Scientific Research

An Improved Fuzzy ISODATA Algorithm for Credit Risk Assessment of the EIT Enterprises*

Jike Yu, Zongfang Zhou#, Hua Zhong, Huizhong Huang

School of Management and Economics, University of Electronic Science and Technology of China, Chengdu, China
Email: #zhouzf@uestc.edu.cn

ABSTRACT

We proposed an improved fuzzy ISODATA algorithm for credit risk assessment of the emerging information technology enterprise in this paper. Firstly, as the uncertainty of the EIT enterprise is relatively large, we set a reference sample and an initial clustering center matrix so that we overcame the shortcomings of traditional ISODATA algorithm and improved the reliability of fuzzy clustering analysis. Secondly, we proposed the steps of evaluating the EIT enterprises' credit risk with improved fuzzy ISODATA algorithm. Last but not least, we assessed 10 EIT enterprises' credit risk of a certain city, which proved the effectiveness and operability.

Keywords: EIT Enterprise; Credit Risk; Reference Sample; ISODATA Algorithm; Fuzzy Clustering

1. Introduction

Emerging Information Technology (EIT) is defined as a technology that can innovate or upgrade the function, product, or service of information technology by using the basic principles and methods of information science, as well as with the technical characteristics of emerging technologies [1]. As there are uncertain elements like the EIT itself, product market of the EIT enterprise and so on, the EIT enterprises are facing large credit risk [2,3].

Credit is the inevitable product of social economic development, and is also an essential part of modern social economy. Credit risk is the possibility that a bond issuer or borrower will default by failing to repay principal and interest in a timely manner, leading to a loss to a bank, or investors. EIT enterprises are the typical venture business. To make objective and comprehensive assessment of their credit risk is not only necessary foundation for a smooth financing, but also the essential part for the EIT enterprises' risk management.

The current credit risk assessment models are Credit Metrics, Credit Risk+, KMV, multi-objective decision making, non-parametric statistical methods, neural network and so on. Using fuzzy clustering analysis for corporate credit risk assessment is non-parametric statistical methods [4, 5]. It's especially in kinds of methods are unsure of the overall distribution function, with good results. The current literatures usually no longer evaluate the merits after

classify the target objectives with fuzzy clustering method, or they consider all indicators as efficiency indicators in the ratings analysis, and the classification is according to the properties of specific target range set. According to this current way, it is possible that the credit rating results are different, if we cluster and rate the same EIT enterprise separately with two (or more than two) groups of the EIT enterprises with different qualifications as one target set. Especially in the case that there is very small number of objects in one group, the possibility that this situation happens is huge, which reduce the objectivity of the EIT enterprise credit risk assessment. On the other hand, as an EIT enterprise is usually in initial period of foundation, there is large uncertainty of its growing and development, and the index data used to assess its credit risk is usually incomplete which need to set up a corresponding reference sample system firstly. Reference sample system is an objective standard for the specific requirement of the EIT enterprise credit risk assessment, which is set as ideal value of every characteristic of every grade, used to study the target clustering. Currently, there are few literatures about the EIT enterprise credit risk assessment.

In view of this, we've improved the traditional fuzzy clustering method through setting the reference sample system of the EIT enterprises, and have proposed an improved fuzzy clustering algorithm to cluster the credit risk of the EIT enterprises. Example shows that the improved algorithm solves the problem of insufficient objectivity of the traditional fuzzy clustering method applied in the credit risk assessment in a certain extent.

2. Fuzzy Cluster Analysis

Fuzzy ISODATA (Iterative Self-Organizing Data Analysis Techniques Algorithm) is an interactive self-organizing data analysis technique for fuzzy cluster [6,7]. Cluster using standard Fuzzy ISODATA works as follow, suppose classes' number has been decided, and choose an original fuzzy cluster matrix, calculate optimal fuzzy cluster matrix and optimal cluster center matrix using iterative operation, then classify the inspected object. The algorithm requires more stringent selection of original fuzzy cluster matrix. Inappropriate selection would cause distortion in iterative process. There are limitations when standard fuzzy ISODATA was used in the scene of rating of target object. The algorithm can only cluster object into specific classes, but can't discriminate whether classes meet the "meaningful distance". Based on this, reference sample system and investigation sample will be collected to be cluster. Improved fuzzy ISODATA algorithm steps are as follows:

1) Establish the original characteristic indicators matrix U^* that descript each attribute value of all inspected object and reference samples. u_{ij}^* is on behalf of the characteristic indicators j of object i.

2) Standardize the data of original characteristic indicators matrix U^* by range method to get U, define $M_j = \max\left(u_{1j}^*, u_{2j}^* \cdots, u_{nj}^*\right); m_j = \min\left(u_{1j}^*, u_{2j}^* \cdots, u_{nj}^*\right)$ for column j of U^*, calculate u_{ij} using formula (1)

$$u_{ij} = \frac{u_{ij}^* - m_j}{M_j - m_j} \tag{1}$$

3) Start iterative operation based on original cluster center matrix $V^{(0)}$ of reference sample system, $l = 0,1,2,\cdots$.

4) Calculate fuzzy classified matrix $R^{(l)}$ using formula (2), where c is on behalf of classes number. And based on

Euclid distance, $\|u_k - V_i\| = \left(\sum_{j=1}^m \left(u_{kj} - v_{ij}\right)^2\right)^{1/2}$,

$$r_{ik}^{(l)} = \left[\sum_{j=1}^c \left(\frac{\|u_k - V_i^{(L)}\|}{\|u_k - V_j^{(l)}\|}\right)\right]^{-2} \tag{2}$$

5) Modify cluster center matrix for $R^{(l)}$, $V^{(l+1)} = \left(V_1^{(l+1)}, V_2^{(l+1)}, \cdots V_c^{(l+1)}\right)^T$, where

$$V_i^{(l+1)} = \frac{\sum_{k=1}^n (r_{ij}^{(l)})^2 u_k}{\sum_{k=1}^n (r_{ij}^{(l)})^2} \tag{3}$$

6) Repeat step 2), compare $R^{(l)}$ and $R^{(l+1)}$, for given precision $\varepsilon > 0$, if $\max\left\{r_{ik}^{(l)} - r_{ik}^{(l+1)}\right\} \le \varepsilon$, iterative operation should be stopped and $R^{(l+1)}$ $V^{(l+1)}$ should be outputted. In opposite condition, $l = l+1$, repeat step 3).

7) Get fuzzy cluster based on optimal cluster center matrix discrimination principle--suppose the optimal cluster center matrix $V^* = \left(V_1^*, V_2^*, \cdots, V_c^*\right)^T$, $\forall u_k \in U$, if $\|u_k - V_i^*\| = \min_{1 \le j \le c}\left(\|u_k - V_j^*\|\right)$, object u_k should be classified to class i.

3. Assessment Steps of Integration of Rough Set and Improved Fuzzy Clustering

The evaluation indicators should be screened at first when we assess the EIT enterprises' credit risk. Based on indicators screening, the application of improved fuzzy ISODATA algorithm for classification of the EIT enterprises credit risk would have a better result. The following are specific steps of assessing and classifying the EIT enterprises credit risk with attribute reduction method from integration of rough set theory and improved fuzzy ISODATA algorithm:

1) Establish an initial set of assessment indicators and sample set to be inspected. Suppose $P = \left(p_1, p_2, \cdots, p_m\right)$ is the initial set of assessment indicators, and $X = \left(x_1, x_2, \cdots, x_n\right)$ is the sample set. a_{ij} is the value of assessment indicator j of sample I;

2) Discretization of data a_{ij};

3) Attribute reduction to the indicator set with application of rough set theory [8,9];

4) Establish a proper reference sample set, and construct an initial clustering center matrix $V^{(0)}$;

5) Add the reference sample set into sample set to be inspected, and cluster all the samples with above-mentioned improved fuzzy ISODATA algorithm in order to achieve risk rating of the EIT enterprises.

4. Case Analysis

We assessed credit risk of 10 EIT enterprises (denoted by A, B, C, D, E, F, G, H, I and J) in certain city. Firstly, choose 4 primary-level indicators and 15 secondary indicators, according to systematic, scientific, operational, objective principle as well as a combination of quantitative and qualitative principle, and referring the evaluation indexes system of emerging technology enterprise credit risk [1]. They are financial benefits status indicator (P_1): ROE (P_{11}), ROA (P_{12}), asset maintenance and appreciation indicator (P_{13}), OPE (P_{14}); assets operating state indicator (P_2): total asset turnover (P_{21}), current assets turnover (P_{22}), inventory turnover (P_{23}), receivable accounts turnover (P_{24}); debtpaying ability indicator (P_3): asset-liability ratio (P_{31}), acidtest ratio (P_{32}), cash flow debt ratio (P_{33}); development state indicator (P_4): sales growth rate (P_{41}), capital accumulation rate (P_{42}), the average growth rate of capital (P_{43}) and the ability of technological innovation and application (P_{44}). The data is from the financial statements of 10 EIT enterprises, except (P_{44}) is through experts grading.

ME

Secondly, make the index data of the EIT enterprise discrete. Use the software, Rosetta, and choose the Entropy Scalar tool in it, which can measure entropy, to accompany this step.

The next step is attribute reduction. With help of *Genetic Algorithm* tool in *Rosetta,* we get the indicators of credit risk characteristics of these 10 EIT enterprises as **Table 1**, denoted by $P' = \{P_{12}, P_{14}, P_{21}, P_{24}, P_{33}, P_{41}, P_{44}\}$, while 8 indicators such as ROA are redundant attributes.

Last but not least, establish the 5-level reference sample system, denoted as K、L、M、N and O, referring to the current 5-level credit risk standard of commercial bank [5] and credit rating standard of the IT corporate in the EIT corporate [10], as **Table 2** shows (*V* indicates the highest credit rating and correspond to the lowest credit risk).

Then the initial clustering centers matrix is got after normalization of the data in **Table 2**:

$$V^0 = \begin{pmatrix} 0 & 0.022 & 0.073 & 0.207 & 0 & 0.531 & 0 \\ 0.004 & 0.138 & 0.219 & 0.322 & 0.039 & 0.595 & 0.500 \\ 0.014 & 0.294 & 0.366 & 0.437 & 0.101 & 0.675 & 0.500 \\ 0.050 & 0.392 & 0.439 & 0.632 & 0.200 & 0.730 & 1.000 \\ 0.117 & 0.487 & 0.512 & 1 & 0.301 & 0.806 & 1.000 \end{pmatrix}$$

Set the accuracy $\varepsilon = 0.001$, and according to the above mentioned improved fuzzy ISODATA algorithm specific steps, through computer multiple iterative operation, get such as shown in **Table 3** the clustering center of the credit rating.

Now the classification results could be got according to **Table 3** and the principle of optimal clustering center

Table 1. Indicators of credit risk characteristics of eit enterprises.

	financial benefits status		assets operating state		debt-paying ability		development state
	ROA P_{12}	OPE P_{14}	total asset turnover P_{21}	receivable accounts turnover P_{24}	cash flow debt ratio P_{33}	sales growth rate P_{41}	ability of technological innovation and application P_{44}
A	0.085	0.234	0.562	2.455	0.085	0.077	1
B	0.134	0.199	0.243	0.142	0.563	0.001	1
C	0.572	0.345	1.365	2.458	0.428	0.052	0
D	0.102	0.313	0.341	2.336	0.225	0.002	1
E	0.878	0.255	0.785	0.183	0.372	0.001	2
F	0.172	0.444	0.759	2.225	0.278	0.089	2
G	0.074	0.755	0.378	3.376	2.064	0.002	1
H	0.254	0.001	0.566	0.904	0.010	0.001	0
I	0.052	0.58	0.567	2.761	0.899	0.001	0
J	0.029	−0.226	0.001	0005	0.039	−0.87	1

Table 2. Credit risk reference sample system for eit enterprises.

	P_{12}	P_{14}	P_{21}	P_{24}	P_{33}	P_{41}	P_{44}	Credit grade
K	−0.084	−0.204	0.100	1.800	−0.082	−0.116	0	I
L	−0.005	−0.091	0.300	2.800	0.001	−0.025	1	II
M	0.004	0.062	0.500	3.800	0.135	0.088	1	III
N	0.036	0.159	0.600	5.500	0.347	0.167	2	IV
O	0.095	0.252	0.700	8.700	0.564	0.275	2	V

Table 3. The clustering centers of every credit grade.

Credit Grade	P_{12}	P_{14}	P_{21}	P_{24}	P_{33}	P_{41}	P_{44}
I	−0.075	−0.202	0.092	1.685	−0.004	−0.134	1.055
II	−0.002	−0.080	0.308	2.753	0.011	−0.002	1.317
III	0.005	0.080	0.522	3.529	0.209	0.067	2.004
IV	0.040	0.187	0.707	5.133	0.396	0.113	2.358
V	0.105	0.358	0.803	7.121	0.661	0.298	2.887

Table 4. The classification results.

Credit Grade	EIT Enterprises
I	B G J K
II	A D L
III	C E I M H
IV	F N
V	O

matrix discrimination, shown in **Table 4**.

In this case, according to the traditional fuzzy ISODATA algorithm that leaves out reference sample system, 10 investigated EIT enterprises must be assigned into 5 credit grades. In the case of V, at least one enterprise would be assigned into this highest level. Actually, only pre-set reference sample existed in grade V, and 7 property values of all cluster centers are close to corresponding attribute values of reference samples, more than that, 5 reference samples are assigned to correct credit grade after clustered with credit risk indicators of other 10 investigated EIT enterprises. Through comparative analysis, results obtained by the improved algorithm stated in this paper are more appropriate for EIT enterprises realities.

5. Conclusion

Setting a reference sample system in this paper overcomes the situation that when a same EIT enterprise forms clustering set separately with other EIT enterprises with different qualifications, it will show different assessment results, and also increase the objectivity of the EIT enterprises' credit assessment. At the same time, it is proved that the possible distortion brought by a random selection of initial fuzzy classification matrix could be avoided, if we accord to the improved method that pre-set the reference sample system and determine the initial clustering center matrix, then calculating the initial fuzzy classification matrix, which increase the reliability of fuzzy clustering analysis for the EIT enterprise credit risk.

REFERENCES

[1] Z. F. Zhou, *et al.*, "The Evolution and Evaluation of Credit Risk for Emerging Technology Companies," Science Press, Beijing, 2010. (in Chinese)

[2] J. Yu, Z. Zhou, *et al.*, "The Ordering Structure of Evaluation Indexes Sequences for the EIT Project," *Journal of Emerging Trends in Computing and Information Sciences*, Unpublished.

[3] J. Yu, Z. Zhou, *et al.*, "Study on Evaluation for EIT Project Based on Membership Cloud Gravity Center," ICISE 2011.

[4] M. Lundy, "Cluster Analysis in Credit Scoring, Credit Scoring and Credit Control," Oxford University Press, New York, 1993.

[5] H. Lin and M. Xia, "Enterprise Credit Rating of Theory and Practice," Foreign Economic and Trade University Press, Beijing, 2003. (in Chinese)

[6] J. C. Bezdek, "Pattern Recognition with Fuzzy Objective Function Algorithms," Plenum Press, New York, 1981.

[7] S. Chen, J. Chen and X. Wang, "The Fuzzy Set Theory and Its Application," Science Press, Beijing, 2005. (in Chinese)

[8] Z. Pawlak, "Rough Set: Theoretical Aspects of Reasoning about Data," Kluwer Academic Publishers, Dordrecht, 1991.

[9] X. Xu, *et al.*, "Based on the Fuzzy Toolbox and Rosetta's Rough Set Data Mining," *Microcomputer Information*, Vol. 23, No. 18, 2007, pp. 174-178.

[10] Y. Zhang, Z. Zhou and H. Tang, "An Selection Method of the ETF's Credit Risk Evaluation Indicators," ICCS 2008, Kraków, 2008.