

Review Article

Application of Several Classical Sorting Algorithms in Early Warning of Payment Risk of Basic Endowment Insurance Fund in China

Xiaohua Chen*

School of Insurance, Central University of Finance and Economics, Beijing, China, 100081

***Corresponding Author**

Xiaohua Chen

Article History

Received: 20.11.2019

Accepted: 27.11.2019

Published: 30.11.2019

Abstract: The financial situation of China's basic endowment insurance fund has begun to deteriorate, and its deterioration trend will accelerate with the deepening of the aging population, so it is urgent to carry out a study on the early warning of the payment risk of this endowment insurance fund. This paper discusses the classification accuracies of C4.5 algorithm, Naive Bayesian algorithm and BP neural network in the warning of basic endowment insurance fund payment risk. It is found that C4.5 algorithm has the best classification effect, with an accuracy of 71.43%; BP neural network takes the second place, with an accuracy of 61.90%; Naive Bayesian algorithm has a poor effect, with an accuracy of 52.38%.

Keywords: basic endowment insurance, payment risk, early warning, sorting algorithm.

INTRODUCTION

According to "China Labor Statistics Yearbook 2018", there are 6 provinces in China where the incomes of basic endowment insurance fund are less than its expenditures, namely Liaoning Province, Jilin Province, Heilongjiang Province, Shandong Province, Hubei Province and Qinghai Province. Among them, Heilongjiang Province has a particularly severe financial situation, with a total deficit of 48.62 billion yuan. These data show that the financial situation of the basic endowment insurance fund in China has gradually deteriorated. Some scholars' studies (Tan and Fan [1], Yu and Zhong [2], Feng and Liu [3], Ai and Zhang *et al.* [4], Liu [5]) also show that the finance of the basic endowment insurance fund is not sustainable or only has a very weak sustainability in the long term, which will directly affect the payment of pensions for retirees and further affect the credibility of the government. Therefore, it is urgent to provide early warning on the finance of the China's basic endowment insurance fund.

LITERATURE REVIEW

The research on early warning of China's basic endowment insurance fund payment risk mainly focuses on early warning's risk factors, indexes construction and models.

Early warning risk factors. From the perspective of operation management, Chen [6] advocated the use of risk control technology to reduce the losses caused by management, psychology, systems, investment, technology and other risks in the process. Cheng [7] believed that risk management of basic endowment insurance fund mainly includes five aspects, namely policy risk, operational risk, fund risk, information risk and moral risk.

Construction of early warning indicators. Li and Wu [8] summarized the index system of early warning of basic endowment insurance system from three aspects: external operation environment, system operation and design. Pan and zheng [9] designed the index system of management risk, moral risk and premium collection risk for basic endowment insurance fund. Chang and Chen [10] designed 10 relevant early warning indicators from the aspects of the operation, design and environment of the basic old-age insurance system, and carried out empirical analysis on the early warning indicators to test the early warning effect.

Early warning models. These studies (Chang and Chen [10], Gao and Gu [11], Ye and Li [12]) applied BP neural network to the risk warning of basic endowment insurance fund, and the results showed that BP neural network model had a good effect of risk

Copyright @ 2019: This is an open-access article distributed under the terms of the Creative Commons Attribution license which permits unrestricted use, distribution, and reproduction in any medium for non commercial use (NonCommercial, or CC-BY-NC) provided the original author and source are credited.

warning. Zhu [13] summarized the research status of risk assessment, risk warning and risk factors of the endowment insurance fund, reconstructed the early-warning indicator system, and conducted empirical study with the improved BP neural network.

On the basis of the above research, this paper first constructs the early warning indicators of fund payment risk from the perspective of the main factors affecting the incomes and expenditures of the basic endowment insurance system. Secondly, it focuses on the application of C4.5 decision tree, Naive Bayesian algorithm and BP neural network in the early warning of the financial status of the insurance fund. Thirdly, the warning effects of these three classification methods are compared, and the best warning classification method is selected. Finally, based on the best method, the financial risk level of the endowment insurance fund in the next 10 years is further predicted.

Early warning indicators construction and data

Early warning indicators

From the perspective of the factors that affect the contributions and expenditures of the endowment insurance fund, the early warning indicators of fund payment risk is constructed.

The main factors affecting the incomes of the fund are: the average annual salary of the employees, the number of employees, the growth rate of the wages of the employees, the contributions paid by the enterprise and the employees, etc., so the following indicators are refined: average annual salary of employees (C1), wage growth rate (C2), growth rate of on-the-job insureds (C3), unit contribution rate (C4) and individual contribution rate (C5). The main factors that affect the pension expenditures are: the level of pension treatment and the proportion of this level to pre-retirement wages, accumulation of individual pension accounts of employees, the situation of the retirees who need to pay the pension, the growth rate of the number of retired insureds, the life expectancy of the retirees and the growth of the pension expenditures as a whole, etc. Therefore, the following indicators can be extracted: bookkeeping rate (P1), the rate of support (P2), growth rate of retired insureds (P3), average life expectancy (P4) and growth rate of pension expenditures (P5). The specific meaning of these indicators is shown in Table 1.

Dependent variable (Y): the payment capacity of fund is the accumulated balance of the basic endowment insurance fund divided by the pension expenditures of the current year, and multiplied by 12 to convert its unit from year to month. In other words, the payment capacity (Y) is the number of months that the fund can still pay pension to retirees in the current year.

According to Ye and Li [12], when $Y \geq 12$, the fund has sufficient capacity to pay, and the payment risk is recorded as class A; when $9 \leq Y < 12$, fund payment has a slight degree of risk, and the risk is recorded as class B; when $6 \leq Y < 9$, fund payment has a medium degree of risk, and the risk is recorded as class C; when $Y < 6$, fund payment has a serious degree of risk, and the risk is recorded as class D.

Table-1: Construction of Early Warning Indicators

type	Indicators	definition of indicators
pension contributions	average annual salary of employees (C1)	the average annual statistical wages of on-the-job employees
	wage growth rate (C2)	the growth rate of per capita disposable income of urban residents
	growth rate of on-the-job insureds (C3)	number of on-the-job insureds in current year minus the number of last year, and then divide by the number of last year
	unit contribution rate(C4)	the employers shall pay the old-age insurance premium rate for the on-the-job employees
	individual contribution rate(C5)	insurance premium rate paid by the staff themselves
pension expenditures	bookkeeping rate (P1)	accounting interest on individual account pension
	the rate of support (P2)	the proportion of the number of retired insureds to on-the-job insureds
	growth rate of retired insureds (P3)	number of retired insureds in the current year minus the number last year, and then divide by the number of last year
	average life expectancy (P4)	the average age a retiree can expect to live
	growth rate of pension expenditures (P5)	the expenditures in the current year minus the expenditures of last year, and then divide by the expenditures of last year

Data of indicators

Since there were few national data from statistical yearbook, relevant representative provinces' data were also selected as samples from all regions of the country. The data of the early warning indicators for 2006-2017 in the whole country, Beijing, Heilongjiang, Shanghai, Henan, Chongqing and Shanxi provinces were selected as samples, totaling 84 samples.

According to the "China Human Resources and Social Security Yearbook 2018" and the corresponding statistical yearbook of each region, data of the average annual salary (C1) and wage growth rate (C2) of employees can be obtained. From the "China Labor Statistics Yearbook 2018", data of growth rate of on-the-job insureds (C3), the rate of support (P2) and growth rate of retired

insureds (P3) were available. According to the State Council Document 38 in 2005, it can be seen that the unit contribution rate (C4) is 0.2 and the individual contribution rate (C5) is 0.08. In 2015 and the previous years, bookkeeping rate (P1) is the average 3.27% of the one-year deposit benchmark interest rate of the People’s Bank of China from 1995 to 2015; in 2016 and the subsequent years, the P1 is the average 7.18% of the 2016-2018 bookkeeping rate published by the government. According to “China population and Employment Statistical Yearbook 2018”, the average life expectancy of the population in 2000 and 2010 in China and each region can be known. Assuming that the growth rate of average life expectancy is the same every year, and then data of the average life expectancy (P4) in 2006-2017 can be obtained. The growth rate of pension expenditures (P5) from 2006 to 2017 can be calculated from the “China Labor Statistics Yearbook 2018”. The national samples are shown in Table 2.

Table-2: National samples of warning indicators

year	pension contributions					pension expenditures					Y
	C1	C2	C3	C4	C5	P1	P2	P3	P4	P5	
2006	21001	0.104	0.073	0.2	0.08	0.0327	0.332	0.061	73.458	0.218	A
2007	24932	0.122	0.073	0.2	0.08	0.0327	0.328	0.068	73.801	0.218	A
2008	29229	0.084	0.087	0.2	0.08	0.0327	0.326	0.070	74.144	0.238	A
2009	32736	0.098	0.075	0.2	0.08	0.0327	0.319	0.094	74.487	0.203	A
2010	37147	0.078	0.091	0.2	0.08	0.0327	0.327	0.085	74.830	0.186	A
2011	42452	0.084	0.104	0.2	0.08	0.0327	0.324	0.082	75.173	0.209	A
2012	47593	0.096	0.071	0.2	0.08	0.0327	0.316	0.090	75.516	0.219	A
2013	52388	0.070	0.058	0.2	0.08	0.0327	0.323	0.079	75.859	0.186	A
2014	57361	0.068	0.059	0.2	0.08	0.0327	0.332	0.068	76.202	0.177	A
2015	63241	0.066	0.036	0.2	0.08	0.0327	0.336	0.063	76.545	0.186	A
2016	68993	0.056	0.072	0.2	0.08	0.0718	0.348	0.105	76.888	0.234	A
2017	76121	0.065	0.062	0.2	0.08	0.0718	0.363	0.091	77.231	0.194	A

Classification effect comparison of several sorting algorithms

C4.5 algorithm, Naive Bayes algorithm and BP neural network are three commonly used classification algorithms. C4.5 algorithm is an extension and optimization based on ID3 algorithm, which selects splitting attributes through information gain rate. As a widely used classical classification method, Naive Bayesian algorithm is constructed on the basis of bayesian theorem and independent hypothesis of feature conditions. BP neural network is widely used in modern scientific research. It can realize arbitrary nonlinear mapping from input to output layer by training according to the error reverse propagation algorithm, and has strong fault-tolerant ability.

Before applying three classical classification methods to the early warning of the payment risk of basic endowment insurance, 84 total samples were randomly divided into two categories, one was training samples and the other was verification samples. The total samples were divided into two categories according to 3:1 by using the random ordinal number in MATLAB, and 63 training samples and 21 verification samples were randomly selected from the total samples.

Classification results of C4.5 algorithm

The C4.5 algorithm was programmed with MATLAB software. First, the training samples were used to generate the decision tree, and then the verification samples were used to test the results. The confusion matrix of the results is shown in Figure 1. Where classes A, B, C, and D correspond to the numbers 1, 2, 3, and 4 in the figure

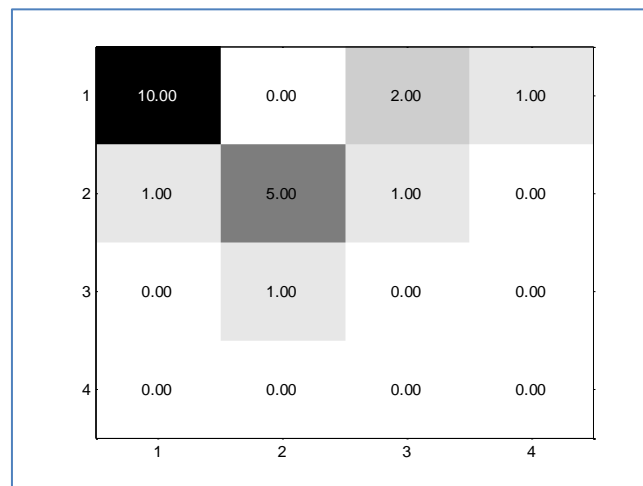


Fig-1: Confusion matrix of C4.5 algorithm classification results

It can be seen that there were 10 results when the risk degree of class A was correctly predicted as class A, and there were 2 and 1 results when the risk degree of class A was wrongly predicted as class C and class D, respectively. There were 5 results when risk degree B was correctly predicted as B, and 1 result when risk degree B was incorrectly predicted as A and C, respectively. The risk degree C class has only one in the verification sample and was incorrectly predicted as the B class. Therefore, the accuracy of the C4.5 algorithm is 71.43%.

Classification results of Naive Bayesian algorithm

Based on the principle of Naive Bayes algorithm, MATLAB software was used for programming, and the confusion matrix of the predicted results can be obtained, as shown in figure 2.

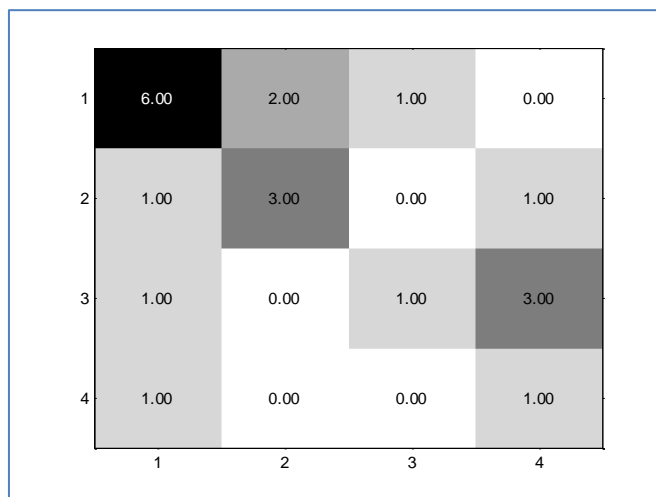


Fig-2: Confusion matrix of Naive Bayesian algorithm classification results

The risk degree class A was correctly predicted to have 6 results for class A, but was incorrectly predicted to be 2 and 1 results for class B and class C, respectively. The risk degree class B was correctly predicted to have 3 results for class B, but it was wrongly predicted to have 1 result for class A and class D, respectively. The risk degree class C was correctly predicted to have 1 result for class C, but it was wrongly predicted to have 1 result for class A and 3 results for class D. The risk degree class D was correctly predicted to have one result for class D, and it was wrongly predicted to have one result for class A. Therefore, the accuracy of Naive Bayesian algorithm is 52.38%.

Classification results of BP neural network

After continuously adjusting the parameters of BP neural network, the prediction accuracy of the model was compared when the number of hidden layers was 10, 11, 12, 13, 14 and 15, and the learning rate was 0.1. When the hidden layer was found to be 11, the BP neural network had a better effect in the early warning of the fund payment risk. The confusion matrix of the results is shown in Figure 3.

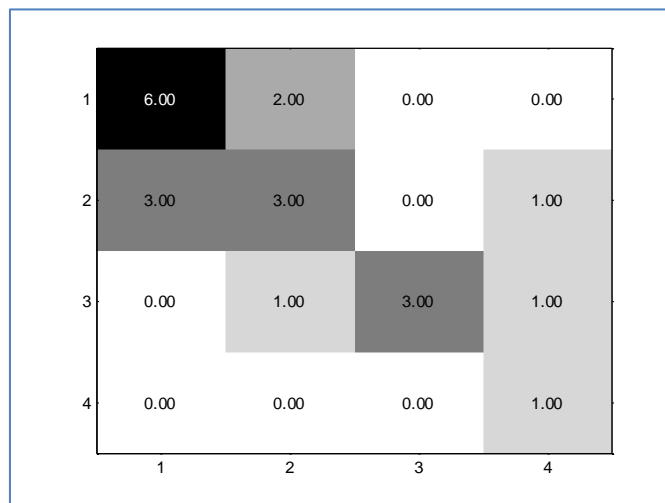


Fig-3: Confusion matrix of BP neural network classification results

The risk degree class A was correctly predicted to have 6 results for class A, but was incorrectly predicted to be 2 results for class B. The risk degree class B was correctly predicted to have 3 results for class B, but it was wrongly predicted to have 3 results for class A and one result for class D. The risk degree class C was correctly predicted as class C with 3 results, but was incorrectly predicted to have one result for class B and class D, respectively. The risk degree class D was all correctly predicted as Class D, with one result. Therefore, the accuracy of the BP neural network is 61.90%.

In conclusion, among the three classical classification algorithms, C4.5 algorithm has the best effect, with an accuracy of 71.43%; BP neural network is the second, with an accuracy of 61.90%; Naive Bayesian algorithm is poor, with an accuracy of 52.38%.

Predicting national financial status of the insurance fund

As can be seen from the above, C4.5 algorithm has the best early warning effect. Based on this best method, the early warning of payment risk of the national basic endowment insurance fund in the next 10 years can be carried out. The data of national early-warning indicators of 2019-2028 were shown in table 3. Among them, the unit contribution rate in 2019-2028 was 16%, which was obtained according to relevant provisions of "Comprehensive Scheme for Reducing Social Insurance Contribution Rates". The individual contribution rate and bookkeeping rate remained unchanged during this period. The values of other warning indicators were obtained by using linear trend extrapolation method.

Table-3: Estimated national early warning indicators for the next 10 years

year	pension contributions					pension expenditures				
	C1	C2	C3	C4	C5	P1	P2	P3	P4	P5
2019	83205	0.048	0.054	0.160	0.080	0.072	0.348	0.093	77.917	0.190
2020	88152	0.044	0.052	0.160	0.080	0.072	0.351	0.094	78.260	0.188
2021	93099	0.039	0.049	0.160	0.080	0.072	0.353	0.096	78.603	0.186
2022	98047	0.034	0.047	0.160	0.080	0.072	0.355	0.098	78.946	0.184
2023	102994	0.029	0.044	0.160	0.080	0.072	0.357	0.099	79.289	0.182
2024	107941	0.025	0.042	0.160	0.080	0.072	0.360	0.101	79.632	0.180
2025	112889	0.021	0.039	0.160	0.080	0.072	0.362	0.102	79.975	0.178
2026	117836	0.016	0.037	0.160	0.080	0.072	0.364	0.104	80.318	0.176
2027	122783	0.012	0.035	0.160	0.080	0.072	0.366	0.106	80.661	0.174
2028	127731	0.007	0.032	0.160	0.080	0.072	0.368	0.107	81.004	0.171

By substituting the early-warning indicators data in table 3 into the decision tree generated by C4.5 algorithm for prediction, the results of payment risk degree of China's basic endowment insurance fund in these 10 years can be obtained, as shown in table 4.

Table 4 Prediction results of payment risk of the national fund in the next 10 years

2019	2020	2021	2022	2023	2024	2025	2026	2027	2028
A	C	C	C	C	C	C	C	C	C

RESEARCH CONCLUSIONS

At present, the financial situation of the basic pension fund in some provinces of China has started to deteriorate. With the deepening of the aging population, the deterioration trend will be more and more obvious. Therefore, it is urgent to give early warning on the payment risk of the basic pension fund in China. This paper discusses the effect of C4.5 algorithm, Naive Bayesian algorithm and BP neural network in early warning of payment risk of basic endowment insurance fund. It is found that C4.5 algorithm has the best prediction effect, with an accuracy of 71.43%; BP neural network takes the second place, with an accuracy of 61.90%; Naive Bayesian algorithm has a poor effect, with an accuracy of 52.38%. Therefore, C4.5 algorithm is more suitable for early warning of payment risk of China's basic endowment insurance fund.

REFERENCES

1. Tan, X., & Fan, G. (2004). The Actuarial Forecasting and Estimates for the Payment Capability of Chinese Pension Insurance System in the Future, *Population & Economics*, 1, 55-58+36.
2. Yu, H., & Zhong, H. (2009). On Sustainable Operation of China's Basic Pension Insurance System: Analysis of Three Simulation Conditions. *Journal of Finance and Economics*, 9, 26-35.
3. Feng, T., & Liu, F. (2010). Study on the Forecasting of the Solvency of Urban Employees' Basic Endowment Insurance Fund, *Northwest Population Journal*, 2, 10-17.
4. Hui, A. I., Yang, Z. H. A. N. G., Chang-yu, Y. A. N. G., & Yan-dong, W. U. (2012). Research on the Financial Sustainability of the

- Pension Insurance Pooling Account in China: Measurement Based on Open System [J]. *Journal of Finance and Economics*, 2.
5. Liu, X. (2014). Study on the Financing Gap and Sustainability of China's Pension System. *China Ind. Econ*, 9, 25-37.
 6. Chen, Y. (2005). Risk and Control of Social Security Management, *China Social Security*, 9, 53-54.
 7. Cheng, L. (2009). Social Security Management Should Focus on Five Major Risks, *China Social Security*, 12, 32-33.
 8. Li, H., & Wu, Y. (2003). Introduction to the Pension Insurance Fund Income and Expenditure Early Warning System, *China Social Security*, 11, 20-21.
 9. Pan, H., & Zheng, J. (2009). Discussion on Risk Early Warning Index System of China's Endowment Insurance Fund, *Journal of Guizhou University (Social Sciences)*, 2, 31-35.
 10. Chang, L., & Chen, L. (2011). The Application of BP Neural Network in the Early Risk Warning of Basic Pension Insurance Fund Pays, *Statistics & Information Forum*, 11, 80-84.
 11. Gao, L., & Gu, J. (2014). Study on Risk Early Warning of Chongqing Basic Pension Insurance Fund, *Reform & Opening*, 9, 60-62+72.
 12. Ye, M., & Li, H. (2018). Shanghai City Endowment Insurance Fund Payment Risk Warning Based on BP Neural Network, *Chinese Journal of Gerontology*, 15, 3827-3832.
 13. Zhu, H. (2009). Risk Assessment and Early Warning Research of Basic Endowment Insurance Fund--Theoretical Model and Numerical Simulation, Southwest Jiaotong University, 2009.