Risk Identification Method of Enterprise Accounting Information Fraud Based on Weighted Association Rule Algorithm

Jianhong Li [1], Tong Yao [2*]

[1]Sichuan University Jinjiang College. Meishan City, Sichuan Province, China. 620860
[2]Sichuan Tourism University. Chengdu, Sichuan, China.610100

Yaotong2011tony@163.com

Abstract. In order to reduce the financial risks existing in enterprises, this paper studies the risk identification of enterprise accounting information fraud with weighted key algorithm. Based on the study of fuzzy association rules, this study analyzes the financial risks of enterprises, puts forward an analysis model of enterprise financial risks based on fuzzy association rules mining algorithm, and makes a qualitative analysis of enterprise financial crises, so as to predict enterprise financial risks and provide reference for enterprises to avoid risks. Comparing the correlation algorithm with the linear regression model, the result shows that the prediction error rate of the linear regression model is higher than that of the correlation algorithm model. Class I misjudgment rate is 40% higher, class II misjudgment rate is flat, overall misjudgment rate is 20% higher, and accuracy linear regression is 20% lower than correlation algorithm. It can be seen that the misjudgment rate of the correlation algorithm model is low, the prediction accuracy is high, and the discrimination and prediction ability of the linear regression model is poor, which proves the effectiveness of the application of the correlation algorithm in this paper.

Keywords: association rule algorithm; Accounting fraud; Fraud risk identification

1 Introduction

Accounting fraud refers to engaging in accounting and financial related work, in order to obtain personal illegal income, through modifying accounting data to achieve the goal, at the expense of violating accounting professional ethics, violating the objective authenticity of accounting information, regardless of national laws and regulations and relevant system norms. While this kind of behavior hinders the development of enterprises, it disturbs the market order and damages the national economy [1]. For listed companies, preventing accounting fraud can reduce their mistakes in investment decisions, and for government departments, it also helps them to supervise and manage listed companies and investment markets. Only when the operating environment of the market is well managed can it help to increase investors' information. Therefore, it is

H. Wu et al. (Eds.): CSIEDE 2022, ACSR 103, pp. 647–655, 2022.
https://doi.org/10.2991/978-94-6463-108-1_72

necessary to prevent accounting fraud, which will be beneficial to the development of listed companies [2]. In addition, listed companies can disclose information to the outside world in regular or irregular time, which can restrain accounting cheating to some extent, but the quality of accounting information of some companies is not good when disclosing. It is therefore necessary to establish a scientific and appropriate basis for the quality of accounting data in order to assess the content of data that sets companies and prevent accounting fraud on a justified basis. Financial reporting is a vehicle for the company to convey information about its business business to the outside world, and it is a list from the company to reflect the financial situation of the company at a time, operating results and cash flow for a billing period [3]. Financial report is constantly evolving and developing with the development of social and economic environment. Compared with the financial report in the early 20th century, today's financial report has undergone tremendous changes. At the same time, due to the development and changes of the social and economic environment, new transactions and events are constantly appearing, and the current financial reports are increasingly exposed to drawbacks and shortcomings, so they need to be constantly adjusted and improved. It can be predicted that today's financial report will change greatly in the near future [4]. For the convenience of research, this study will focus on the problem of financial report fraud under the current financial accounting statement system. The current financial reporting system consists of balance sheet, profit statement, cash flow statement and its schedules and notes. Concepts related to financial report fraud include accounting fraud, accounting information distortion and earnings management. There is some confusion in the academic field because of the similarity in words and the correlation in fact of the concepts of financial report fraud, earnings management and accounting information distortion. Therefore, this paper clearly explains the relationship among them. Accounting fraud refers to the company's subjective practice of falsifying and concealing false accounts, and it is the behavior that the accounting subject arbitrarily violates accounting standards and tramples on the relevant accounting laws and regulations of the state to create false accounting information in order to achieve its subjective purpose [5]. Obviously, the meanings of accounting fraud and financial report fraud are similar.

2 Literature review

Kakaria, A. et al. said that the re-identification of records and the risk of inferring sensitive information are important factors in the process of anonymous technology and parameter decision. If properly evaluated, measuring the risk of re-identification and reasoning will help to optimize the balance between data protection and practicality, because too aggressive anonymization will make data unusable, and it is troublesome to publish data with high risk of de-anonymization. This method proposes a new privacy measure (ITPR) based on information theory to evaluate the risk of re-identification of data sets and inference of sensitive information. The proposed indicators are compared with the existing information theory indicators and their ability to assess the risks of different situations characterizing data sets. The results show that

ITPR is the only index that can effectively quantify the risk of re-reporting and inferencing sensitive information [6]. The purpose of Jiang, G.J. et al.' s research is to check whether the earnings risk in the quarter at the end of fiscal year (FYE) increases compared with other quarters of the same company, and more importantly, whether this earnings risk is unique. In addition, these drugs can reduce the risk of risk. The evidence from the study shows that data at risk for quarterly FYE returns will not be explained by other risks. Solutions to mitigate this will strengthen corporate governance and a tighter financial disclosure structure. The investigation shows that from 1984 to 2015, the coefficient reaction of FYE's quarter was significantly lower than that of the non-FYE quarter. In addition, strong corporate governance and a streamlined financial disclosure model could help reduce uncertainty associated with such data, whether by voluntarily reducing the use of personal data from companies or reforming the FASB Code (such as FASB 145) [7]. The goal of Mohammad I, S., et al.'s research is to examine the impact of corporate responsibility (CSR) of Iranian listed companies on the financial accounting scheme, including stock returns, fund managers, asymetry data, and financial performance. This is descriptive correlation and demand for research. The population statistics of the study include all the companies listed on the Tehran Stock Exchange, and the research extends from 2012 to 2018. Samples from 150 companies were selected in the screening process. More regression and Eviews 10 software are used for data analysis and hypothesis testing. The results show that corporate social responsibility has a significant impact on stock returns, but not for income management. Corporate responsibility has a big impact on asymmetry data and financial performance. The study was the first study on the concept of corporate responsibility and finance in Iran [8].

3 Methods

The first algorithm has an antimonotony of the model, and the grid model is often used to collect all the products. The flow of the pre-algorithm is available in figure 1 [9]. In addition, a data with D differences will simultaneously generate 2D frequency products and R rights, including:

$$R = \sum_{k=1}^{d-1}\left[\binom{d}{k} \times \sum_{j=1}^{d-k} \frac{d-k}{j}\right] = 3^d - 2^{d+1} + 1 \tag{1}$$

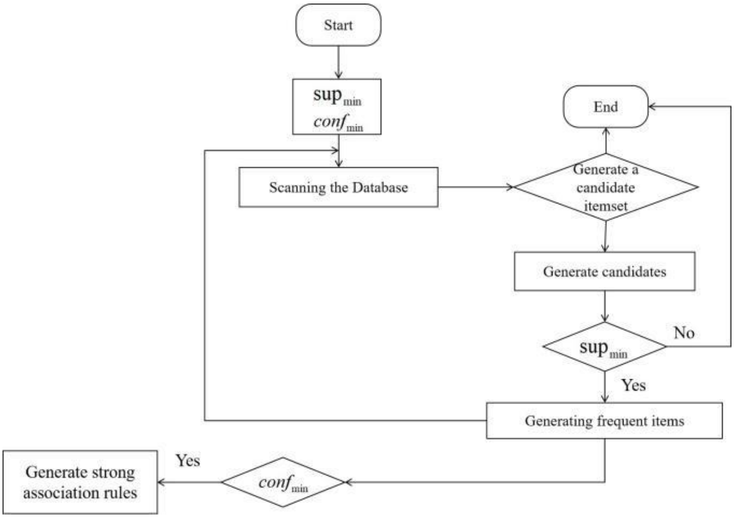

Fig. 1. Apriori algorithm flow

In order to test the influence of financial indicators for financial companies, the indicators of profitability, management, growth, solvency and cash flow among companies are chosen in this form. On this basis, following the correlation analysis of financial indicators, some correlated financial indicators are eliminated and the model is simplified [10]. The correlation coefficient of each bond index can be determined by formula:

$$r_{x,y} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - \left(\sum x_i\right)^2}\sqrt{n\sum y_i^2 - \left(\sum y_i\right)^2}} \quad (2)$$

Where: x and y are two variables, and $r_{x,y}$ is the correlation coefficient of the variables, satisfying $-1 \le r_{x,y} \le 1$. When $\left|r_{x,y}\right| = 1$, x is completely linearly related to y; When $r_{x,y} = 1$, x perfect positive correlation in y; When $r_{x,y} = -1$, x is completely negatively correlated with y; When $r_{x,y} = 0$, x and y are not related; When $-1 < r_{x,y} < 1$, x and y have a linear relationship [11].In addition, exclude the indicators that have good or negative correlation to reducing the collinearity of financial indicators. Reducing the deviation from the company's financial risk analysis, it is necessary to clean up data collection data and eliminate the abnormal values of all financial indicators. At the same time, as the next mining agency policy, it is necessary to discretize the continuous index data based on the level of funding. Considering that the financial data set consists of financial data from many companies, the variables of financial indicators are regularly distributed [12-13]. Therefore, the area's division is used in the wings to discretize continuous variables. According to the quantiles 1/5, 2/5, 3/5, and 4/5 of each variable distribution, each silver index variable division makes five degrees, each described as blind level (grade 1), slowing levels (level 2), level of the wrong behavior (grade 3), crisis level (grade 4), and death level (grade 5).

The pre-algorithm, which is based on generations and test model candidates, is used to determine the model's content. First, the candidate's model set is processed in parallel and the time the series's data is received [14]. Secondly, the continuous attributes of the data are discretized to determine the relevant fuzzy attribute data set. The specific process of the algorithm is described as follows. Input: minimum fuzzy support $\sup_{\min}$ minimum fuzzy trust $conf_{\min}$; Output: the association rule set is S_{ar}

Step1, making the parallel processor p_1, p_2, L, pn;

Step2, dividing the database into a plurality of partitions and distributing them to different processors;

Step3, clustering different processors by using fuzzy FCM clustering algorithm, and generating new data sets; meanwhile, discretizing attributes to obtain time series, and constructing a document tree according to $\sup_{\min}$ and $conf_{\min}$;

Step 4, perform counting operation on each local processor;

Step 5: Calculate the global count according to the local count and generate an output rule set.

As time constraints fulfilled by the front and back attributes of the law, the rules are filtered out to get the right time. Use the evolution of the law to determine the degree of corporate crisis and conduct a qualitative analysis of the corporate financial crisis [15]. Ultimately, by calculating the coefficient crisis, the crisis stage of the company is determined and the quantitative analysis of the financial crisis of the company is carried out. The rare precursors and high precursors of the laws exacerbate the human crisis; Otherwise, the crisis could be reduced. If the law is always in the first phase, the corporate crisis is relatively small; If in phase three, the crisis is a little different; If it's in the fifth phase, the company is on the verge of bankruptcy. In addition, the coefficient crisis is introduced to calculate the degree of corporate financial crisis:

$$F = F\left(\bar{x}_i, \sup\left(x_i\right), conf\left(x_i\right)\right) = \frac{1}{n}\sum_{i=1}^{n} \bar{x}_i conf\left(x_i\right) \tag{3}$$

Where: n is the number of rules, and $\bar{x}_i$ is the data variable after discretization.

The question is considering a company listed as a product of research and takes annual and quarterly data from 2003 to 2020 as a data center and runs 32 financial indicators [16-17]. Classify data samples from financial indicators and eliminate abnormal values. At the same time, the correlation between financial indicators is calculated using previous theory, and financial indicators with a simple correlation coefficient in the group's financial indicators are chosen, and the indicators are discretized based on the level of funding, and the reconstructed database of financial indicators is preserved. In addition, each company's revenue ratio is summarized as the daily series in 12 quarters, and the discretized data set is used as input on the agency's policy-only algorithm. In addition, policy organizations are used to weather revenue ratios and warn of crises at an early stage.

4 Results and analysis

The selection of indicators is based on the signs of fraud, combined with the characteristics of listed companies in China capital market and the characteristics of fraud, and taking into account the availability of indicators. Finally, according to the classification of fraud identification variables, the final selected indicators are divided into two categories: one is the financial indicators that discover the possibility of violations; The other is the index to measure the motivation or ability of the company to violate the rules [18-19]. The research results of the model construction show that there are related indicators that can be used to identify the risks of financial report fraud, which makes the model construction significant, and then financial report users can use the model to measure and avoid risks, making contributions to standardizing the efficiency of the capital market. The significant statistical indicators of identifying fraud risk in linear regression algorithm model and correlation algorithm model can be summarized as follows. Table 1 shows the comparison of model discrimination effect.

Table 1. Comparison of discriminant effects of models

Model construction	Class I misjudgment rate	Class II misjudgment rate	Overall misjudgment rate	accuracy rate
Linear regression	31.25%	12.90%	22.22%	77.78%
Association algorithm	9.375%	6.45%	7.94%	92.06%

As can be seen from Table 1, compared with the discriminant effect of the models, the misjudgment rate of the linear regression model is generally higher than that of the correlation algorithm model. The misjudgment rate of Class I is 21.875% higher, that of Class II is 6.45% higher, that of the whole is 14.28% higher, and that of the accuracy of linear regression is 14.28% lower than that of Logistic regression. Among them, the misjudgment of Class I is that fraudulent financial reports are judged as non-fraudulent financial reports. Class II misjudgment refers to the judgment of non-fraudulent financial reports as fraudulent financial reports, and identifying the mistakes made by the model with two kinds of misjudgments will bring different costs and different degrees of losses to auditing and decision-making [20]. First of all, judging fraudulent financial reports as non-fraudulent financial reports may lead to incorrect decisions made by financial report users, and the decisions made through financial reports will be deviated and lead to mistakes, resulting in serious economic losses and reputation losses. Secondly, judging non-fraudulent financial reports as fraudulent financial reports will lead to additional investigations by regulatory authorities and internal audit departments at the cost of manpower, financial resources and material resources, and may make financial report users lose better investment opportunities, resulting in loss of opportunity costs. Table 2 shows the comparison of the prediction effects of the models.

Table 2. Comparison of prediction effects of models

Model construction	Class I mis-judgment rate	Class II mis-judgment rate	Overall mis-judgment rate	accuracy rate
Linear regression	40%	10%	25%	75%
Association algorithm	0	10%	5%	95%

It can be seen from Table 2 that compared with the prediction effect of the models, the prediction error rate of the linear regression model is higher than that of the correlation algorithm model, with class I error rate being 40% higher, class II error rate being equal, the overall error rate being 20% higher, and the accuracy rate of linear regression being 20% lower than that of the correlation algorithm. It can be seen that the correlation algorithm model has low misjudgment rate, high prediction accuracy, and poor discrimination and prediction ability of linear regression model. Therefore, in practical application, the correlation algorithm model with high accuracy can be selected for application, and other methods that help to identify financial report fraud can be used for comprehensive evaluation, so as to improve the prediction accuracy and reasonably avoid the misjudgment rate.

5 Conclusion

The purpose of this paper is to analyze which variables or indicators can reflect the signs of fraud and the characteristics of fraudulent companies, and establish a model to identify financial report fraud on this basis, so as to improve the ability of financial report users to identify fraud. The results show that fraudulent companies have higher debt ratio and other receivables than non-fraudulent companies, worse short-term solvency, lower accounts receivable turnover rate, adjusted cash per share and equity concentration, and it is easier for certified public accountants to issue non-standard unqualified opinions. Among them, the symbol of ownership concentration obtained by empirical test proves that moderate ownership concentration can produce the effect of interest convergence, which proves the popular saying that equity dispersion is always required. In the other step of demonstration, the selected effective indicators will be constructed into an association algorithm model, and a fraud identification model with low cost and good effect will be obtained.

References

1. Lestari, R. . (2020). Analysis of quality of management accounting information system to improve the implementation of good corporate governance principles (case of sharia commercial banks in bandung city), 32(6), 154-158.
2. Shbail, M. , Al-Trad, E. B. , Alshurafat, H. , Ananzeh, H. , & Kurdi, B. . (2021). Factors affecting online cheating by accounting students: the relevance of social factors and the fraud triangle model factors. Data in Brief, 55(17), 69-74.

3. Tapanjeh, A. , & Altarawneh, A. R. . (2020). Applicability of Forensic Accounting to Reduce Fraud and its Effects on Financial Statement of Jordanian Shareholding Companies From the Perspective of Judiciary and Certified Public Accountant, 116(8), 21-26.
4. Wang, M. , Zhao, W. , & Zhang, W. . (2022). Can reform of information disclosure by an exchange restrain corporate fraud? evidence from china. Asia-Pacific journal of financial studies(2), 51-54.
5. Bkakria, A. , Cuppens, F. , Cuppens, N. , & Tasidou, A. . (2021). Information theoretic-based privacy risk evaluation for data anonymization. Journal of Surveillance, Security and Safety, 2(1), 83-102.
6. Jiang, G. J. , Zhu, K. X. , & Chen, L. H. . (2022). Information uncertainty of fiscal year end quarter earnings. Review of Accounting and Finance, 21(2), 83-108.
7. Mohammadi, S. , & Saeidi, H. . (2022). Corporate social responsibility and financial accounting concepts: evidence from an emerging market. Sustainability Accounting, Management and Policy Journal, 13(3), 740-764.
8. Ngugi, B. K. , Hung, K. T. , & Li, Y. J. . (2022). Reducing tax identity theft by identifying vulnerability points in the electronic tax filing process. Information & computer sceurity(2), 30-36.
9. Guo, J. , Guo, J. , Yang, C. , Wu, Y. , Sun, L. , & Guo, J. . (2021). Shing: A Conversational Agent to Alert Customers of Suspected Online-payment Fraud with Empathetical Communication Skills, 6(10), 19-24.
10. Liu, S. . (2021). Analysis of Accounting Fraud in Enterprises Based on Game Theory——Take the Case of Geeya Technology Company Limited Fraud as an Example, 06(19), 17-24.
11. Wang, Y. , Yu, M. , & Gao, S. . (2022). Gender diversity and financial statement fraud. Journal of accounting and public policy(2), 41-46.
12. Hwa, B. , Yca, C. , Jl, D. , & Xz, A. . (2022). Financial fraud risk analysis based on audit information knowledge graph - sciencedirect. Procedia Computer Science, 27(1), 3-9.
13. Heneke, E. , Valentine, R. , & Jourdan, Z. . (2021). Predictive factors in financial fraud and malfeasance from 1950-2018. Journal of business and economic perspectives (1), 48-52.
14. Nan, T. , Bimpong, P. , Ansah, R. K. , Danso, E. , Tabiri, A. , & Arhin, I. . (2020). Predicting bankruptcy and financial fraud of both listed and unlisted commercial banks in ghana using 3 z-score models. (12), 67-71.
15. Verykios, V. S. , Stavropoulos, E. C. , Zorkadis, V. , Katsikatsos, G. , & Sakkopoulos, E. . (2022). Sensitive data hiding in financial anti-fraud process. International journal of electronic governance(1/2), 14-19.
16. Khersiat, O. M. . (2020). The impact of joint audit on fraud detection in financial statements from the point of view of auditors. (1), 2-6.
17. Wang, H. B. , & Gao, Y. J. . (2021). Research on parallelization of apriori algorithm in association rule mining. Procedia Computer Science, 183(4), 641-647.
18. Xma, B. , Hao, S. , Xz, C. , & Jl, A. . (2022). Financial fraud detection using the related-party transaction knowledge graph - sciencedirect. Procedia Computer Science, 97(4), 114-118.
19. Zhamiyeva, R. M. , Sultanbekova, G. B. , Abzalbekova, M. T. , Zhakupov, B. A. , & Kozhanov, M. G. . (2022). The role of financial investigations in combating money laundering. International journal of electronic security and digital forensics: IJESDF(2), 14-19.
20. Souza, D. D. , & Jr, C. B. . (2021). Ensemble and mixed learning techniques for credit card fraud detection, 20(1), 71-77.