Financial distress prediction for listed enterprises using Fuzzy C-Means

Lu Wang

School of Information Management and Engineering, Zhejiang University of Finance and Economics

Abstract

Enterprise financial distress prediction model is an important tool for enterprise risk management. Fuzzy C-Means (FCM) model is more suitable for the classification calculation in reality because of introducing the fuzzy membership. This article adopts the FCM model to predict financial distress. In the experiment, 90 distressed enterprises and 90 normal enterprises were selected as samples, which are all from the same time and the same industry in Chinese Shenzhen and Shanghai Stock Exchange, and 12 financial indicators of these enterprises as a sample dataset. After the data are pre-processed, we identified 11 financial indicators with a significant difference to make the financial distress prediction. Last, it can be concluded based on the FCM results that the performance is better in the case of distressed enterprises.

Keywords: financial distress prediction, financial indicator, fuzzy C-Means

Introduction

With economic globalization, enterprises are more vulnerable to economic crises. Once business bankrupt happens, it not only makes the enterprises' investors face a huge economic risk but also makes the government suffer great economic loss. In order to make enterprises operate healthily and prevent them from going bankrupt in periodical crises, enterprises must pay attention to risk management, strengthen crises monitoring and improve the accuracy of distress prediction. Each enterprise is able to survive an economic crisis only by using proper management of the financial indicators and ensuring normal operation. Therefore, financial distress prediction model can master the financial situation, predict financial distress, and avoid the situation of enterprise failure or restructuring. Many experts and scholars adopt many kinds of models and algorithms to study financial distress prediction.

Fuzzy C-Means (FCM) is an improved algorithm based on the thought of K-means. This algorithm is more suitable for the classification calculation in reality than the K-means. In this article, we adopt the FCM algorithm to predict financial distress and select the financial indicators of Chinese listed companies as the sample data. The data which are pre-processed are calculated by the FCM algorithm. The results can testify the good performance of the FCM algorithm.

The article is organized as follows: Section 2 shows the literature review about the financial distress prediction. Section 3 illustrates the foundation theories related to this paper. Section 4 is the empirical research that describes the process of the financial distress prediction by the FCM algorithm. Section 5 is the conclusion of this article.

Literature review

In the early time, statistical theory and classical econometrics were introduced in order to predict a business failure. Fitzpartrick first made research on financial distress prediction by analysing financial ratios (Fitzpartrick, 1932). Then, Beaver (1966) used a univariate model on financial ratios to predict enterprises bankruptcy. Finally, multiple discriminant analysis (MDA) (Altman, 1968) and regression analysis (Ohlson, 1980; Zmijewski 1984) were used to improve financial distress prediction. But these methods demand data to conform to strict hypothesis condition, which is hard to be realized in the case of complex financial data. With the development of artificial intelligence technology, artificial neural networks are commonly used to predict financial distress (Lahmiri and Bekiros, 2019). Besides, there are other prediction methods, such as support vector machine (Kim, Mun and Bae, 2018), fuzzy sets theory (Sanz et al., 2015), decision tree (Xia, Liu and Li, 2017), case-based reasoning (Leonardi, Portinale and Artusio, 2017), and so on. As undemanding hypothesis condition, these methods are more widely applied for the purpose of financial distress prediction than previous prediction methods.

FCM model is a classic unsupervised clustering algorithm, and has been widely applied in the fields of image processing (Zhao, Ha and Zhao, 2019), agriculture (Anter, Hassenian and Oliva, 2019), pattern recognition (Golsefid, Zarandi and Turksen, 2016), medicine0 (Singh and Bala, 2019), and so on. In terms of financial distress prediction, Liu and Wu (2019) used kernel-based fuzzy c-means to organize samples of companies and designed a hierarchical selective ensemble model for business failure prediction. De Andres et al. (2011) proposed a hybrid system, which combined fuzzy clustering and MARS. Both models were especially suitable for the bankruptcy prediction problem, due to their theoretical advantages when the information used for forecasting was drawn from company financial statements (De Andres, Lorca and Juez, 2011). Liu and Wu (2018) investigated the dynamic process of the company's financial status over years and proposed to design a financial path using fuzzy c-means (FCM) approach.

Foundation Theory

The fuzzy C-means algorithm was first formally introduced by Dunn (1973), which embedded the fuzzy c-partitions into K-means and adopted the degree of membership to determine which cluster an element belongs to. As the typical clustering techniques, the FCM is based on minimizing the following objective function (De Andreas, Lorca and Juez, 2011):

$$J_m(U,V)=\sum_{k=1}^{c}\sum_{i=1}^{n}u_{ik}{}^{m}\left\|x_i-v_k\right\|^2 \tag{1}$$

where x_i is the ith of input data in an D-dimensional space, v_k is the kth of D-dimension centre of the cluster, and c is the number of total clusters. $\|*\|$ is the Euclidean distance between the input x_i and the centre v_k. u_{ik} is the degree of membership of x_i in the cluster c, and need satisfy Eq. (2). m is any real number greater than 1, which determines the amount of fuzziness of the resulting classification.

$$U\in\left\{u_{ik}\in[0,1]\left|\sum_{k=1}^{c}u_{ik}=1,\forall k\ \ and\ \ 0<\sum_{i=1}^{n}u_{ik}<n,\forall i\right.\right\} \tag{2}$$

Then, the FCM algorithm would be iteratively optimized to the degree of membership u_{ik} and the centre v_k for the minimization of Eq. (1). The concrete update equations are shown in the following equations:

$$u_{ik}=\sum_{p=1}^{c}\left[\frac{\left\|x_i-v_k\right\|}{\left\|x_i-v_p\right\|}\right]^{\frac{-2}{m-1}} \tag{3}$$

$$v_k=\frac{\sum_{i=1}^{n}u_{ik}{}^{m}\cdot x_i}{\sum_{i=1}^{n}u_{ik}{}^{m}} \tag{4}$$

This iteration will not stop until the difference of the degree of membership between iterations is smaller than one very small constant.

Empirical research

The selection of samples and financial indices

China Securities Supervision and Management Committee specially treated (ST) the quoted companies that had negative net profit in two consecutive years. This article selects such companies as distressed samples and chooses non-ST companies as normal samples. All samples should be selected from the same time and the same industry so that a reasonable comparison was possible. Therefore, we selected 180 samples in the manufacture industry from Shenzhen and Shanghai Stock Exchange between 2012 and 2016, which included 90 distressed samples and 90 normal samples. In addition, the financial reports of the companies in question are published at the beginning of next year, then China Securities Supervision and Management Committee decides whether they should be specially treated in the same year. So, in this paper, financial data were selected in the year of T-2 when a company experienced a business failure in the year of T.

According to the description above, 12 financial indicators of these enterprises are used as candidate variables. They are shown in Table 1.

Tab. 1: Candidate financial indicators and definition

Variables	Definition	Variables	Definition
X1	Current ratio	X7	Operating profit ratio
X2	Quick ratio	X8	Growth rate of main income
X3	Asset-liability ratio	X9	Growth rate of net profit
X4	Inventory turnover ratio	X10	Cash to sale ratio
X5	Total assets turnover ratio	X11	Net cash flow per share
X6	Rate of return on total assets	X12	Undistributed profits per share

Source: Own processing.

Experimental design

After financial indicators have been confirmed, first, we need to pre-process these data because of the strong or weak correlations in which much repeated information exist and make many unnecessary matters. Then, the accuracy of distressed samples, the accuracy of normal samples and the accuracy of the entirety samples are regarded as the evaluation index for FCM model. Last, to enhance reliability of experiment results, the calculation process is repeatedly carried out 50 times. The final statistical results are average values based on 50 times repeated operation.

Experimental result

The basis predicting of financial distress is the obvious difference of the data between distressed and normal companies. The significance test is used to determine whether the differences of data in different groups are accidental or existential. This method consists of two steps: the first step is a judgment on the distribution of the whole dataset. As the second step, the analysis should be further made according to different situations of the distribution.

In the first step, the Kolmogorov-Smirnov Test (K-S Test) was used to make the normality test for the 12 financial indicators. The results are shown in Table 2.

Tab. 2: K-S test of ratio variables

Indicator	K-Value	P-Value	Indicator	K-Value	P-Value
X1	4.889	0.000	X7	3.474	0.000
X2	4.124	0.000	X8	1.819	0.003
X3	1.100	0.178	X9	1.982	0.001
X4	2.402	0.000	X10	1.695	0.006
X5	1.251	0.087	X11	0.688	0.731
X6	4.832	0.000	X12	0.727	0.666

Source: Own processing.

We can see from the table above that the P-values of X3, X5, X11 and X12 are bigger than 5%, which means that these four indicators are the normal distribution, and the others cannot be regarded as the normal distributed data.

In the second step, first, the T test is carried out on the financial indicators X3, X6, X11 and X12, which are the normal distribution. The results are shown in Table 3.

Tab. 3: T test of ratio variables

Indicator	T-Value	df	Sig.
X3	3.617	178	0.000
X5	-11.293	178	0.000
X11	-4.118	178	0.000
X12	2.998	168.444	0.003

Source: Own processing.

It results from the table above that the Sig. values are smaller than 5%, which means that these four indicators pass the T test, and there is a significant difference in these indicators. Therefore, these four financial indicators should be retained.

Then, the other 8 financial indicators, which are not the normal distribution, are tested by the Mann-Whitney U test. The results are shown in Table 4.

Tab. 4: Mann-Whitney U test of ratio variables

Indicator	Mann-Whitney U	Sig.	Indicator	Mann-Whitney U	Sig.
X1	5430.000	0.000	X7	5126.000	0.002
X2	5079.000	0.003	X8	4703.500	0.061
X4	6298.000	0.000	X9	7822.000	0.000
X6	7080.000	0.000	X10	6358.000	0.000

Source: Own processing.

The table above clearly shows that X8, which is the growth rate of main income, has the Sig. value 0.061>5%, which means that this indicator cannot pass the Mann-Whitney U test, and there is no significant difference for these indicators. Therefore, X8 (growth rate of main income) should be deleted. Finally, we included 11 financial indicators into FCM model to make financial distress prediction.

According to the description of the previous section, we obtain the prediction outcomes of the FCM model, which are shown in the Table 5. The prediction accuracy of the FCM model is 84.78% for the sake of completeness. In terms of the normal companies prediction, the FCM shows only 78.26% accuracy. On the contrary, in the case of the distressed companies prediction, the FCM shows better accuracy (91.30%).

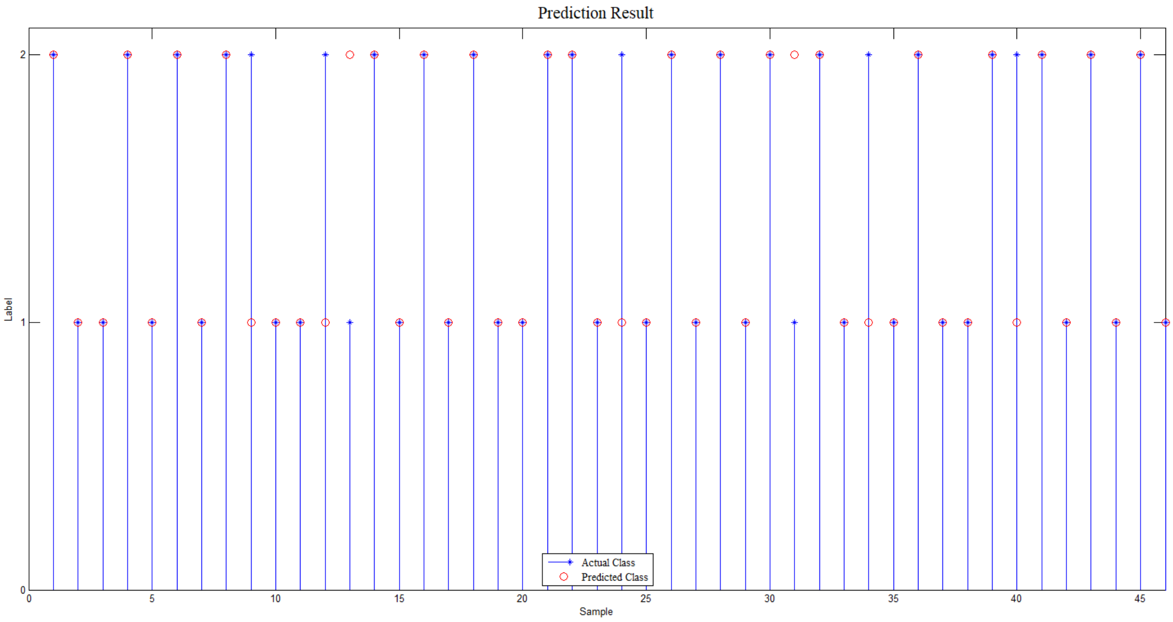
Tab. 5: Prediction accuracy of FCM model

Model	Normal (%)	Distressed (%)	Total (%)
FCM	78.26	91.30	84.78

Source: Own processing.

The specific prediction results of samples are shown in Figure 1 as follow.

Figure 1: Prediction results



Source: Own processing.

In Figure 1 the horizontal axis expresses the sample, and the vertical axis expresses the sample label. The blue star with the blue line expresses the actual classes of samples, and the red circle expresses the predicted classes of samples. Therefore, when the red circle coincides with blue star, the predicted results are correct. In other cases, the predicted results are wrong.

The above prediction results indicate that the FCM model´s performance is better in case of making predictions for predicting the distressed enterprise. In fact, if one distressed company is misclassified as a normal company, it is easy to make more serious conclusions because it does not consider risk prevention. Therefore, the FCM model appears to be more suitable to make the financial distress predictions.

Conclusion

In this article, we adopted the FCM model to predict financial distress. First, the calculated process of the FCM model was introduced. Then, 90 distressed companies and 90 normal companies, which were all from Chinese Shenzhen and Shanghai Stock Exchange, were selected as samples, and 12 financial indicators of these enterprises were chosen as a sample dataset. In the experiment, we pre-processed the sample data and deleted the indicator without a significant difference between the two classes, and introduced the 11 financial indicators into the FCM model. Last, we obtained the final results using the FCM model.

It can be concluded that the FCM model is more suitable for predicting the distressed companies, which is more important for the financial distress prediction.

Acknowledgement

The research in this article has been supported by the National Natural Science Foundation of China under Grant No. 71801187.

References

ALTMAN, E. I., 1968. Financial ratios discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance.* **23**(4), 589-609. ISSN 1540-6261.

ANTER A. M., HASSENIAN, A. E. and OLIVA, D., 2019. An improved fast fuzzy c-means using crow search optimization algorithm for crop identification in agricultural. *Expert Systems with Applications*. **118**, 340-354. ISSN 1873-6793.

BEAVER, W. H., 1966. Financial Ratios as Predictors of Failure, Empirical Research in Accounting: Selected Studies*. Journal of Accounting Research.* **4**, 79-111. ISSN 1475-679X.

DE ANDRES, J., LORCA, P., JUEZ, F. J. D. and SANCHEZ-LASHERAS, F., 2011. Bankruptcy forecasting: A hybrid approach using Fuzzy c-means clustering and Multivariate Adaptive Regression Splines (MARS). *Expert Systems with Applications*. **38**(3), 1866-1875. ISSN 0957-4174.

DUNN, J. C. S., 1973. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*. **3**(3), 32-57. ISSN 1087-6553.

FITZPARTRICK, P. J., 1932. A comparison of ratios of successful industrial enterprises with those of failed firm. *Certified Public Accountant*. **10**, 598-605. ISSN 0732-8435.

GOLSEFID, S., ZARANDI, M. H. and TURKSEN, I. B., 2016. Multi-central general type-2 fuzzy clustering approach for pattern recognitions. *Information Sciences.* **328**, 172-188. ISSN 0020-0255.

KIM S., MUN, B. M. and BAE, S. J., 2018. Data depth based support vector machines for predicting corporate bankruptcy. *Applied Intelligence*. **48**(3), 791-804. ISSN 1573-7497.

LAHMIRI S. and BEKIROS, S., 2019. Can machine learning approaches predict corporate bankruptcy? Evidence from a qualitative experimental design. *Quantitative Finance*. **4**, 1-12. ISSN 1469-7696.

LEONARDI, G., PORTINALE, L. and ARTUSIO, P., 2017. Recommending Personalized Asset Investments through Case-Based Reasoning: The SMARTFASI System. In: Bourbakis, N., Esposito, A., Mali, A., Alamaniotis, M., *IEEE 28th International Conference on Tools with Artificial Intelligence*. San Jose, CA: IEEE, s. 804-811. ISBN 978-1-5090-4459-7.

LIU, J. M. and WU, C., 2019. Hybridizing kernel-based fuzzy c-means with hierarchical selective neural network ensemble model for business failure prediction. *Journal of* *forecasting.* **38**(2), 92-105. ISSN 1099-131X.

LIU, J. M. and WU, C., 2018. Company financial path analysis using fuzzy c-means and its application in financial failure prediction. *Journal of business economics and management*. **19**(1). 213-234. ISSN 2029-4433.

OHLSON, J. A., 1980. Financial ratios and probabilistic prediction of bankruptcy. *Journal of Accounting Research*. **18**, 109-131. ISSN 1475-679X.

SANZ, J. A, BERNARDO, D., HERRERA, F., BUSTINCE, B. and HAGRAS, H., 2015. A Compact Evolutionary Interval-Valued Fuzzy Rule-Based Classification System for the Modeling and Prediction of Real-World Financial Applications with Imbalanced Data. *IEEE Transactions on Fuzzy Systems*. **23**(4), 973-990. ISSN 1941-0034.

SINGH C. and BALA, A., 2019. A transform-based fast fuzzy C-means approach for high brain MRI segmentation accuracy. *Applied Soft Computing*. **76**, 156-173. ISSN 1872-9681.

XIA, Y., LIU, C. and LI, Y., 2017. A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring. *Expert Systems with Applications*. **78**, 225-241. ISSN 1873-6793.

ZHAO Y., HA, Q. L. and ZHAO, Y. D., 2019. Soil pore identification with the adaptive fuzzy C-means method based on computed tomography images. *Journal of Forestry Research.* **30**(3), 1043-1052. ISSN 1993-0607.

ZMIJEWSKI, M., 1984. Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research*. **22**, 59-82. ISSN 1475-679X.

Contact address of the author:

Lu Wang, Ph.D., School of information management and engineering, Zhejiang university of finance and economics, Hangzhou, Zhejiang, 310000, China, e-mail: wanglu_hit@126.com