

Internet Big Data Information Analysis and Power Intelligent Automation Risk Prediction Based on Case Based Reasoning

Yu-Hui Xu^{1, a}, Xiao-Yun Tian^{2, b}

¹College of Economics and Management, Hengyang Normal University, Hengyang, Hunan, 421002, China

²School of Economics and Management, Zhejiang Normal University, Jinhua, Zhejiang, 321004, China

^afrank5102@163.com, ^b1608418196@qq.com

Keywords: internet big data analysis; power intelligent automation risk prediction; NT-CBR

Abstract. The development of the Internet makes it easier to get data. However, it is more urgent to obtain effective information on the large data which is redundant. Based on the risk prediction system of power intelligent automation, this paper tries to solve the problems of enterprise risk management, and tries to make human beings free from the heavy human and mental labor. In order to solve the problem of large data analysis, this paper constructs a close triangle case based reasoning model based on the data mining technology of NT-SMOTE and the case based reasoning technology. Compare with NT-MDA, NT-Logit, NT-Probit, NT-DT, NT-SVM and other methods, the experimental results show that NT-CBR is more effective than other methods of intelligent risk prediction.

Introduction

Big data is the inevitable product of the development of the internet. Data information is generated per minute. Statistics and analysis of large data become more and more important. Depending on the traditional human and brain works, we have been unable to handle much redundant complex data, but extracting key useful information becomes more important. In order to classify and analyze the large data information quickly and accurately predict the future development of large data, this paper uses data mining technology and case based reasoning technology to carry out the relevant research work. In order to realize the intelligent processing of large data information, this paper studies the data mining technology and the case based reasoning technology, and collects the data onto the internet about small samples and large data such as Chinese tourism and hotel industry as the research sample. In order to solve the problem of financial risk management in tourism and hotel industry, we get the relevant data from the Internet, and then use the improved data processing method to balance the data, and then solve the problem of enterprise's intelligent risk management by Case based reasoning and so on. Finally, we use the case based reasoning method to establish the risk prediction system based on the power intelligent automation system to solve the basic risk management problem.

Literature Review

Manyika J, Chui M, et al. (2011) [1] pointed out that big data is one of the future direction of development. Retailers can improve operating profit margins by big data, health care industry can use big data to improve efficiency and quality, reduce medical expenses, government departments can use big data to improve the operational efficiency, personal user service data can help companies understand the market demand, big data can help collect and analyze data to make better decisions, can improve the development of the next generation of products and services. Koo D, Piratla K, et al. (2015) [2] use wireless sensor network technology to build a large data collection and analysis system which can be connected to the downstream users and the upstream data, and uses data mining technology to achieve higher levels of sustainable water supply. In order to solve

the problem of trust in the supply chain partners, Li H, Sun J, et al. (2012) [3] proposed the integrated supply chain trust diagnosis method based on inductive reasoning method, and compared with multivariate discriminant analysis, logistic regression, single euclidean case-based reasoning, and single inductive case-based reasoning, the experimental results show that this method has obvious advantages, which greatly promotes the improvement and application of case based reasoning technology. Li H, Sun J (2012) [4] proposed a method to improve the case based reasoning method to improve the classification performance of case based reasoning technology, and compared with discriminant analysis multivariate, regression logistic, the classical CBR algorithm and, it was found that the integrated case based reasoning method greatly improved the performance of prediction. In order to carry out effective classification of small sample data, Xu Y H, Li H, et al. (2014) [5] proposed a new NT-SMOTE data mining technology based on the spatial triangle and over sampling technology, which effectively solves the problem of large data distribution of small sample. Zhang M (2015) [6] studied the use of the Internet for children and adolescents through large data analysis techniques. Research shows that the Internet preference will lead to education inequality. High social status is closely related to network capital enhancement, low social status is closely related to the network entertainment. Reasonable selection and application of large data network information becomes very important. Yan A, Wang D (2015) [7] propose a retrieval-based revision method to achieve better classification performance of case-based reasoning classifiers. The use of attribute weight algorithm greatly improves the classification results of case reasoning technology, which will help to solve the problem of the credibility of the problem.

Pre preparation

In the first step, we need to collect and sort out the original real data sets we need from the Internet. This job is very time consuming and difficult. Because it is very difficult to get all useful data sets from the Internet. We get the best possible data information we need through the major financial websites and related database information providers. Effective data acquisition is a common problem faced by all of the big data researchers. This is very lucky, we have collected through various channels as soon as possible complete real data sets. In fact, we get the data information is noisy, redundant and unbalanced due to various reasons. In order to dig out the data information that we can use from the redundant data, we need to carry on the second step processing for the data information, which is the data preprocessing. In this paper, we first remove the null values of the sample and the related classification properties. We perform data preprocessing for data that is collected and sorted from the Internet, using the most commonly used three times standard deviation boundary big data mining processing method. In the end, we obtain 195 negative samples and 11 positive samples in t-3, 208 negative samples and 11 positive samples in t-2, and 221 positive samples and 11 positive samples in t-1, by data preprocessing method. Negative sample is a sample of the enterprise annual financial data which is operating in good condition. Positive sample is a sample of the enterprise annual financial data which is in trouble. The annual financial data sample is composed of 41 financial indicators. Negative samples and positive samples are used as the source case database in Case based reasoning system.

Method

NT-SMOTE is an effective method to improve the typical minority class imbalance data problem by using the idea of the adjacent triangle space. This paper addresses the problem of classification and identification of small sample large data classification problems using NT-SMOTE. Traditional classification and identification method for data sampling and classification and identification are based on the balanced data. To help solve the problem of imbalanced data classification and identification in tourism and hotel industry, this paper uses the NT-SMOTE method to carry out a balanced sampling and processing, which lays the foundation for the prediction of the intelligent classification. Case based reasoning method (CBR) is one of the most popular methods of the field

of artificial intelligence. Establish a class of empirical database according to the similarity between case classes, and to identify and analyze the new case by case base data information. In order to optimize and improve the classification effect of intelligent classifier, this paper combines NT-SMOTE method and CBR method to solve the problem of risk identification and assessment. In order to compare the advantages of the improved case based reasoning method, NT-CBR, the NT-SMOTE method and MDA, Logit, Probit, DT and SVM are combined to construct a new classification recognition model: NT-MDA, NT-Logit, NT-Probit, NT-DT, and NT-SVM. So as to understand the advantages and disadvantages of various classification models on enterprise risk identification and evaluation.

Model

This paper is to build an intelligent risk forecasting system for power intelligent automation with NT-CBR technology, to solve the huge data analysis task faced by enterprises. The core of this paper is NT-CBR technology. According to the 4:6 ratios, We divide the data sample into the test case database and the source case database, which are processed by data preprocessing. In order to obtain a balanced distribution of the source case database, this paper uses the NT-SMOTE method to simulate the data simulation. In order to reduce the spatial dimension of sample data, this paper uses the normal distribution test, significant difference test, multiple linear test and other methods to reduce the sample financial data, and excludes those of the impact on the forecast results are not obvious indicators of data. The purpose is to build a financial index system which can reduce the accuracy of financial risk. In order to guarantee the stability and reliability of the experimental results, the 30 random partitioning and experiment were performed repeatedly. We establish risk identification assessment models using CBR, MDA, Logit, Probit, DT, SVM and other methods, and conduct experiments to evaluate and analyze the financial risks of enterprises. Schematic diagram of NT-CBR and other methods shown in Fig 1.

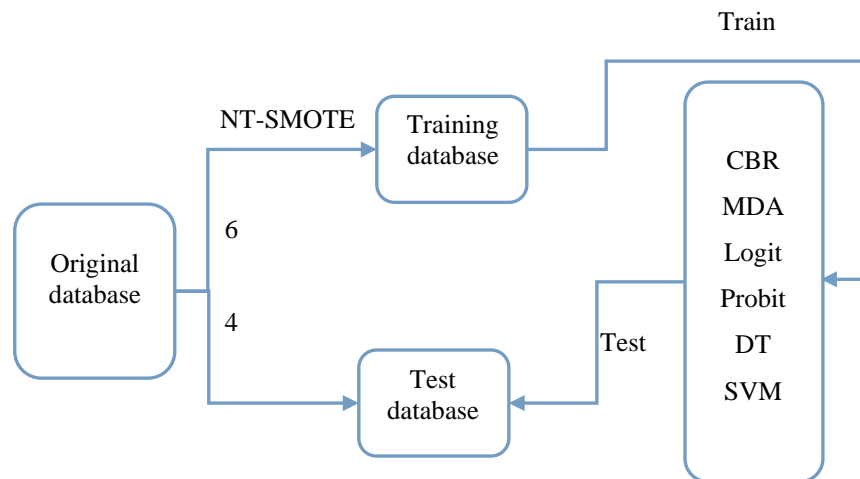


Figure 1. Schematic Diagram of NT-CBR and Other Methods

Case based reasoning is like a basic model of human learning and thinking, which can be used to infer the new case by using the similar principle. Case based reasoning method is helpful to improve the intelligence assistant level and the system's flexibility and adaptability, and can overcome the uncertainty of the traditional decision making, in order to reduce the decision-making risk, and enhance the robustness and intelligence. MDA is strict with the requirements of sample data, and the change of the variable has a greater impact on the identification results. Logit and Probit are used to solve the sample classification problem using different distribution functions. Both of the changes in the data sample variables do not have a huge difference, and the operation efficiency is faster. Decision tree is a kind of mapping relationship between the object attributes and object values, which is a kind of intelligent forecasting method which can be used to test the classification of test samples. However, it is difficult to predict and classify in the continuous

density or tight time series data. Support vector machine is a kind of classification and prediction algorithm, which has a distinct advantage in the high dimensional nonlinear data, which can effectively map the high dimensional data onto a higher dimension. However, it has poor performance in multi-classification forecasting.

Research results

In real life, the data size is increasing, the constraint conditions are increasing, the nonlinear is serious, the environment is more complex, which makes the system more difficult to be optimized. In order to better to compare the advantages and disadvantages of various optimization methods, this paper compares the advantages and disadvantages of various optimization methods of the repeated 30 times. Experimental results are shown in Fig 2, Fig 3, Fig 4 and Table 1

The experimental results show that the NT-MDA has excellent classification performance and good stability in the case of variable data. NT-Logit and NT-Probit data classification results are poor in different sampling conditions, and the sensitivity of risk identification is poor. Only the enterprise data deterioration index reached a certain threshold, variables, classification performance can be effectively improved. The classification performance of NT-DT is generally poor, the stability is also poor. The performance of NT-SVM is also general, the classification result is changed greatly. The performance of NT-CBR is excellent in classification performance, stability and sensitivity. From the experimental results, NT-CBR and NT-MDA showed superior classification performance in a small number of sample identification, while other methods performed poorly. NT-CBR and NT-MDA are less than other methods in most of the classification and identification of samples. Therefore, a case based reasoning method is superior to the classification of large data.

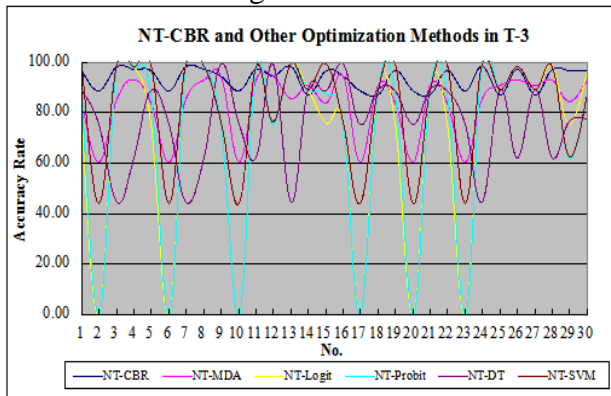


Figure 2. NT-CBR and Other Methods in T-3

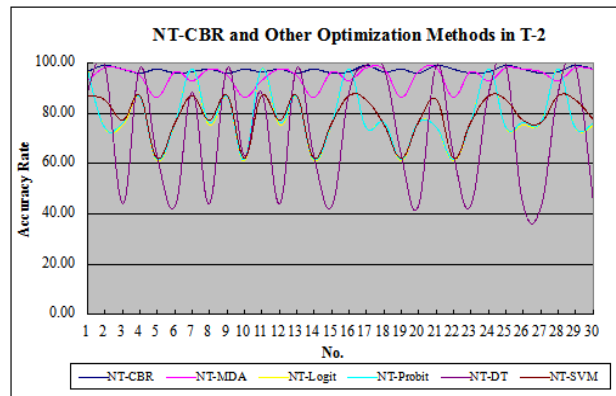


Figure 2. NT-CBR and Other Methods in T-2

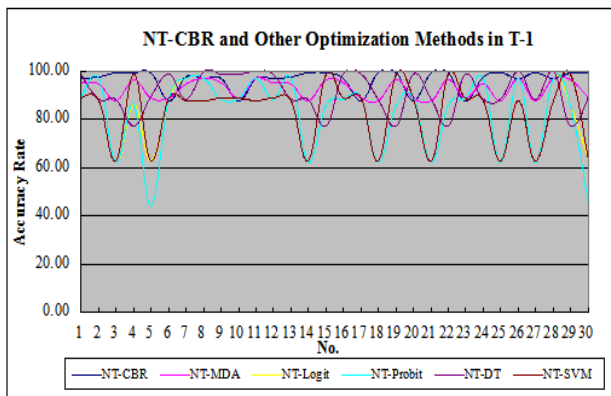


Figure 2. NT-CBR and Other Methods in T-1

Table 1. Classification Performance Evaluation System of Various Methods

Time	t-3			t-2			t-1		
Methods \Indices	Sensit ivity	Specit ivity	G	Sensiti vity	Speciti vity	G	Sensit ivity	Specit ivity	G
NT-CBR	92. 67	94. 83	93. 56	100. 0	94. 46	97. 19	96. 67	96. 36	96. 44
NT-MDA	81. 33	88. 10	83. 49	96. 67	92. 80	78. 57	91. 33	94. 04	83. 35
NT-Logit	64. 67	97. 36	70. 46	66. 67	94. 52	78. 76	74. 67	95. 76	82. 13
NT-Probit	66. 67	97. 47	71. 67	66. 67	95. 05	94. 61	73. 33	95. 76	92. 45
NT-DT	62. 00	97. 87	76. 15	59. 33	97. 04	72. 38	85. 33	97. 83	91. 01
NT-SVM	74. 00	98. 22	82. 59	66. 00	95. 97	78. 97	72. 67	97. 42	83. 08

Summary

To solve and optimize the process of enterprise management, improve the level of intelligence, it is an important issue in management theory and practice. Internet has entered the era of big data information. How to quickly obtain the effective data information, and provide intelligent decision support, is becoming more and more important in the rapidly changing environment. Case based reasoning technology for the Internet big data mining and analysis, for us to further promote the intelligent innovation, science and technology to change the life provides a reference and reference. Based on NT-CBR technology, the retrieval efficiency is greatly improved, the accuracy and stability of the system is also improved. Case based reasoning technology has important research value and prospect for the field of intelligent science and technology in the future. In this paper, a decision system based on intelligent automation of electric power is established using NT-CBR Method. We hope this study can play a role in the future development of intelligent machine system. We hope that this study will broaden the idea of the researcher, provide a reference to optimize the design of the power intelligent automation system, and to provide the theory and practice based on the application of case based reasoning technology.

References

- [1] Manyika J, Chui M, Brown B, et al. Big data: The next frontier for innovation, competition, and productivity[J]. 2011.
- [2] Koo D, Piratla K, Matthews C J. Towards Sustainable Water Supply: Schematic Development of Big Data Collection Using Internet of Things (IoT)[J]. Procedia Engineering, 2015, 118: 489-497.
- [3] Li H, Sun J, Wu J, et al. Supply chain trust diagnosis (SCTD) using inductive case-based reasoning ensemble (ICBRE): The case of general competence trust diagnosis[J]. Applied Soft Computing, 2012, 12(8): 2312-2321.
- [4] Li H, Sun J. Case-based reasoning ensemble and business application: A computational approach from multiple case representations driven by randomness[J]. Expert Systems with Applications, 2012, 39(3): 3298-3310.
- [5] Xu Y H, Li H, et al. Neighborhood Triangular Synthetic Minority Over-sampling Technique for Imbalanced Prediction on Small Samples of Chinese Tourism and Hospitality Firms[C]//Computational Sciences and Optimization (CSO), 2014 Seventh International Joint Conference on. IEEE, 2014: 534-538.
- [6] Zhang M. Internet use that reproduces educational inequalities: Evidence from big data[J]. Computers & Education, 2015, 86: 212-223.
- [7] Yan A, Wang D. Trustworthiness evaluation and retrieval-based revision method for case-based reasoning classifiers[J]. Expert Systems with Applications, 2015, 42(21): 8006-8013.