S.I.: SPIOT 2020

Early warning of enterprise finance risk of big data mining in internet of things based on fuzzy association rules

Hongyu Shang[1] · Duan Lu[2] · Qingyuan Zhou[3,4]

Abstract

As the big data, Internet of Things, cloud computing, and other ideas and technologies are integrated into social life, the big data technology can improve the corporate financial data processing. At the same time, with the fiercer competition between enterprises, investors and enterprises have paid more attention to the role of financial crisis warning in corporate management. The work selected the multiple financial indicators based on big data mining in Internet of Things. The rules between all financial indicators were found to choose more representative financial risk indicators. Then the frequent fuzzy option set was determined by FCM (fuzzy cluster method), parallel rules, and parallel mining algorithm, thus obtaining the fuzzy association rules that satisfy the minimum fuzzy credibility. Finally, the relevant data of listed companies were selected to analyze the corporate financial risks, which verified the method proposed in the work.

1 Introduction

The rapid economic development of science and technology has pushed the humankind into an era of information explosion. In 2012, the Obama administration announced a high-profile plan called "Big Data Research and Development," and big data became a hot word. Big data, also known as huge amount of data, refers to collecting, storing, analyzing, and processing data of different structures and types in a more cost-effective and efficient manner in the age of information explosion. Besides, it can obtain relevant information with decision value.

In recent years, with the development of society and the widespread application of Internet of Things, the big data has been applied to various fields due to its diverse types, fast acquisition speeds, low acquisition and storage costs, and wide sources. It enables companies to achieve efficient and high-quality financial management, with advantages including increasing the use value of financial data, improving the processing of financial data, and providing managers with useful information for decision-making.

The financial status of enterprise is the focus of all the stakeholders of enterprise including operators, corporate creditors, and investors. In a fierce market competition, no enterprise can avoid risk. Financial crisis early warning and risk aversion should be considered in making market management decisions, credit ratings of financial and insurance industry, and making investment decisions.

Under the big data environment, a series of changes have taken place in the financial risk management of companies. The broader access to information enables companies to have more references when making decisions; the faster access to information enables companies to focus on real-time information affecting financial risk. The cloud computing, mobile computing and other technologies have improved the financial analysis. At the same time, the big data is of great significance to the early warning of corporate financial risks. Companies use the data mining to pick out valuable information from massive amounts of

✉ Qingyuan Zhou
  zqy2131@czimt.edu.cn

[1] Nanjing Zijin Huicai Technology Co., Ltd, Nanjing 210093, China

[2] Business School, Nanjing University, Nanjing 210093, China

[3] School of Economics and Management, Changzhou Vocational Institute of Mechatronic Technology, Changzhou 213164, China

[4] Changzhou Key Laboratory of Industrial Internet and Data Intelligence, Changzhou 213164, China

data, which is analyzed to predict potential risks. This method has been applied to the financial management, helping managers to grasp the current status of companies, and take timely measures to reduce the possibility of financial risks with minimized losses.

From the US subprime mortgage crisis to the European debt crisis, the research on financial risk crisis warning is particularly important. In harsh market competition, the requirements of companies for risk management are increasing. The study of corporate financial risk originated in the 1930s, and Fitzpatrick [1] used the single variable analysis to predict the crisis. The use of a single financial indicator for prediction has created a precedent for empirical research on early warning of financial crisis. In the 1960s, Atman, American scholar, used the multiple variable analysis model to discuss the early warning of corporate financial crisis. Besides, he used the Z-score model to establish a multivariate linear function formula to predict financial crisis [2]. In the 1970s, Meyer and Pifer [3] used the LPM (linear probability model) to analyze the financial crisis warning of banking industry. Linear probability model is a special case of multivariate analysis model to estimate the probability of business failure. Later, Laitinen et al. in [4] applied the LPM to corporate financial crisis.

Huaiyi Zhu and Yong Gao in 2002 introduced the artificial neural system to crisis early warning system. Moreover, they designed a crisis early warning system of core competency strategy for timely detection and anticipation of deviations in order to implement control and ensure the sustainability and efficiency of core competency strategies [5]. Hua Ren and Xusong Xu used the fuzzy optimization and BP neural network to theoretically derive and design the data analysis subsystem and alarm subsystem, and constructed a crisis warning index system [6]. Yingyu Wu et al. established a corporate financial crisis identification system based on financial and non-financial perspectives, using principal component analysis and neural network technology [7]. Yang et al. introduced the Benford's law into the financial-risk early-warning Logistic model. It increases the effective variables representing the quality of financial data to improve the prediction accuracy of the early warning model [8]. The use of genetic algorithm models to predict corporate bankruptcy is based on the optimization of parameters constrained in many aspects. Condition ratios and qualitative variables of financial indicators can be used for the conditional discrimination and rule extraction, with clear structure [9].

Koyuncugil AS and Ozgulba N used the neural network technology to predict financial risks. A financial risk warning model is established, including five financial ratio indicators. The accuracy of prediction results has been significantly improved [10]. Banerjee, Arindam, Prachi

et al. used three multivariable models, SVM and ANN to study the financial risks of companies. The financial risk estimated by SVM is closest to the real result [11]. Cao introduced the relevant variables of option pricing model during the establishment of the enterprise financial risk early-warning model, focusing on the option variables related to corporate financial risk [12]. Xiao, Yang, Pang et al. used the genetic algorithm and neural network method in the construction of financial risk early warning system, which greatly improves the accuracy of prediction results. The system has gradually become a dynamic system [13].

Fang, Shyng, Lee et al. applied the rough set method in predicting corporate financial risk, which improves the speed of data analysis. It provides time and space for companies to choose how to solve financial risks [14]. Maimon and Rokach reiterated the concept of data mining and identified the high-density, time-efficient, and easy-to-understand information that can provide basis for enterprise managers' decisions from massive data. Selecting and storing information belongs to the generalized data mining [15].

Biao Song, Jianming Zhu, and Xu Li used the information collected through Internet for emotional analysis and processing, based on which a financial risk early warning model was established. The financial risk predicted by this model based on big data has a small deviation from the actual risk [16]. Liang Zhang, Lingling Zhang, and Yibing Chen used the logistic regression model and SVM model and established a financial risk early warning model based on information fusion. This method has improved the feasibility of the early warning model, and its accuracy is far higher than one of the above methods alone [17].

With the rapid development of capital market, the requirements of companies for risk management are increasing. It has become a research hotspot and a difficult point to evaluate the financial risks in enterprise management and provide timely warnings. For the foregoing studies, there are problems such as many assumptions, inability to handle massive data, failure to consider the time continuity of financial indicators, and failure to track the fluctuations and trends of financial indicators. The work analyzed the corporate financial risk based on fuzzy association rules and dynamic maintenance methods. First, for the time series generated by a complex system, we studied the correlation characteristics of the internal or local morphology of time series, and the division boundaries of the time domain attribute universe were softened through fuzzy clustering algorithms. Then, an improved parallel mining algorithm of Boolean attribute association rules was used to find the frequent fuzzy attribute sets and determine the fuzzy association rules. Based on this, an

enterprise financial risk analysis model was established based on fuzzy association rule mining algorithm.

2 Problem description

Association rules mining is to find the correlation between different items in the same event [18, 19]. The used mining strategy includes the generation of frequent itemsets and rules. The former is to find all frequent itemsets that meet the minimum support threshold, and the latter is to extract the rules with high confidence from frequent itemsets. These rules are called strong rules. Association rule is an important topic in data mining, and people have done a lot of work [20]. At present, Apriori and FP_Growth are representative in association rule mining algorithms.

2.1 Apriori algorithm based on candidate pattern generation and testing

Based on the support-confidence framework, the Apriori [21] algorithm proposed by Agrawal et al. uses the iteration to generate frequent pattern sets of all lengths. The Apriori algorithm has the anti-monotonicity of frequent patterns, and the lattice structure is often used to enumerate all possible itemsets. A data set containing d different items may simultaneously generate 2^d frequent itemsets and R rules:

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

An original method for finding frequent itemsets is to determine the support count for each candidate itemset in the lattice structure. If an itemset is in frequent pattern, all its subsets must also be the same, which is called the anti-monotonicity of frequent pattern. Conversely, if the selection is infrequent, the entire subgraph containing the selection can be immediately pruned. In the generation of rules, the support degree indicates the probability of simultaneous occurrence of itemsets A and B in the database, with a certain statistical significance. Confidence indicates the probability when itemsets A and B occur simultaneously. Besides, it stands for the strength of rule.

L_k and C_k are supposed to be the frequent pattern set and candidate pattern set with the length of k, respectively. The database is scanned to generate the candidate 1-itemset C_1. Then the anti-monotonicity of Apriori is used for pruning after support count comparison, with frequent 1-itemset L_1 generated. Frequent 1-itemsets are linked with themselves to generate candidate 2-itemset C_2, and the pruning is performed after comparison. Frequent itemset L_k with the

length of $k\,(k \geq 1)$ is obtained until no more frequent itemsets are generated [22].

2.2 Time series data mining

Time series data mining is an in-depth study of the advancement of things by analyzing the time characteristics of data. Knowledge is obtained from the data with time characteristics. A large amount of time series data is used to extract potential, unpredictable rules, which are closely related to time characteristics. These rules can be used to predict the short-term, medium-term or long-term development trends of time data.

Let Y denote a time series, which can be represented by

$$Y = f(T, S, C, e) \quad (2)$$

where T is the long-term trend, which indicates that the predicted value steadily increases, decreases, or remains at a certain level according to a certain rule over time; S the seasonal change, which means the predicted value has a regular periodic change within a certain time; C the cyclical variation, which represents that the predicted value cyclically changes over a long period; e the random term, which indicates the impact of an unexpected and accidental factor on time series.

In order to discover the regularity of data, the time series data needs to be smoothed and anti-seasonally processed. The processing steps are as follows:

Step 1: Estimate the long-term trend term T to get the product $Se = \frac{Y}{T}$ of seasonal variation term and error term. Use a 6-month central moving average for monthly data to smooth the data:

$$\hat{y}_{m6} = \frac{(0.5y_{t-6} + y_{t-5} + y_{t-4} + \ldots + y_{t+4} + y_{t+5} + 0.5y_{y+6})}{12} \quad (3)$$

Use a 2-month central moving average for quarterly data:

$$\hat{y}_{q2} = \frac{(0.5y_{t-2} + y_{t-1} + y_t + y_{t+1} + 0.5y_{y+2})}{4} \quad (4)$$

The moved data is

$$\frac{y}{\hat{y}} = S \times e \quad (5)$$

There is no seasonality in the moved data, where S is the normalized seasonal factor.

Step 2: Remove the error term and estimate the seasonal term S. The numbers corresponding to different seasons are called seasonal factors are standardized. The data $\frac{y}{\hat{y}}$ after removing the long-term trend includes the seasonal and random error terms. By averaging data from the same season in different years, the error term can be removed,

with seasonal term left. In order to ensure that the average of seasonal index is 1, the seasonal factor needs to be normalized.

Normalize the monthly data:

$$zb_{im} = \frac{z_i \times 12}{\sum_{j=1}^{12} z_j} \tag{6}$$

Normalize the quarterly data:

$$zb_{iq} = \frac{z_i \times 4}{\sum_{j=1}^{4} z_j} \tag{7}$$

Step 3: Remove the seasonal terms from original data to obtain the seasonally adjusted data.

2.3 Fuzzy FCM clustering

The FCM clustering algorithm was first proposed by Bezdek [23]. Compared with other clustering algorithms, it is the most effective and simplest to calculate, widely used in industrial processes.

Given the data sample set $\phi = \{\phi(1), \phi(2), \ldots, \phi(N)\}$, the FCM clustering algorithm obtains the membership matrix $\mathbf{U} = [\mu_{i,k}]_{c \times N}$ and clustering center $\mathbf{V} = [v_1, v_1, \ldots, v_c]$ by minimizing the objective function. When the number c of clusters is constant, the objective function of FCM clustering algorithm can be expressed as

$$J(U, V, \Phi) = \sum_{q=1}^{c} \sum_{k=1}^{N} \mu_{q,k}^{w} ||\phi(kT) - v_q||^2 \tag{8}$$

where w is the index weight that affects the fuzzification of membership matrix. $w \in (1, \infty)$, and it is often set to 2. $\mu_{i,k}$ satisfies

$$\begin{cases} \sum_{q=1}^{c} \mu_{q,k} = 1, & k = 1, 2, \ldots, N \\ 0 < \sum_{k=1}^{N} \mu_{q,k} < N, & q = 1, 2, \ldots, c \end{cases} \tag{9}$$

The objective Eq. (8) is minimized to obtain

$$v_q = \frac{\sum_{k=1}^{N} \mu_{q,k}^{w} \phi(kT)}{\sum_{k=1}^{N} \mu_{q,k}^{w}}, \quad q = 1, 2, \ldots, c \tag{10}$$

$$\mu_{q,k} = \frac{1}{\sum_{j=1}^{c} \left(\frac{||\phi(kT) - v_q||}{||\phi(kT) - v_q||}\right)^{2/(w-1)}}, \quad q = 1, 2, \ldots, c, \tag{11}$$

$$k = 1, 2, \ldots, N$$

(10) and (11) cannot obtain the analytical solutions. The FCM clustering algorithm provides an iterative algorithm to approximately obtain the minimum value of objective function.

Step 1: Giving the data sample set $\mathbf{\Phi}$, the number c of clusters, and the arbitrary initial membership matrix $\mathbf{U}_0$.

Step 2: Calculate the cluster center vector $v_q, q = 1, 2, \ldots, c$ according to Eq. (10).

Step3: Recalculate the subjection degree $\mathbf{U}$ from v_q obtained from Step 2 and Eq. (11). If $q = j$, and $||\phi(kT) - v_q|| = 0$, then $\mu_{j,k} = 1, \mu_{q,k} = 0, \forall q \neq j$.

Step4: Repeat the above steps until the given convergence index is satisfied. For example, $||\mathbf{U}_l - \mathbf{U}_{l-1}|| \leq \varepsilon$ where $|| \bullet ||$ is the norm; l the iteration; ε the index for terminating iteration. When $\varepsilon = 0.01$, the satisfactory accuracy can be achieved.

After iteration, the membership matrix $\mathbf{U}$ and cluster center $\mathbf{V}$ can be obtained. That is, with the given number c of clusters, the parameter v_q to be identified is determined, and s_q can be determined by the nearest neighbor heuristic algorithm:

$$s_q = \left[\frac{1}{p} \sum_{l=1}^{p} (c_q - c_l)^T (c_q - c_l)^{1/2} \right], \quad q = 1, 2, \ldots, c \tag{12}$$

where p is the number of nearest neighbors of the q-th cluster, and $c_l (l = 1, 2, \ldots, p)$ is the cluster center of each nearest neighbor of c_q.

After pre-processing the time series with the length of n according to the method in Sect. 2.2, the attributes of used FCM clustering algorithm are divided. The monthly fluctuations relative to the average trend are taken as a new time series instead of original one, and then the above method is used for fuzzy clustering. Each of these local sequences is softened to a representative form.

Supposing $X = \{X_1, X_2, \ldots, X_n\}$ be a new time series, the time window with the width of ω is applied to X, thus forming the subsequence $Y_i = \{X_i, X_{i+1}, \ldots, X_{i+\omega-1}\}$ with the length of ω. The time window is slipped in a single step from the beginning to the end of time series X to form a series of subsequences $M_1, M_2, \ldots, M_{N-\omega+1}$ with the width of ω. It is denoted as $W(X, \omega) = \{M_i | i = 1, 2, \ldots, N - \omega + 1\}$, which is the set of subsequences that the time series X slips using the sliding window with the width of ω. $W(X, \omega)$ is regarded as $N - \omega + 1$ points in ω-dimensional Euclidean space, and the FCM method is used for fuzzy clustering.

3 Early warning model of corporate financial risk is based on fuzzy association rules

Corporate financial risk analysis is to select the appropriate risk analysis models and risk analysis indicators [23, 24]. The degree of risk is quantitatively described, defining the level of corporate financial risk [25, 26]. Management takes measures to control risks and provides theoretical and practical evidence.

3.1 Selection of financial indicators and correlation analysis

In order to examine the impact of financial indicators on corporate financial risks, the financial indicators in the work were closely related to financial risks, and selected from corporate profitability indicators, corporate operation ability indicators, corporate growth ability indicators, corporate debt-paying ability indicators and corporate cash flow indicators. On this basis, after the correlation analysis of these financial indicators, some highly relevant financial indicators are eliminated to simplify the model. The correlation coefficient of each financial indicator can be determined by

$$r_{x,y} = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}} \qquad (13)$$

where x, y are two variables, and $r_{x,y}$ is the correlation coefficient of variables, satisfying $-1 \le r_{x,y} \le 1$. When $|r_{x,y}| = 1$, x, y are in completely linear correlation; $r_{x,y} = 1$ indicates x, y are in completely positive correlation, and $r_{x,y} = -1$ in completely negative correlation. $r_{x,y} = 0$ indicates x, y are in non-correlation. $-1 < r_{x,y} < 1$ indicates x, y are in linear relationship.

Indicators with high positive or negative correlations are excluded to reduce the collinearity between financial indicators.

3.2 Data preprocessing and reconstruction

In order to reduce the deviation of corporate financial risk analysis, it is necessary to clean up the collected sample data and eliminate all abnormal values of financial indicators. At the same time, in order to mine the association rules in the next step, the continuous financial indicator data needs to be discretized according to the financial risk grade.

William Weitzel et al. divided the process from recession to final death of enterprise into five stages: blind stage, sluggish stage, wrong-action stage, crisis stage and extinction stage. Therefore, all indicators can be divided into 5 stages according to the principle of enterprise life cycle, which are represented by 1, 2, 3, 4 and 5, respectively. According to the distribution of indicator values, the financial crisis early-warning indicators are divided into five sub-areas (See Fig. 1).

In the recession and metamorphosis periods, the enterprise must continuously reform to seek metamorphosis, or it will die out. Therefore, the establishment of appropriate early warning mechanisms is necessary for modern enterprises.

The database reconstruction is to discretize the continuous data of financial crisis warning indicators into financial indicator data suitable for association rule mining. Due to the data set consisting of financial data of different companies, each financial indicator variable is basically in normal distribution. Thus, the equal area partition in normal distribution is used to discretize continuous variables. The work discretized each financial indicator variable into 5 grades according to the 1/5, 2/5, 3/5, and 4/5 quantiles of distribution function of each variable.

3.3 FARM algorithm (fuzzy association rules big data mining algorithm)

The work used the Apriori algorithm based on candidate pattern generation and testing to determine the frequent pattern sets, and parallelized the candidate pattern sets. They were counted after being divided on each processor, and the processors communicated through message passing. The time series was obtained after processing by parallel algorithm. The continuous attributes were discretized to obtain new fuzzy-attribute data sets, which were divided on each processor. During data scanning, the processor could calculate the local fuzzy supporting number asynchronously and independently. The synchronization was maintained at the end of each scan, and the same candidate sets were saved by processor. The inputs were the minimum fuzzy support degree $\sup_{\min}$ and minimum
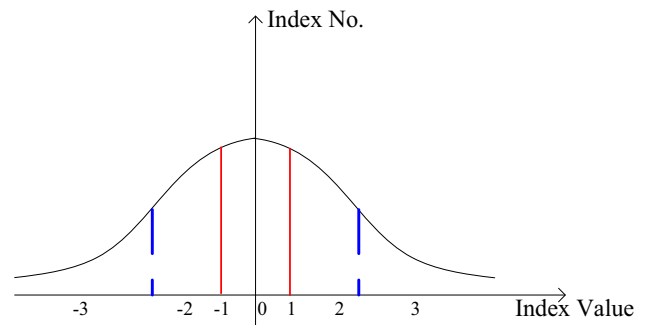

Fig. 1 Equal area partition of warning indicators of financial crisis

fuzzy trust degree $conf_{\min}$; the output was the association rule set S_{ar}. The algorithm steps are as follows:

Step 1: Set up the parallel processor $p_1, p_2, \ldots, p_n$.

Step 2: Divide the transaction database into multiple partitions and allocate them to each processor separately.

Step 3: Cluster each processor using FCM algorithm and transform it into a new data set. Use the continuous-attribute discretization technology to transform the obtained time series into a new database, and a prefix tree is constructed according to $\sup_{\min}, conf_{\min}$.

Step 4: Perform a local count on each processor according to Step 3. For each transaction in the transaction database and each item in the candidate itemset, if an item belongs to a transaction on a processor, a local count is performed. Propagate them to other sites.

Step 5: Calculate the global count, and generate a rule set.

The rules are filtered according to the timing constraints to be satisfied by the antecedent and posterior attributes of rules to obtain the timing rules. The development trend of rules is used to determine the enterprise crisis degree, with a qualitative analysis of enterprise financial crisis. By calculating the crisis coefficient, the enterprise crisis stage is finally determined, which realizes the quantitative analysis of enterprise financial crisis. If with low antecedent and high consequent of rules, the corporate crisis is aggravated; otherwise, the crisis is reduced. If the rules are always at the first stage, the enterprise crisis is relatively light; if at the third stage, the crisis is moderate; if at the fifth stage, the enterprise is on the bankruptcy verge.

Crisis coefficient is introduced to calculate the specific degree of enterprise financial crisis:

$$
\begin{aligned}
F &= F(\overline{x}_i, \sup(x_i), conf(x_i)) \\
&= \frac{1}{n}\sum_{i=1}^{n} \overline{x}_i \sup(x_i) + \frac{1}{n}\sum_{i=1}^{n} \overline{x}_i conf(x_i)
\end{aligned}
\tag{14}
$$

where n is the number of rules, and $\overline{x}_i$ the discretized variable data.

4 Example simulation

A minor financial crisis may be just a temporary difficulty in capital turnover, while a serious financial crisis is an unsuccessful operations or bankruptcy liquidation. It is a development process from financial crisis to corporate bankruptcy. If taking appropriate measures, companies may resolve the financial crisis [27].

The work selected the Chinese ST listed company as the research object, and the annual and quarterly statements of from 2003 to 2018 as the data source, collecting a total of 32 financial indicators (see Table 1).

The data samples of selected financial indicators were sorted out to remove outliers. Equation (13) was used to calculate the correlation coefficients between financial indicators, excluding the financial indicators with higher absolute correlation coefficients in the same group of financial indicators. They were return on net assets X3, net profit X5, main business income per share X10, number X12 of receivables turnover days, number X14 of inventory turnover days, number X16 of current assets turnover days, principal business income growth rate X17, net asset growth rate X19, quick ratio X24, and cash flow ratio X32. The remaining indicators were discretized according to the financial risk level to obtain a reconstructed financial indicator database.

The 12 financial indicators of each enterprise's 12 quarters were summarized as a record for time series analysis. The discretized data set was coded at the granularity of each level of each financial indicator every quarter, which was taken as the input of association rule mining algorithm. The association rules were performed with the sliding window parameters $\omega(\omega = 3)$ and the clustering class number of 4. Then the algorithm proposed was used to mine the rules, thus obtaining the fuzzy association rule set. Besides, association rules were used for predicting financial indicators and crisis warning. The level attribute set of financial index of analysis target can be matched in whole or in part in the mined antecedent of rule base rules to obtain the corresponding crisis warning information.

The financial index data of Chinese ST listed companies from 2013–2018 was selected, with a total of 720 data records for 30 companies. Each record contained 22 financial indicators after excluding high-correlation ones. The proposed algorithm and traditional Apriori algorithm were used to generate frequent pattern sets, and Fig. 2 shows the running time of algorithm.

Coordinate X is the support threshold, with the variation range of 0.25–0.15, and the step size of 0.01. Coordinate Y is the running time for computing frequent pattern sets. With different support thresholds, the proposed algorithm has shorter running time and higher operating efficiency.

Figure 3 shows the number of rules under different support thresholds and confidence thresholds. Coordinate X is the support degree; coordinate Y the confidence degree; coordinate Z the rule number. When different support thresholds and confidence thresholds are selected, different association rules can be obtained according to financial-indicator database.

The data mining of financial indicators shows that when the listed companies have financial risks, some key financial indicators always appear frequently (See Table 2). Main key indicators are the key factors for judging whether the enterprise has financial risk, and their fluctuation determines the risk level of enterprise.

Table 1 Financial indicators

No	Index	No	Index	No	Index
1	Earnings X1 per share	12	Number X12 of receivable turnover days	23	Current ratio X23
2	Net asset X2 per share	13	Inventory turnover rate X13	24	Quick-action ratio X24
3	Yield X3 of net asset	14	Number X14 of inventory turnover days	25	Cash ratio X25
4	Cash flow X4 per share	15	Turnover rate X15 of current assets	26	Interest payment multiple X26
5	Net profit X5	16	Number X16 of current-asset turnover days	27	Asset-liability ratio X27
6	Yield X6 of net asset	17	Income growth rate X17 of main business	28	Ratio X28 of net operating cash flow to sales income
7	Net profit ratio X7	18	Net profit growth rate X18	29	Rate X29 of return on operating cash flow of assets
8	Gross profit rate X8	19	Net-asset growth rate X19	30	Ratio X30 of net operating cash flow to net profit
9	Operating income X9	20	Total asset growth rate X20	31	Ratio X31 of net operating cash flow to debt
10	Main business income X10 per share	21	EPS growth rate X21	32	Cash flow ratio X32
11	Receivable turnover rate X11	22	Shareholders' equity growth rate X22		

Fig. 2 Performance comparison between Apriori and FARM algorithm
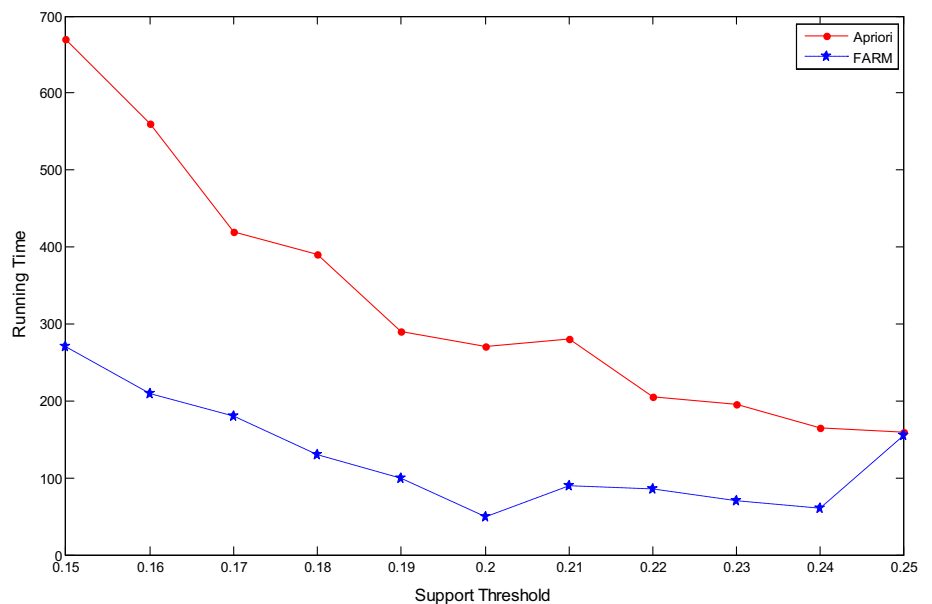
5 Conclusions

With the development of the economy and the era of big data, the enterprises can collect and store all business activity data. The work proposed a fuzzy association rule mining algorithm of time series based on FCM clustering, which used FCM clustering algorithm for the fuzzy discretization on the cleaned time data. A parallel mining algorithm of association rule was used to obtain the frequent fuzzy option sets, and multiple processors in parallel generated the fuzzy association rules satisfying minimum fuzzy trust degree. The rules between all financial indicators were mined to determine more representative financial risk indicators. The big data mining algorithm of Internet of Things was established based on fuzzy association rules to obtain the model of corporate financial-risk analysis. The rules between financial indicators were found to predict corporate financial crisis. The method in the work has been verified by the experiment, and the fluctuation of key indicators determined the enterprise risk degree.

Fig. 3 Relation of association rules, support threshold and confidence threshold
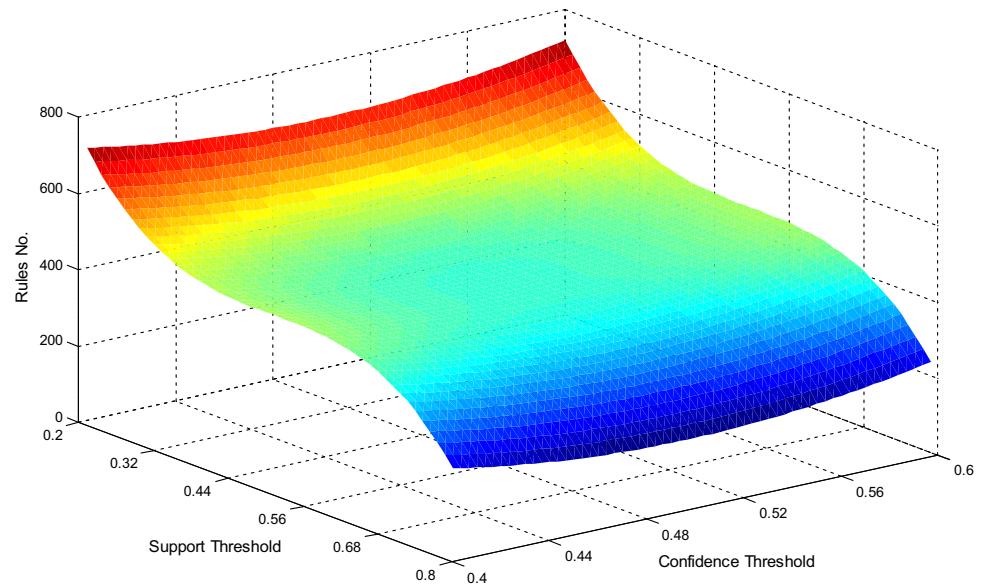

Table 2 Key financial indicators

No	Indicator	No	Indicator	No	Indicator
1	Earnings X1 per share	5	Net-profit growth rate X18	9	Ratio X28 of net operating cash flow to sales income
2	Yield X6 of net asset	6	Total asset growth rate X20	10	Rate X29 of return on operating cash flow of assets
3	Receivable turnover rate X11	7	Current ratio X23		
4	Inventory turnover rate X13	8	Cash ratio X25		

Compliance with ethical standards

Conflict of interest The authors declared that they have no conflicts of interest to this work. We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

References

1. Fitzpatrick PJ (1932) A comparison of ratios of successful industrial enterprises with those of failed firms [J]. Account Publishing Computer 2:589–605
2. Altman EI (1968) Financial ratios, discriminant analysis, and the prediction of corporate bankruptcy [J]. J Finance 23(4):589–609
3. Meyer PA, Pifer HW (1970) Prediction of bank failures [J]. J Finance 25(4):853–868
4. Laitimen EK (1993) Financial predictors for different phases of the failure process [J]. Omega 21(2):215–228
5. Zhu HY, Yong G (2002) The Designing Processing of a Crisis Early Warning System of Core Competence Strategy [J]. Soft Sci 16(5):13–16
6. Ren H, Xu XS (2003) A warning system of enterprise's crisis management [J]. J WUT (Inf Manage Eng) 25(06):153–157
7. Wu YY, Cai QP, Wu F (2008) Pre-warning study of corporations' financial crisis based on ANN technique [J]. J Southeast Univ (Philos Soc Sci) 10(1):22–26
8. Yang GJ, Zhou YM, Sun LL (2009) Enterprise financial early warning method based on Benford-logistic model [J]. J Quantitative Tech Econ 10:149–165
9. Sivapathasekaran C, Mukherjee S, Ray A, Gupta A, Sen R (2010) Artificial neural network modeling and genetic algorithm based medium optimization for the improved production of marine biosurfactant [J]. Biores Technol 101(8):2884–2887
10. Koyuncugil AS, Ozgulba N (2012) Financial early waiving system model and data mining application for risk detection [J]. Expert Sys Appl 39(6):62–65
11. Banerjee A et al (2014) Data analytics: hyped up aspirations or true potential [J]. Vikalpa J Decis Mak 38(4):1–11
12. Cao Y (2012) Aggregating multiple classification results using Choquet integral for financial distress early warning [J]. Expert Sys Appl 39(2):112–123
13. Xiao Z, Yang X, Pang Y, Dang X (2012) The prediction for listed companies' financial distress by using multiple prediction methods with rough set and Dempster-Shafer evidence theory. Knowl-Based Syst 26:196–206. https://doi.org/10.1016/j.knosys.2011.08.001
14. Fang SK, Shyng JY, Lee WS, Tzeng GH (2012) Exploring the preference of customers between financial companies and agents based on TCA. Knowl-Based Syst 27:137–151. https://doi.org/10.1016/j.knosys.2011.09.003
15. Maimon O, Rokach L (2009) Introduction to Knowledge Discovery and Data Mining. In: Maimon O, Rokach L (eds) Data

Mining and Knowledge Discovery Handbook. Springer, Boston, MA

16. Song B, Zhu JM, Li X (2015) The research of enterprise financial early warning based on big data [J]. J Cent Univ Finance Econ 06:55–64

17. Zhang L, Zhang L, Teng W, Chen Y (2013) Based on information fusion technique with data mining in the application of finance early-warning. Procedia Computer Sci 17:695–703. https://doi.org/10.1016/j.procs.2013.05.090

18. Creighton C, Hanash S (2003) Mining gene expression databases for association rules. Bioinformatics 19(1):79–86. https://doi.org/10.1093/bioinformatics/19.1.79

19. Li Li, Donghai Zhai and Fan Jin (2003) GRG: an efficient method for association rules mining on frequent closed itemsets, Proceedings of the 2003 IEEE International Symposium on Intelligent Control, Houston, TX, USA, 2003, pp. 854–859 https://doi.org/10.1109/ISIC.2003.1254748

20. Ceglar A, Roddick JF (2006) Association mining. ACM Computing Surveys (CSUR) 38(2):5. https://doi.org/10.1145/1132956.1132958

21. Bodon F. (2003) A fast APRIORI implementation. In: Proceedings of the ICDM workshop on frequent itemset mining implementations (FIMI '03), Melbourne, Florida, USA.

22. Agrawal R, Srikant R. (1994) Fast algorithms for mining association rules. In: Proc. 20th int. conf. very large data bases, VLDB. pp. 487–499.

23. Babuška R, Verbruggen H (2003) Neuro-fuzzy methods for nonlinear system identification. Annual Rev Control 27(1):73–85

24. Zhou Q, Lou J, Jiang Y (2019) Optimization of energy consumption of green data center in e-commerce. Sustain Computing Inf Sys 23:103–110. https://doi.org/10.1016/j.suscom.2019.07.008

25. Zhou Q, Zhang Z, Wang Y (2020) Research on safety management system optimization of B2C e-commerce intelligent logistics information system based on data cube[J]. J Intell Fuzzy Sys 38(2):1585–1592. https://doi.org/10.3233/JIFS-179522

26. Zhou Q, Zheng Xu, Yen NY (2019) User sentiment analysis based on social network information and its application in consumer reconstruction intention. Comput Hum Behav 100:177–183. https://doi.org/10.1016/j.chb.2018.07.006

27. Lin F, Liang D, Chen E (2011) Financial ratio selection for business crisis prediction. Expert Syst Appl 38(12):15094–15102