A Study on the Analysis of Stability Indicators in Financial Statements Using Fuzzy C-Means Clustering

Min Sun Kim*, Eun Soo Choi*, Jae Yeul Lee, Min Soo Kang*****

Department of Medical IT Marketing, Eulji University,
553 Sanseong-Daero, Sujeong-gu, Seongnam-si, Gyeonggi-do, 13135, Korea.
**Da Vinchi College of General   Education, Chung-Ang University*
*** Department of Medical IT, Eulji University,*
553 Sanseong-Daero, Sujeong-gu, Seongnam-si, Gyeonggi-do, 13135, Korea.

Co-corresponding author: Eun Soo Choi.
Co-corresponding author: Min Soo Kang.

Abstract

IoT and A.I researches are active as the 4th industrial revolution comes. Innovative companies have already applied artificial intelligence to some functions. In addition, there will be an increasing number of companies that apply artificial intelligence to business analysis, and in the future, business management such as management and management finance experts will be replaced by artificial intelligence. In this paper, we selected 334 corporations which are corporations with total assets of 12 billion more out of the list specified by Korea Information & Communication Construction Association. We calculated the debt ratio, current ratio, BIS ratio from financial statement. These three things are stability indices. The debt ratio, the current ratio, and BIS ratio, which are key business analysis indicators of the stability indices, are clustered and analyzed by applying the Fuzzy C-Means clustering algorithm based on Fuzzy theory. As a result of clustering, we could see that it was divided into a certain section.

Keywords: Stability Indicators, Machine Learning, Fuzzy C-Means , Clustering

INTRODUCTION

Recently, IoT and A.I. researches are active as the 4th industrial revolution comes. Innovative companies have already applied artificial intelligence to some functions. In addition, there are many companies that want to develop artificial intelligence in the field of business analysis. One of them, SMACC, is starting service from start-up to large corporation by applying artificial intelligence [1]. According to the Korea Employment Information Service, 47.2% of administration and management support staff will be replaced by artificial intelligence in 2025, and 55.2% [2]. The Korea International Trade Association has introduced the AI-based data analysis and management system for enterprise management such as human resource management, finance, marketing, and production planning beyond the boundary between IT industry and non-IT industry As the technique was introduced, it was expected that efficiency would be significantly improved [3]. In this paper, 334 corporations which have total assets over 12 billion were selected and applied. we applied and analysis Fuzzy theory based Fuzzy C-means clustering algorithm.

A STUDY ON FUZZY THEORY BASED BUSINESS ANALYSIS DATA CLUSTERING

Business analysis data sets

The dataset was sampled from 334 corporations in the IT and construction industry with over 12 billion assets disclosed in the Data Analysis, Retrieval and Transfer System [4].

$$Debt\ Ratio = \frac{Total\ Debt}{Total\ OwnerShip\ Interest} \times 100 \qquad (1.1)$$

$$Current\ Ratio = \frac{Current\ Assets}{Current\ Debt} \times 100 \qquad (1.2)$$

$$BIS\ Ratio = \frac{Equity\ Capital}{Total\ Assets} \times 100 \qquad (1.3)$$

The data used in the financial statements are Total Debt, Total Ownership Interest, Current Assets, Current Debt, Equity Capital, and Total Assets. The data of the financial statements are calculated by the above equations (1.1), (1.2), and (1.3), and the data set includes the Debt Ratio, the Current Ratio, and the BIS Ratio, which are the stability indices of the management of the IT and construction industry [5].

Fuzzy C-Means

Fuzzy Theory is an attempt to solve the difficulties of bringing a computer closer to a human being. In the case of a computer, numerical and symbol processing is used to perform unambiguous processing, whereas human behavior is to use a lot of ambiguous information This is a mathematical

theory that deals with ambiguity [6]. Zadeh wrote "Fuzzy Sets" to solve precisely and complicated control theory [7]. Fuzzy C-Means (FCM) is one of the typical clustering algorithms that have been used since the 1970s [8].

Table 1. Algorithm of Clustering

	Fuzzy C-Means [8]	**K-Means [9]**	**DBSCAN [10]**
Range of Clustering	Whole	Whole	Range of Eplison
Judgement of Clustering	Distance	Distance	Range of Eplison Range of minPts
Number of Clustering	C (1≦C)	K (1≦K)	N (0≦N)
Probability of Clustering	0 ~ 1	0 or 1	NULL or 0 or 1

Table 1 compares the clustering algorithms K-Means and DBSCAN. The method of clustering based on distance and Centroid as a whole is similar to that of K-means, but it is different in that the probability of belonging to a cluster is stochastically approaching from 0 to 1.

FCM clustering is an algorithm that divides a set of N vectors $X_i (i = 1, 2, ..., N)$ into C clusters and finds the center value in each cluster generated when the objective function is minimum.

$$J(V, U \mid X) = \sum_{i=1}^{N} \sum_{k=1}^{C} u_{ik}^m d_{ik}^2 \qquad (2.1)$$

$$d_{ik} = d(x_k - v_i) = \left[\sum_{i=1}^{l} (x_k - v_i)^2 \right]^{l/2} \qquad (2.2)$$

$$v_i = \frac{\sum_{k=1}^{N} (u_{ik})^m x_k}{\sum_{k=1}^{N} (u_{ik})^m} \qquad (2.3)$$

$$u_{ik} = \left(\sum_{i=1}^{C} \left(\frac{d_{ik}}{d_{ij}} \right)^{2/(m-1)} \right)^{-1} \qquad (2.4)$$

The objective function of the FCM algorithm is shown in Equation (2.1). $m(1 \le m \le \infty)$ is a constant indicating the degree of fuzzification, d_{ik} is a measure value indicating the distance between x_k and v_i, and Euclidean distance is used in FCM. Equations (2.2), (2.3), and (2.4) show parameter values in the objective function. The objective function of Equation (2.1) is optimized by using an optimization algorithm, and the update equations (2.3) and (2.4) can be obtained by using the Largrange Multiplier Method.

EXPERIMENT AND RESULT

Experiment Environment

The dataset was sampled from 334 corporations in the IT and construction industry with over 12 billion assets disclosed in the Data Analysis, Retrieval and Transfer System .

In addition, for the management ratios, Equation 1-3 was applied to test the csv file in conjunction with R. Experimental results were performed using R and libraries e1071 and scatterplot3d [11].
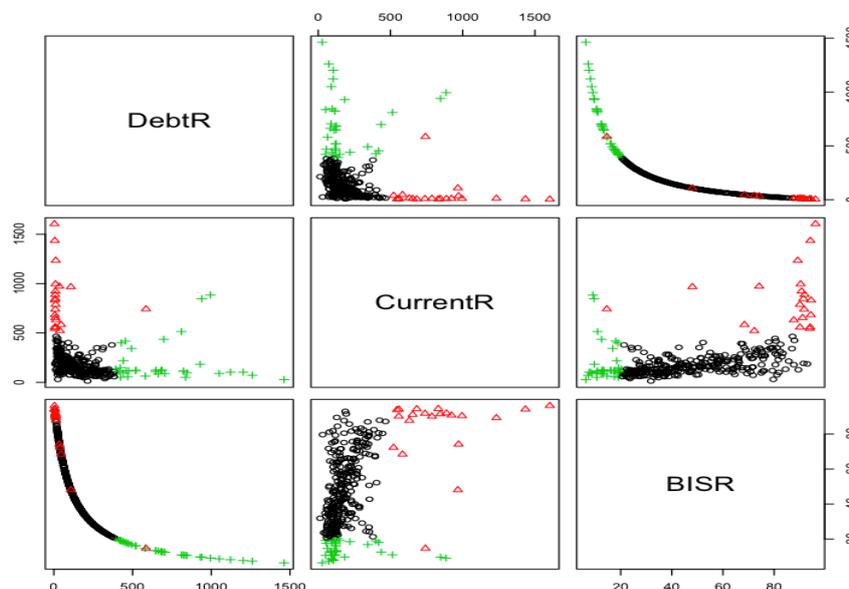
Experiment Result

Figure 1. Cluster of Business Analysis

Figure 1 above shows the results of the management analysis using Fuzzy C-Means. As a characteristic point of the distribution of data, it can be seen that the debt ratio and the current ratio have an inverse function. The higher the equity ratio, the lower the debt ratio.

Table 2. Range of Cluster

	Debt & Current	Debt & BIS	Current & BIS
Cluster 1	Current <500 Debt<400	20<BIS<90	20<BIS<90
Cluster 2	500<Current <1500+	90<BIS	90<BIS
Cluster 3	400<Debt <1500+	BIS<20	BIS<20

In Table 2 and Figure 1, when the debt ratio and the current ratio are appropriate, the debt ratio and the current ratio cluster are characteristic points. When the debt ratio is noticeably large, Can be seen. It can be seen that when the debt ratio and the current ratio are clustered by the FCM, other clusters are detected as the other clusters. In addition, it can be seen that when the BIS ratio, the debt ratio, and the current ratio are clustering, the same interval is created based on the capital adequacy ratio, so that a company that falls outside the cluster scope may have a management problem.
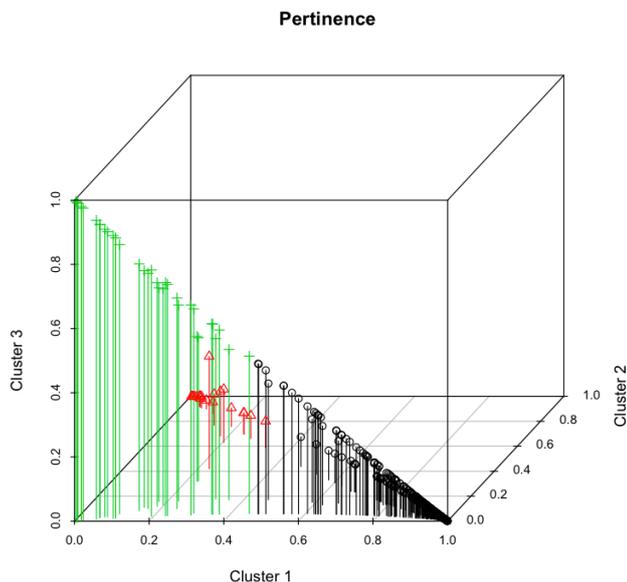

Figure 2. Pertinence of Clustering

Figure 2 shows Pertinence measurements after applying Fuzzy C-Means. As a result, cluster 2 is less likely to be transferred to another cluster, but some cluster 1 and cluster 3 are more likely to be transferred.

CONCLUSION

In this Paper, We used the Fuzzy C-Means clustering algorithm to analyze the debt ratio, current ratio and BIS ratio of the stability indices. As a result of clustering, we could see that it was divided into a certain section. Through this section, we were able to predict corporations with management problems. Based on this research, future studies will be conducted to forecast the management status of the company through the cumulative multi - year trend of the company.

REFERENCES

[1]   https://www.smacc.io/en/

[2]   Korea Employment Information Service, A Study on the Effects of Jobs on Technological Change, (2016)

[3]   Korea International Trade Association, Present Situation and Response Strategy of Korean Artificial Intelligence Startup, (2017)

[4]   https://dart.fss.or.kr/

[5]   Korea Information & Communication, Construction Association, 2016 KICA Business Analysis, (2017)

[6]   Klir, George, and Bo Yuan., Fuzzy sets and fuzzy logic. Vol. 4. New Jersey: Prentice hall, (1995).

[7]   Zadeh, Lotfi A., Fuzzy sets , Information and control 8.3 (1965), 338-353p.

[8]   Bezdek, James C., Robert Ehrlich, and William Full., FCM: The fuzzy c-means clustering algorithm, Computers & Geosciences 10.2-3 (1984), 191-203p.

[9]   Hartigan, John A., and Manchek A. Wong., Algorithm AS 136: A k-means clustering algorithm, Journal of the Royal Statistical Society. Series C (Applied Statistics) 28.1 (1979), 100-108p.

[10]   Ester, Martin, et al., A density-based algorithm for discovering clusters in large spatial databases with noise, Kdd. Vol. 96. No. 34. (1996).

[11]   Team, R. Core. "R language definition." Vienna, Austria: R   foundation   for   statistical computing (2000).