

	Accuracy	RMSE	MAE	KS	AUC
Neural network quantile regression	--	1.301e-08	7.12e-09	--	--
Logistic regression	0.880	0.129	0.063	0.702	0.907
Random forest	0.859	0.199	0.131	0.672	0.827

It can be seen from Table 4 that the test effect of the model designed in this study is good, and it is more suitable for the classification early warning of financial risks.

5.Conclusion

In conclusion, the classification early warning model of financial risk designed based on neural network quantile regression can accurately predict financial risks. Compared with BP neural network and generalized neural network model, the MAE and RMSE of neural network quantile regression are only 7.12e-09 and 1.301e-08, which are completely suitable for the classification early warning of financial risks. However, there are still shortcomings in the design. For example, in data screening, some important indicators have been removed due to data quality problems, which indicates that there is still room for improvement in the accuracy of prediction. Therefore, the next step of research is to optimize this aspect.

Conflict of Interest

The authors declare that they have no conflicts of interest to this work.

References

Liu, X., & Zhang, X. (2022). Research on construction of financial risk early warning system based on data of credit investigation system. *Credit Investigation*, 4005.

Yang, K., & Fu, F. (2021). Financial risk early warning system based on big data. *Journal of Changchun University of Science and Technology*, 401.

Dang, Y., Miao, Z., & Zhang, T. (2022). Progress in the application of big data in the monitoring and early warning of systemic financial risk. *Financial Development Research*, 02.

Zhao, Y. (2022). Early warning analysis of systemic financial risk level based on entropy method. *Industrial Innovation Research*, 07.

Li, T., Zhou, T., & Hu, Q. (2022). Early warning analysiPrediction of Ship Energy Consumption Interval Based on Quantile Regression of Gradient Lifting Decision Tree s of systemic financial risk level based on entropy method. *Journal of Shanghai Maritime University,* 432.

Kang, N., & Liu, T. (2022). Research on RMB Exchange Rate Flows and Prediction Based on Portal Quantile Autoregression Model. *Journal of Fuyang Normal University*, 3902.

He, D., Li, R., & Sun, S. (2022). Prediction of Gas-liquid Two-phase Pressure Rise in Centrifugal Pump Based on Machine Learning. *Transactions of the Chinese Society of Agricultural Engineering*, 3804.

Zhuo, M., & Lin, W. (2022). Prediction model of leakage Location of cross-sea pipeline based on machine learning. *Progress in Science and Technology of Water Resources and Hydropower*, 4203.

Xiang, S., Zeng, Y., & Yan, X. (2022). Research on bond default risk monitoring and early warning based on support vector Machine method. *Financial Economics*, 01.

Cui, H., & Ma, P. (2022). Wind Power Interval Prediction Based on Optimized BP Neural Network and Nonparametric Estimation. *Electronic Design Engineering*, 3013.

How to Cite: Rehman, H., Yao, Z., An, J., & Tang, Y. (2022). Financial Risks Classification Early Warning Analysis of Data Mining Technology. *Journal of Global Humanities and Social Sciences*, *3*(3), 56–59. https://doi.org/10.47852/bonviewGHSS2022030305