International Journal of Mathematics And its Applications
Volume 5, Issue 4–C (2017), 421–427.
ISSN: 2347-1557
Available Online: http://ijmaa.in/

International Journal of Mathematics And its Applications

Fuzzy Clustering Analysis of Selected IT Companies in India

Research Article

C. Loganathan[1], Annakkodi[1]* and S. Rangasamy[2]

1 Department of Mathematics, Maharaja Arts and Science College, Coimbatore, Tamilnadu, India.

2 Department of Economics, Maharaja Arts and Science College, Coimbatore, Tamilnadu, India.

Abstract: This paper applies fuzzy clustering algorithm in classifying IT companies in India according to some general financial indexes, such as Equity capital, Earning per Share, Price to book value, net profit margin and dividend payout. 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 IT companies is shown clearly that different clustered results change gradually with the threshold reducing, and then, it's shown that there is similar relationship with the prices of those companies in the stock market. In this way, it is greatly valuable in contrasting the IT companies financial conditions in order to identify some good opportunities of investments.

Keywords: Fuzzy clustering algorithm, Data Mining, Financial Analysis.

1. Introduction

At present, IT companies play a predominant role in the development of the economy in India. In this scenario, it is very necessary to carryout research on IT companies and investment decisions on them. Fuzzy clustering analysis is just one of the important mathematical areas of research which will throw right on criteria to take mathematics oriented and sound financial decisions in respect of investments on stocks and shares of IT companies. Investors are required to take decisions in a highly uncertain and volatile situations, very often looming large in stock market. In order to facilitate them to take profitable and pertinent decisions, fuzzy clustering appears to be a more appropriate approach. Therefore the selected IT companies have been divided into different clusters according to the financial similarity, and many valuable conclusions could be arrived at on the basis of the outcome of clustering and then applied to the IT industry as a whole. For example, some chances of long term investments can be found out by identifying the similar financial conditions of the companies in the same cluster, and the reasons for the fluctuations of stock prices in the stock market. Clustering will enrich the experience and enable the investors to make some wise investment decisions in the midst of significantly discernible uncertainties in the volatile stock market. This paper, accordingly analyses several financial factors, investigates with reference to some typical IT companies in India, and makes fuzzy clustering which is of great value to the investors for identifying the possibilities of earning maximum profit on investments in IT companies. In Section 2, Fuzzy clustering analysis is introduced, and then in Section 3, Fuzzy clustering algorithm is applied to selected IT companies. The last section deals with some final conclusions and presents some practical implications of applications.

* *E-mail: akkodi99@gmail.com*

2. Fuzzy Clustering Algorithm

Clustering is such a procedure wherein attributes are distinguished or classified in accordance with their similarity. A formal mathematical definition of clustering is as follows:
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 Kgroups C_k such that data that belong to the same group are more "alike" than data in different groups. Each of the Kgroup is called a cluster. The result of the algorithm is an injective mapping $X \mapsto C$ of data items X_i to clustersC_k. As one of the important clustering algorithms, fuzzy clustering analysis which obtains the uncertain degree of samples belonging to each class and expresses the intermediate property of the members, can trace back to the concept of fuzzy partition. With this concept some typical fuzzy clustering algorithms, such as methods based on the similarity and fuzzy relations, the transitive closure of fuzzy equivalent relation, the convex decomposition of data, or the dynamic programming. At present, the research on fuzzy clustering analysis of operating companies for explaining relationship with stock prices is very little, and on account of great role that IT companies have played in the modern society in India, this paper will make use of the newest data in the stock market and apply that approach to cluster the IT companies according to their financial similarity and will be valuable for guiding the corporate policy makers as well as investors in the stocks and shares of the clustered companies.

3. Application of Fuzzy Clustering Analysis on IT Companies

3.1. Establishing Initial Partition Matrix

Ten IT companies, which are listed in the stock markets of India have been selected. Fuzzy clustering method is applied to those ten companies according to five general financial parameters. The companies selected and analyzed are given below:

Companies:

X_1 : TCS

X_2 : Infosys

X_3 : Wipro

X_4 : HCL Technology

X_5 : Tech Mahindra

X_6 : Oracle Financial Services

X_7 : Mphasis Group

X_8 : Mind Tree

X_9 : Polaris

X_{10} : Rolta India

Attributes:

A_1 : Equity Capital

A_2 : Earning per share

A_3 : Price to book value

A_4 : Net Profit Margin

A_5 : Dividend payout

In this way, the initial partition matrix can be constructed where in the IT companies are displayed horizontally and their financial parameters are displayed vertically. They are shown below:

Company	Attributes				
	A_1	A_2	A_3	A_4	A_5
X_1	19.8	109.7	10.5	22.2	32
X_2	28.6	41.8	1.12	24.79	8.3
X_3	49.32	17.1	2.25	30.90	4
X_4	14.01	35.65	3	21.4	5
X_5	24	25.52	1.36	8.75	4
X_6	4.2	31.98	0.64	35.05	0
X_7	21.01	3.03	0.17	15.65	1.7
X_8	8.37	7.65	0.61	11.9	0.55
X_9	4.992	2.154	0.2	10.62	1.25
X_{10}	16.13	2.14	0.15	21.76	0.225

Table 1. **Initial Partition Matrix**

3.2. Normalizing the Partition Matrix

Under fuzzy matrix, the data should be converted into the same dimension for comparison, and compressed within the range [0,1]. In the formula 1 below given, the partition matrix is normalized and the result is shown in the table 2

$$x'_{ik} = \frac{x_{ik} - \min\{x_{ik}\}}{\max\limits_{1\leq i\leq n}\{x_{ik}\} - \min\limits_{1\leq i\leq n}\{x_{ik}\}}\ (k = 1, 2, \ldots, m) \tag{1}$$

where $x'_{ik} \in [0,1]$

Company	Attributes				
	A_1	A_2	A_3	A_4	A_5
X_1	0.1785	1.0000	0.0957	0.2024	0.2917
X_2	0.2607	0.3781	0.0102	0.2260	0.0757
X_3	0.4496	.1559	0.0205	0.2817	0.0365
X_4	0.1277	0.3250	0.0273	0.1951	0.0456
X_5	0.2188	0.2326	0.0124	0.0798	0.0365
X_6	0.0383	0.2915	0.0058	0.3195	0
X_7	0.1915	0.0276	0.0015	0.1427	0.0155
X_8	0.0763	0.0697	0.0056	0.1085	0.0050
X_9	0.0455	0.0196	0.0018	0.0968	0.0114
X_{10}	0.1470	0.0195	0.0014	0.1984	0.0021

Table 2. **Normalized Partition Matrix**

3.3. 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 has been built. There are many methods to calculate the similarity. Some common formulas are shown below:

(a). **Euclidean distance**

$$r_{ij} = \begin{cases} 1, & i = j; \\ 1 - c\sqrt{\sum\limits_{k=1}^{m}\left(x'_{ik} - x'_{jk}\right)^2}, & i \neq j. \end{cases} \tag{2}$$

$i, j = 1, 2, \ldots, n.$ c is endowed with proper value for $0 \leq r_{ij} \leq 1$.

(b). **Minkowski Metric**

$$r_{ij} = \begin{cases} 1, & i = j; \\ 1 - c\left\{\sum\limits_{k=1}^{m}\left|x'_{ik} - x'_{jk}\right|^p\right\}^{\frac{1}{p}}, & i \neq j. \end{cases} \tag{3}$$

$i, j = 1, 2, \ldots, n$. c is endowed with proper value for $0 \leq r_{ij} \leq 1$. P is positive integer, when $p = 2$, it is Euclidean distance

(c). **Dot Product**

$$r_{ij} = \begin{cases} 1, & i = j; \\ 1 - \frac{1}{M} \sum_{k=1}^{m} x'_{ik}.x'_{jk}, & i \neq j. \end{cases}$$

$i, j = 1, 2, \ldots, n$, $M = \max \left(\sum_{i \neq j} x'_{ik} \; o \; x'_{jk} \right)$.

(d). **Cosine**

$$r_{ij} = \frac{\sum_{k=1}^{m} x'_{ik} \; o \; x'_{jk}}{\sqrt{\sum_{k=1}^{m} {x'_{ik}}^2} \; o \; \sqrt{\sum_{k=1}^{m} {x'_{jk}}^2}}$$

$i, j = 1, 2, \ldots, n$.

(e). **Correlation**

$$r_{ij} = \frac{\sum_{k=1}^{m} \left| x'_{ik} - \bar{x}'_{jk} \right| \left| x'_{jk} - \bar{x}'_{jk} \right|}{\sqrt{\sum \left(x'_{ik} - \bar{x}'_{ik} \right)^2 o \left(\sqrt{\sum \left(x'_{jk} - \bar{x}'_{jk} \right)^2} \right)}}$$

$i, j = 1, 2, \ldots, n$; $\bar{x}'_i = \frac{1}{m} \sum_{k=1}^{m} x'_{ik}$; $\bar{x}'_j = \frac{1}{m} \sum_{k=1}^{m} x'_{jk}$.

This paper follows Minkowski metric formula to build 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 3 below.

Company	Attributes									
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
X_1	1.0000	0.7512	0.6622	0.7299	0.6930	0.7164	0.6110	0.6278	0.6078	0.6078
X_2	0.7512	1.0000	0.9062	0.9468	0.9345	0.9109	0.8598	0.8757	0.8555	0.8565
X_3	0.6622	0.9062	1.0000	0.8708	0.9018	0.8354	08961	0.8504	0.8381	0.8788
X_4	0.7299	0.9468	0.8708	1.0000	0.9506	0.9491	0.8811	0.8979	0.8778	0.8778
X_5	0.6930	0.9345	0.9018	0.9506	1.0000	0.9014	0.9180	0.9307	0.9111	0.9143
X_6	0.7164	0.9109	0.8354	0.9491	0.9014	1.0000	0.8923	0.9027	0.8864	0.8910
X_7	0.6110	0.8598	0.8961	0.8811	0.9180	0.8923	1.0000	0.9539	0.9416	0.9768
X_8	0.6278	0.8757	0.8504	0.8979	0.9307	0.9027	0.9539	1.0000	0.9798	0.9626
X_9	0.6078	0.8555	0.8381	0.8778	0.9111	0.8864	0.9416	0.9798	1.0000	0.9544
X_{10}	0.6078	0.8565	0.8788	0.8778	0.9143	0.8910	0.9768	0.9626	0.9544	1.0000

Table 3. **Fuzzy Similar Matrix (R)**

3.4. Obtaining Fuzzy Equivalent Matrix

According to the fuzzy clustering approach , the matrix must have three qualities: "reflexivity", "Symmetry" and " transitivity", which are not allowed by fuzzy similar matrix. However, through transitive closure, Fuzzy similar matrix R can be transformed into fuzzy equivalent matrix, which has not only "reflexivity" and "Symmetry", but also is provided with "transitivity". The transitive closure $t(R)$ is gained by means of the "squared method" and hence fuzzy similar matrix is squared gradually.

$$R \rightarrow R^2 \rightarrow R^4 \rightarrow \cdots \rightarrow R^{2^i}$$

$$(R^2 = Ro\, R = \bigvee_{k=1}^{n} (r_{ik} \wedge r_{kj}))$$

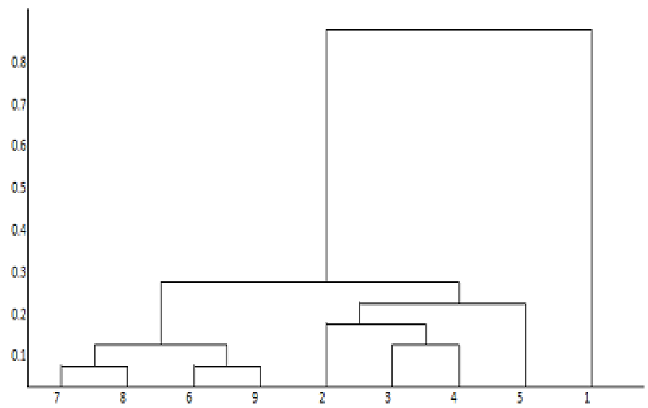
where $R^k o\, R^k = R^{2k}$, which means that R^k is provided with " transitivity", and R^k is just requisite transitive closure which is also the "optimal" fuzzy equivalent matrix . The optimal fuzzy equivalent matrix is shown in Table 4 below.

Company	Attributes									
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
X_1	1.0000	0.7512	0.6622	0.7299	0.6930	0.7164	0.6110	0.6278	0.6078	0.6078
X_2	0.7512	1.0000	0.9062	0.9468	0.9345	0.9109	0.8598	0.8757	0.8555	0.8565
X_3	0.6622	0.9062	1.0000	0.8708	0.9018	0.8354	0.8961	0.8504	0.8381	0.8788
X_4	0.7299	0.9468	0.8708	1.0000	0.9506	0.9491	0.8811	0.8979	0.8778	0.8778
X_5	0.6930	0.9345	0.9018	0.9506	1.0000	0.9014	0.9180	0.9307	0.9111	0.9143
X_6	0.7164	0.9109	0.8354	0.9491	0.9014	1.0000	0.8923	0.9027	0.8864	0.8910
X_7	0.6110	0.8598	0.8961	0.8811	0.9180	0.8923	1.0000	0.9539	0.9416	0.9768
X_8	0.6278	0.8757	0.8504	0.8979	0.9307	0.9027	0.9539	1.0000	0.9798	0.9626
X_9	0.6078	0.8555	0.8381	0.8778	0.9111	0.8864	0.9416	0.9798	1.0000	0.9544
X_{10}	0.6078	0.8565	0.8788	0.8778	0.9143	0.8910	0.9768	0.9626	0.9544	1.0000

Table 4. **Optimal Fuzzy Equivalent Matrix** (R^*)

3.5. Description of 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^*.



It means that in some level (λ) some IT companies can gather together depending on the financial condition, such as the companies $\{X_7, X_8\}$, $\{X_6, X_9\}$ and $\{X_3, X_4\}$ are similar at the threshold. Assume that $\lambda = 0.9539$ what more, with the decrease of λ the clustered companies will increase gradually.

When $\lambda = 1$, the IT 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 $\lambda = 0.9539$ the IT companies can be classified into nine clusters:

$$\{X_7, X_8\},\ \{X_1\},\ (X_2\},\ \{X_3\},\ \{X_4\}, \{X_5\}, \{X_6\}, \{X_9\},\ \{X_{10}\}$$

When $\lambda = 0.8864$ the IT companies can be classified into eight clusters:

$$\{X_7, X_8\},\ \{X_6, X_9\},\ \{X_1\},\ \{X_2\},\ \{X_3\},\ \{X_4\},\ \{X_5\},\ \{X_{10}\}$$

When $\lambda = 0.9201$ the IT companies can be classified into seven clusters:

$$\{X_7, X_8, X_6, X_9\},\ \{X_1\},\ \{X_2\},\ \{X_3\},\ \{X_4\}, \{X_5\},\ \{X_{10}\}$$

When $\lambda = 0.8708$ the IT companies can be classified into six clusters:

$$\{X_7, X_8, X_6, X_9\},\ \{X_3, X_4\},\ \{X_1\},\ (X_2\},\ \{X_5\},\ \{X_{10}\}$$

When $\lambda = 0.8885$ the IT companies can be classified into five clusters:

$$\{X_7, X_8, X_6, X_9\},\ \{X_2, X_3, X_4\},\ \{X_1\}, \{X_5\},\ \{X_{10}\}$$

When $\lambda = 0.9092$ the IT companies can be classified into four clusters:

$$\{X_7, X_8, X_6, X_9\},\ \{X_2, X_3, X_4, X_5\}, \{X_1\},\ \{X_{10}\}$$

When $\lambda = 0.8978$ the IT companies can be classified into three clusters:

$$\{X_7, X_8, X_6, X_9,\ X_2, X_3, X_4\},\ \{X_1\},\ \{X_{10}\}$$

When $\lambda = 0.7538$ the IT companies can be classified into one cluster:

$$\{X_7, X_8, X_6, X_9,\ X_2, X_3, X_4,\ X_1,\ X_{10}\}$$

4. Conclusion

Stock market is invariably and very often erratically volatile. Policy makers to regulate the market and investors to make investment decision in highly volatile scenario, some guiding force is required. Fuzzy clustering built up and recommended in this article will help very much to identify the companies with similar attributes and substantiate the reasons for why certain companies should be preferred for investment. Fuzziness calculated in between the extremes of volatility can be narrowed down step by step and the clustering of companies on the basis of similar attributes will help the investors to minimize the risk of loss and maximize the chances of earning significant profit. Fuzzy Logic investment support has been exclusively elicited by taking into account of capital structure, earning per share, price to book value, net profit margin and dividend payouts the research may further be expanded to include, apart from these endogenous parameters, exogenous parameters like GDP (Gross Domestic Product), per-capita disposable income, financial policy of the government, inflation etc. There is still enormous scope for making the research, all inclusive of endogenous and exogenous parameters and more pertinent clustering can be done to guide the investors specifically and policy makers in general.

References

[1] Gao Bo and Wang Huilong, *An Empirical Analysis of the Impact on Residents Consumption of Housing Price Fluctuation in Yangtze River Delta*, Industrial Economics Research, 1(2011), 5-14.

[2] K.Mali and S.Mitra, *Clustering of symbolic data and its validation*, Advances in Soft Computing, Springer-Verlag, Heidelberg, (2002), 339-344.

[3] E.H.Ruspini, *A new approach to clustering*, Inf. Cont., 15(1969), 22-32.

[4] S.Tamra et al., *Pattern classification based on fuzzy relations*, IEEE SMC, 1(1)(1971).

[5] GAO Xinbo, *Advances in theory and applications of fuzzy clustering*, Chinese Science Bulletin, 45(11)(2000), 961-970.

[6] L.Zkim, *Fuzzy Relation Composition and Pattern Recognition*, Inf. Scin., 89(1996), 107.

[7] Sani Susanto, Igu Suharto and Paulus Sukaptio, *Using fuzzy Clustering Algorithm for Allocation of Students*, World Transaction on Engineering and Technology Education, 1(2002), 245-248.

[8] Jijian Xie, *Fuzzy Mathematics Method and Application*, People Publising Company, (2005).