World Academy of Science, Engineering and Technology
International Journal of Economics and Management Engineering
Vol:5, No:7, 2011

Fuzzy Clustering Analysis in Real Estate Companies in China

Jianfeng Li, Feng Jin and Xiaoyu Yang

Open Science Index, Economics and Management Engineering Vol:5, No:7, 2011 publications.waset.org/7791.pdf

***Abstract*—**This paper applies fuzzy clustering algorithm in classifying real estate companies in China according to some general financial indexes, such as income per share, share accumulation fund, net profit margins, weighted net assets yield and shareholders' equity. By constructing and normalizing initial partition matrix, getting fuzzy similar matrix with Minkowski metric and gaining the transitive closure, the dynamic fuzzy clustering analysis for real estate companies is shown clearly that different clustered result change gradually with the threshold reducing, and then, it's shown there is the similar relationship with the prices of those companies in stock market. In this way, it's great valuable in contrasting the real estate companies' financial condition in order to grasp some good chances of investment, and so on.

***Keywords*—**Fuzzy clustering algorithm; Data Ming; Real Estate Company, Financial Analysis

I. Introduction

At present, real estate has played great role in the development of economy in China, in this way, it's very necessary to research on real estate companies themselves[1]. Fussy clustering Analysis is just one of important subjects.

Real estate companies can be divided into different cluster according to the financial similarity, and many valuable things can be found out on the basis of clustering outcome, for example, some chance of investment in the long term can be gotten because of the similar financial condition for the companies in the same cluster, and the stock prices maybe fluctuate similarly in the stock market. What's more, some good or bad experience may be found out by contrasting the companies in the same cluster, which is valuable for the manager to make some wise management policy. This paper, according to several financial factors, investigates into some typical companies in China, and makes fuzzy clustering analysis of them, which is of a great value for research on real estate industry.

The remainder is presented as follows. In section 2, fuzzy clustering analysis is introduced, and then in section 3, fuzzy clustering algorithm is applied to some real estate companies. The last section draws some final consideration and presents some practical implication.

Jian-feng Li is with the Department of Management Science and Engineering, Dalian Maritime University, Dalian City, China (phone: 86-0411-84729568; e-mail: lijianfeng_vvv@ yahoo.com.cn).

Feng Jin is with the Department of Management Science and Engineering, Dalian Maritime University, Dalian City, China (phone: 86-0411-84729568; e-mail: student_dlmu@ yahoo.com.cn).

Xiaoyu Yang is with Liaoning Technical College of Construction, Liaoyang City, China (phone: 86-0419-2302406; e-mail: Hello_vvv@sohu.com).

II. Fuzzy Clustering Approach

Clustering is such a procedure that objects are distinguished or classified in accordance with their similarity. A formal mathematical definition of clustering is in the follow [2]: let $X \in R^{n \times m}$ a set of data items representing a set of n points x_i in R^m. The goal is to partition X into K groups C_k such every data that belong to the same group are more "alike" than data in different groups. Each of the K groups is called a cluster. The result of the algorithm is an injective mapping $X \mapsto C$ of data items X_i to clusters C_k.

As one of important clustering approaches, fuzzy clustering analysis which obtains the uncertainty degree of samples belonging to each class and expresses the intermediate property of their memberships, can trace back to the concept of fuzzy partition which is proposed by Ruspini[3,4]. With this concept, some typical fuzzy clustering algorithms, such as methods based on the similarity and fuzzy relations [5], the transitive closure of fuzzy equivalent relation[6], the convex decomposition of data[7], or the dynamic programming and indistinguishable relation are developed one after the other by the scholars[8]. It has been applied in many aspects:

For example, in the educational aspect, students can be allocated into some number of classes by means of fuzzy clustering algorithm based on each student's achievement of prerequisite subjects (Susanto, 2002) [9], in the agricultural aspect, the soil samples can be classified on base of the concentration of thirteen chemical elements through Gustafson-Kessel fuzzy clustering algorithm (Costel, 2006)[10], and in the documental aspect, similar documents can be found through a kind of fuzzy clustering approach with the predefined fuzzy clusters being used to find extract feature vectors of the related documents (Ridvan, 2006) [11] , and so on.

At current, the research on fuzzy clustering analysis of real companies for relation with stock prices is very little, and on account of great role that real estate companies have played in the modern society in China, this paper will make use of the newest data in the store market and apply that approach to cluster the real estate companies according to their financial similarity. It is valuable for showing the development of real estate companies in China.

III. Application of Fuzzy Clustering Analysis on Real Estate Companies

A. Establishing Initial Partition Matrix

Ten real estate companies are selected, which are the listed companies in the stock market of China. Fuzzy clustering method is applied to those companies according to ten general financial indexes (attributes). They are shown below:

Companies:
X1: CHINA VANKE CO.,LTD
X2: SHENZHEN ZHENYE(GROUP) CO.,LTD
X3: SHENZHEN PROPERTIES & RESOURCES DEVELOPMENT(GROUP)
X4: CHINA MERCHANTS PROPERTY DEVELOPMENT CO.,LTD
X5: CHINA UNION HOLDINGS LTD
X6: YIHUA REAL ESTATE CO.,LTD
X7: RONGAN PROPERTY CO,.LTD
X8: CONGQING YUKAIFA CO.,LTD
X9: JIANGSU ZHONGNAN CONSTRUCTION GROUP CO.,LTD
X10: POLY REAL ESTATE GROUP CO.,LTD

Attributes:
A1: Income per Share
A2: Net assets per Share
A3: Retained earnings per share
A4: Share accumulation fund
A5: Sales margin
A6: Business profitability
A7: Net profit margins
A8: Weighted net assets yield
A9: Diluted net assets yield
A10: Shareholders' equity

In this way, the initial partition matrix can be constructed (Time: at April 2011), in which, the real estate companies are displayed horizontally and their financial indexes is displayed vertically. It is shown below:

TABLE I
INITIAL PARTITION MATRIX

Companies	*Attributes*									
	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}
X_1	0.11	4.14	1.3347	0.8044	48.95	20.69	15.12	2.68	2.65	19.35
X_2	0.069	3.29	0.7644	0.8333	46.08	14.27	10.54	2.12	2.11	30.98
X_3	0.4775	1.9456	0.7275	0.1074	72.69	38.52	29.12	27.99	24.55	38.16
X_4	0.33	10.93	4.3688	4.9406	45.14	27.12	19.82	3.04	2.99	30.17
X_5	0.022	1.52	0.2771	0.1537	59	27.83	15.31	1.48	1.47	41.79
X_6	0.0038	2.28	0.1662	1.0576	61.82	18.16	11.2	0.17	0.17	71.45
X_7	0.0498	1.8454	0.8454	0	51.42	26.03	19.82	2.74	2.7	24.41
X_8	0.0057	3.29	0.4765	1.733	50.52	0.89	6.71	0.17	0.17	48
X_9	0.17	4.12	2.4229	0.5297	24.45	12.98	9.81	4.19	4.1	19.78
X_{10}	0.15	6.43	2.4271	2.9022	50.2	26.11	19.45	2.37	2.38	17.21

B. Normalizing the Partition Matrix

Under the require of fuzzy matrix, the data should be converted into the same dimension for comparison, and compressed into the range [0,1] Following formula 1 below, the partition matrix is normalized and the result is shown in the tableII

$$x_{ik}^{'} = \frac{x_{ik} - \min\limits_{1 \le i \le n}\{x_{ik}\}}{\max\limits_{1 \le i \le n}\{x_{ik}\} - \min\limits_{1 \le i \le n}\{x_{ik}\}} \text{(k=1, 2, } \cdots \text{, m)} \quad (1)$$

Where $x_{ik}^{'} \in [0,1]$

TABLE II
NORMALIZED PARTITION MATRIX

Companies	*Attributes*									
	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}
X_1	0.224	0.278	0.278	0.163	0.508	0.526	0.375	0.09	0.102	0.039
X_2	0.138	0.188	0.142	0.169	0.448	0.356	0.171	0.07	0.08	0.254
X_3	1	0.045	0.134	0.022	1	1	1	1	1	0.386
X_4	0.689	1	1	1	0.429	0.697	0.585	0.103	0.116	0.239
X_5	0.038	0	0.026	0.031	0.716	0.716	0.384	0.047	0.053	0.453
X_6	0	0.081	0	0.214	0.775	0.459	0.2	0	0	1
X_7	0.097	0.035	0.162	0	0.559	0.668	0.585	0.092	0.104	0.133
X_8	0.004	0.188	0.074	0.351	0.54	0	0	0	0	0.568
X_9	0.351	0.276	0.537	0.107	0	0.321	0.138	0.145	0.161	0.047
X_{10}	0.309	0.522	0.538	0.587	0.534	0.67	0.568	0.079	0.091	0

C. Building up fuzzy similar matrix

In order to cluster the companies, the similarity of vectors in the sample is needed. In this way, fuzzy similar matrix R can be gotten. There are many methods to calculate the similarity. Some common formulas are shown below:

(a) Euclidean distance

$$r_{ij} = \begin{cases} 1 & \text{i = j} \\ 1 - c\sqrt{\sum\limits_{k=1}^{m}\left(x_{ik}^{'} - x_{jk}^{'}\right)^2} & \text{i} \ne \text{j} \end{cases} \quad (2)$$

$i, j = 1, 2, \cdots, n$

C is endowed with proper value for. $0 \le r_{ij} \le 1$

(b) Minkowski metric

$$r_{ij} = \begin{cases} 1 & \text{i = j} \\ 1 - c\left\{\sum\limits_{k=1}^{m}\left|x_{ik}^{'} - x_{jk}^{'}\right|^p\right\}^{\frac{1}{p}} & \text{i} \ne \text{j} \end{cases} \quad (3)$$

$i, j = 1, 2, \cdots, n$

c is endowed with proper value for. $0 \le r_{ij} \le 1$

p is positive integer, when p=2, it is *Euclidean distance*

(c) Dot product

$$r_{ij} = \begin{cases} 1 & i = j \\ 1 - \frac{1}{M} \sum_{k=1}^{m} x'_{ik} \cdot x'_{jk} & i \neq j \end{cases} \quad (4)$$

$i, j = 1, 2, \cdots, n$

$$M = \max(\sum_{i \neq j} x'_{ik} \cdot x'_{jk})$$

(d) Cosine

$$r_{ij} = \frac{\sum_{k=1}^{m} x'_{ik} \cdot x'_{jk}}{\sqrt{\sum_{k=1}^{m} {x'_{ik}}^2} \cdot \sqrt{\sum_{k=1}^{m} {x'_{jk}}^2}} \quad (5)$$

$i, j = 1, 2, \cdots, n$

(e) Correlation

$$r_{ij} = \frac{\sum_{k=1}^{m} \left| x'_{ik} - \overline{x'_{ik}} \right| \left| x'_{jk} - \overline{x'_{jk}} \right|}{\sqrt{\sum_{k=1}^{m} \left(x'_{ik} - \overline{x'_{ik}} \right)^2} \cdot \sqrt{\sum_{k=1}^{m} \left(x'_{jk} - \overline{x'_{jk}} \right)^2}} \quad (6)$$

$i, j = 1, 2, \cdots, n$

$$\overline{x'_i} = \frac{1}{m} \sum_{k=1}^{m} x'_{ik}, \ \overline{x'_j} = \frac{1}{m} \sum_{k=1}^{m} x'_{jk}$$

This paper follows formula 3 (*Minkowski metric)* to built up fuzzy similar matrix, where the parameters p and c are taken separately as 6 and 0.4. The result is shown in the table III below.

TABLE III
Fuzzy Similar Matrix(R)

Companies	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
X_1	1	0.9	0.58	0.63	0.83	0.62	0.9	0.76	0.8	0.83
X_2	0.9	1	0.55	0.6	0.85	0.7	0.83	0.85	0.81	0.8
X_3	0.58	0.55	1	0.52	0.54	0.51	0.57	0.48	0.54	0.58
X_4	0.63	0.6	0.52	1	0.53	0.55	0.54	0.58	0.62	0.78
X_5	0.83	0.85	0.54	0.53	1	0.78	0.87	0.71	0.7	0.74
X_6	0.62	0.7	0.51	0.55	0.78	1	0.65	0.8	0.6	0.6
X_7	0.9	0.83	0.57	0.54	0.87	0.65	1	0.71	0.76	0.75
X_8	0.76	0.85	0.48	0.58	0.71	0.8	0.71	1	0.75	0.7
X_9	0.8	0.81	0.54	0.62	0.7	0.6	0.76	0.75	1	0.76
X_{10}	0.83	0.8	0.58	0.78	0.74	0.6	0.75	0.7	0.76	1

D. Obtaining fuzzy equivalent matrix

According to the fuzzy clustering approach, the matrix must have three qualities: "reflexivity", "symmetry" and "transitivity", which isn't all owed by fuzzy similar matrix, however, through transitive closure, Fuzzy similar matrix R can be transformed into fuzzy equivalent matrix, which not only have "reflexivity" and "symmetry", but also is provided with "transitivity".

The transitive closure $t(R)$ is gained by means of the "square method" that fuzzy similar matrix is squared gradually.

$$R \to R^2 \to R^4 \to \cdots \to R^{2^i} \to \cdots \quad (7)$$

$$(\ R^2 = R \circ R = \vee_{k=1}^{n} (r_{ik} \wedge r_{kj}) \ )$$

When $R^k \circ R^k = R^k$, it means that R^k is provided with "transitivity", and R^k is just requisite transitive closure which is also the "optimal" fuzzy equivalent matrix R^*. The optimal fuzzy equivalent matrix is shown in table IV below.

TABLE IV
Optimal Fuzzy Equivalent Matrix(R^*)

Companies	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
X_1	1	0.9	0.58	0.78	0.87	0.8	0.9	0.85	0.81	0.83
X_2	0.9	1	0.58	0.78	0.87	0.8	0.9	0.85	0.81	0.83
X_3	0.58	0.58	1	0.58	0.58	0.58	0.58	0.58	0.58	0.58
X_4	0.78	0.78	0.58	1	0.78	0.78	0.78	0.78	0.78	0.78
X_5	0.87	0.87	0.58	0.78	1	0.8	0.87	0.85	0.81	0.83
X_6	0.8	0.8	0.58	0.78	0.8	1	0.8	0.8	0.8	0.8
X_7	0.9	0.9	0.58	0.78	0.87	0.8	1	0.85	0.81	0.83
X_8	0.85	0.85	0.58	0.78	0.85	0.8	0.85	1	0.81	0.83
X_9	0.81	0.81	0.58	0.78	0.81	0.8	0.81	0.81	1	0.81
X_{10}	0.83	0.83	0.58	0.78	0.83	0.8	0.83	0.83	0.81	1

E. Describing dynamic fuzzy clustering

Setting the threshold to λ, which changes from the big to the small, the dynamic fuzzy clustering can result from the above optimal fuzzy equivalent matrix R^*.

Supposing $\underset{\sim}{C} \in P(U)$, for $\forall \lambda \in [0,1]$,

$$(\underset{\sim}{C})_\lambda = C_i \triangleq \{x | \underset{\sim}{C}(x) \ge \lambda\}, \quad (8)$$

The last, with the different threshold λ being set on, the dynamic fuzzy clustering for the real estate companies based on finance condition is shown according to the formula 8. According to the above analysis, the dynamic fuzzy clustering map can be gained, which is shown in Fig1.

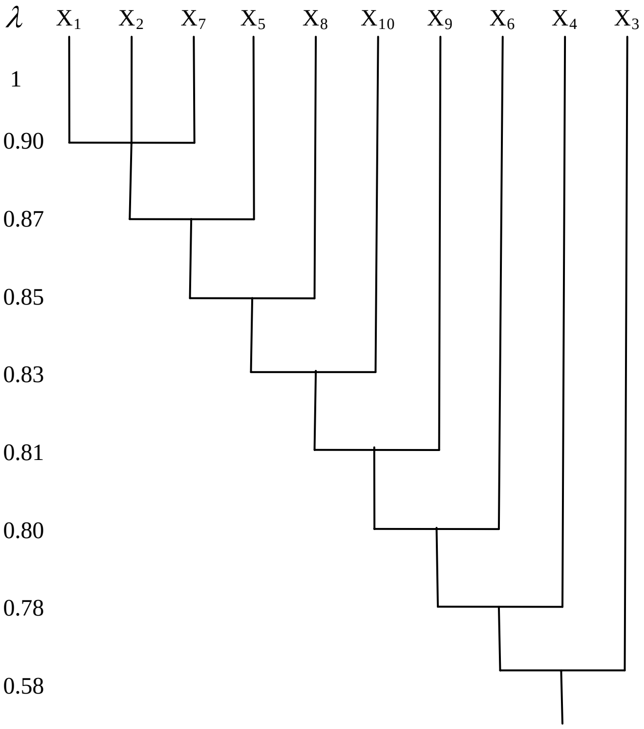

Fig1 Dynamic fuzzy clustering map

It means that in some level (λ) some real estate companies can gather together depending on the financial condition, such as the companies X_1, X_2 , X_7 are similar at the threshold where λ =0.90, what's more, with the decrease of λ , the clustered companies will increase gradually

When $\lambda = 1$, the real estate companies can be classified into ten clusters: $\{X_1\}, \{X_2\}, \{X_3\}, \{X_4\}, \{X_5\},\{X_6\},\{X_7\},\{X_8\}, \{X_9\}, \{X_{10}\}$.

When $\lambda = 0.9$, the real estate companies can be classified into eight clusters: $\{X_1, X_2, X_7\}, \{X_5\}, \{X_8\},\{X_{10}\},\{X_9\},\{X_6\}, \{X_4\}, \{X_3\}$.

When $\lambda = 0.87$, the real estate companies can be classified into seven clusters: $\{X_1, X_2, X_7, X_5\}, \{X_8\},\{X_{10}\},\{X_9\},\{X_6\}, \{X_4\}, \{X_3\}$.

When $\lambda = 0.85$, the real estate companies can be classified into six clusters: $\{X_1, X_2, X_7, X_5, X_8\},\{X_{10}\},\{X_9\},\{X_6\}, \{X_4\}, \{X_3\}$.

When $\lambda = 0.83$, the real estate companies can be classified into five clusters: $\{X_1, X_2, X_7, X_5, X_8, X_{10}\},\{X_9\},\{X_6\}, \{X_4\}, \{X_3\}$.

When $\lambda = 0.81$, the real estate companies can be classified into four clusters: $\{X_1, X_2, X_7, X_5, X_8, X_{10},X_9\},\{X_6\}, \{X_4\}, \{X_3\}$

When $\lambda = 0.80$, the real estate companies can be classified into three clusters: $\{X_1, X_2, X_7, X_5, X_8, X_{10},X_9,X_6\}, \{X_4\}, \{X_3\}$

When $\lambda = 0.78$, the real estate companies can be classified into two clusters: $\{X_1, X_2, X_7, X_5, X_8, X_{10},X_9,X_6, X_4\}, \{X_3\}$

When $\lambda = 0.58$, the real estate companies can be classified into one clusters: $\{X_1, X_2, X_7, X_5, X_8, X_{10},X_9,X_6, X_4, X_3\}$

Approximately, it is related to the average prices in 30 day of those companies in stock market(May 6, 2011), which is shown in Table V and Fig 2 below.

TABLE V
STOCK PRICES OF REAL ESTATE COMPANIES

Companies	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
Price	8.63	8.04	8.69	18.16	4.36	5.66	8.51	8.67	12.56	13.64

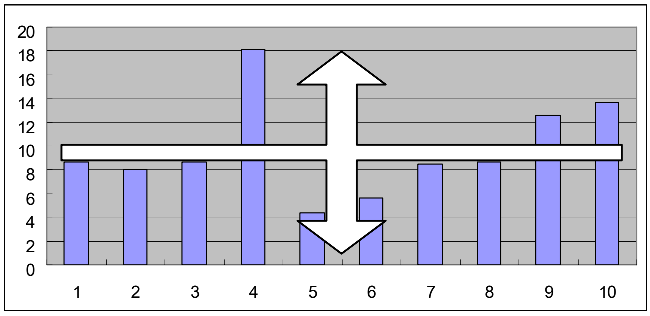

Fig. 2 The Stock Prices of those Companies

Here, most companies' prices fluctuation is similar with the dynamic fuzzy clustering map. For example, the stock prices of companies X1, X2, X7 and X8 are similar just like they are near in the dynamic fuzzy clustering map. What's more, they are close to the companies X10, X9, X6 and X4 with the decrease of threshold λ in the map.

That novel and potentially discipline is shown clearly with enough accuracy, although some companies such as X5 and X3 are not totally fit for that discipline, but there are some reasons to explain it. For the company X5, its tradable share is very large, on account of which, the stock price is lower relatively, what's more, there are large change on that company's operation, so it belongs to a different example in a special time. In addition, for the company X3, its current stock price is rising up gradually, and it means good investment opportunity, which is shown in Fig 3.

Fig. 3 The Stock Price of Company X3

In a word, dynamic fuzzy clustering analysis of real estate companies is with good precision above, and it's valuable to grasp the investment opportunity and so on.

IV. Conclusion

As an important mathematic tool, fuzzy clustering is always receiving great attention from enterprisers and scholars, and it has been applied broadly in many aspects. In this paper, it is used to classify the real estate companies based on the some general financial indexes. The dynamic fuzzy classification of companies with the change of different threshold is shown clearly, which has a big value on contrasting the real estate companies' financial condition in order to grasp the chance of investment and so on.

References

[1] Gao Bo; Wang Huilong. "An Empirical Analysis of the Impact on Residents Consumption of Housing Price Fluctuations in Yangtze River Delta". *Industrial Economics Research*, Vol 1, 2011, pp5-14
[2] T. Graepel. "Statistical physics of clustering algorithms". *Technical report 171822*, FB Physik, Institut fur Theoretische Physic, 1998.
[3] K.Mali, S.Mitra, "Clustering of symbolic data and its validation, Advances in soft computing-AFSS 2002", *Springer-Verlag*, Heidelberg, 2002, pp339-344
[4] Ruspini, E. H., "A new approach to clustering", *Inf. Cont*., Vol.15, 1969, pp22
[5] Tamra, S. et al., "Pattern classification based on fuzzy relations", *IEEE SMC*, Vol.1, No.1, 1971, pp217
[6] Zkim, L, "Fuzzy relation compositions and pattern recognition", *Inf. Sci.*, Vol.89, 1996, pp107
[7] R.-N.P. Singh, W.H. Bailey, "Fuzzy logic applications to multisensor multitargent correlation", *IEEE Transa. Aero space Electron. Systems*, Vol.33, No3, 1997, pp752-769
[8] GAO Xinbo "Advances in theory and applications of fuzzy clustering". *Chinese science bulletin*. Vol.45. No.11, 2000, pp961-970
[9] Sani Susanto, Igu Suharto, Paulus Sukapto, "Using fuzzy clustering algorithm for allocation of students", *World transaction on engineering and technology education*, Vol.1, 2002, pp245-248
[10] Costel Sarbu, Katharina Zehl, "Fuzzy divisive hierarchical clustering of soil data using gustafson-kessel algorithm", *Laboratory systems*, August, 2006
[11] Ridvan Sracoglu, Novruz Allahverdi, "A fuzzy clustering approach for finding similar documents using a novel similarity measure". *Expert systems with applications*, July, 2006