



Capturing Corporate Attributes in a New Perspective Through Fuzzy Clustering

Yusuke Matsumoto^(✉), Aiko Suge, and Hiroshi Takahashi

Graduate School of Business Administration, Keio University,
Hiyoshi 4-1-1, Kohoku-ku, Yokohama-shi, Kanagawa-ken, Japan
{yusukeM-aichi, aikosuge}@keio.jp,
htaka@kbs.keio.ac.jp

Abstract. Although industrial classification plays an important role in various contexts, it is rarely questioned. However, as diversification and business transformation are ongoing, it is becoming difficult to recognize company's real business. Therefore, there is not enough to allocate one type of business class to express the situation of the company, and a new type of industrial classification system is required. Through the analysis, we construct a new industrial classification system with Fuzzy C Means (FCM). This study also confirms the validity of proposed method through composite variance and absolute prediction error (APE). As the result, we present that there is a possibility that we are able to represent one company with overlapping industry, so to speak, assign one company more than two industries.

Keywords: Fuzzy C Means · Composite variance · Absolute prediction error · Industrial classification

1 Background

Industrial classification is utilized not only in practical situations but also in various research fields. For example, industrial classification is used in areas such as benchmarks for company selection criteria and financial conditions. Indicators used as criteria are set in each country. For example, in Japan, we often use the Japan Standard Industrial Classification (20 in the large classification, 99 in the middle classification) established by the Ministry of Internal Affairs and Communications (MIC), the Nikkei classification (36 in the middle classification, 256 class in the small classification) and TOPIX Sector Indices (10 in the large classification, 33 in the middle classification). In addition, globally, we use GICS (Global Industry Classification Standard) as international indicators established by Standard & Poor's and Morgan Stanley Capital International.

Traditionally, the existing industrial classifications constructed by various institutions give only one classification to one company, and do not permit overlapping industries. Moreover, the existing industrial classification is defined based on sales alone. However, in recent years, the business domain of companies has undergone major changes due to the increase of M&A and aggressive business transformation. Therefore, industrial identification is becoming more ambiguous than before. For example, in a diversified company, it is difficult to say that classified industrial identification accurately

reflects the enterprise when classified business is lower in sales or profits than other businesses. With these backgrounds, there is a limit to describing one company only one industry and representing the real situation of a company. Therefore, a new industrial classification is required. This paper examines whether such a new industrial classification technique can be utilized as an alternative to traditional industrial classification.

There are at least two merits to assign one company several industries. Firstly, it is easier to compare segments with other companies. How to divide the business segments shown in the securities report has been left to the discretion of each company. Therefore, when we compare the segments with other companies, we think that it is possible to compare segments among several companies on a unified scale. Secondly, there is a possibility that we can conduct empirical analysis in more detailed. Now, the existing industrial classification systems assign one type of industry to one company. However, looking at the breakdown of one company, there is a limit to using sales as the standard, as in the case of high market share even though sales are low. We think it is possible to take into account such influences and to enable more robust control of industry characteristics.

The composition of this paper is as follows: In Sect. 2, we detail the previous research. In Sect. 3, we describe data employed in this analysis. In Sect. 4, we show our method. In Sect. 5, we show our results. In Sect. 6, we select to analyze for manufacturing industry in Japan Standard Industrial Classification and show our results. Finally, we summarize this paper and describe the issues and discussion of this paper in Sect. 7. Again, the purpose of this thesis is to explore the possibility of a new industrial classification technique.

2 Related Work

There are very few research papers on industrial classification. Regarding these studies, there are mainly two areas; the reliability of the existing system and the construction of new system.

In the United States, there are several studies about the reliability of industrial classification. For example, Elton and Gruber [1] stated that there is no guarantee that existing industrial classification form homogeneous groups, and the industrial classification method constructed using statistical methods may be more accurate. Hrazdil, Trottier and Zhang [2] analyzed which index, such as Standard Industrial Classification (SIC), the North American Industry System (NAICS) or the Global Industry Classification Standard (GICS) are effective and homogeneous for the group stocks with similar operating characteristics. Weiner [3] stated that as a result of performing cluster analysis based on the company's financial information, the industrial classification constructed by himself was more accurate than the existing industrial classification. On the other hand, in Japan, there are very few studies about the reliability of industrial classification. Studies in Japan about the reliability of industrial classification have only been done by Kimura [4], Shintani [5], and Nakaoka [6].

Various methods have been proposed to address construction of a new industrial classification. For example, Sasaki and Shinno [7] and Ando and Shirai [8] have proposed a method of acquiring industry type information from web pages and classifying

companies based on that information. Isogai and Dam [9] proposed methods of classifying companies based on stock price fluctuations. Peneder [10] classified the companies through the statistical cluster analysis and presented that the use of cluster analysis provided valuable tools for the industrial classification. Lee, Ma, and Wang [11] focused on “co-search” on the Internet and proposed a new method to recognize economically related peer companies.

Lewellen [12] builds an industrial classification that allows overlapping based on which companies the competitor are. There are Yang [13] and Budayan, Dikmen and Birgonul [14] as studies that classified companies using the clustering method. Yang [11] has built an industrial classification system that allowed overlapping by conducting latent class analysis for companies in Japan and Malaysia. Budayan, Dikmen and Birgonul [14] classified construction companies in Turkish by using clustering methods; K-means, self-organizing map and Fuzzy C Means. They set the number of clusters to 3 and they drew conclusions after they experimentally compare the three methods. As the result, they reported that the clustering methods such as self-organizing maps or Fuzzy C Means have a possibility to provide more valuable results.

However, in these papers, there can be pointed as three limitations. First, comparison with the existing industrial classification have not been conducted thoroughly. Second, the verification of validity of results have not been conducted thoroughly. Third, analysis using stock fluctuation depended on the heavy econometric model. Moreover, comparison with the existing industrial classification and verification of validity of results have not been conducted thoroughly. It remains to be seen whether cluster analysis that allows for overlapping can lead to more accurate results and whether reliability can be guaranteed.

The purpose of this paper is to construct a new industrial classification system that allows overlapping, and compare it with the existing industrial classification.

3 Data

We used financial data in 2016, for companies listed on the First Section of the Tokyo Stock Exchange. The number of companies analyzed is 1210. Data was obtained from Nikkei NEEDS. In this paper, we use the companies belonging to 17 industries among the large industrial classifications of the MIC. We exclude (1) companies that belong to compound service, (2) companies belonging to government except elsewhere classified, and (3) companies belonging to industries unable to classify. The reason why we exclude the three industries is that there no companies which belong to those three industries. In addition to this, we utilize the companies belonging to the manufacture industry, which is one of the 17 industries. These companies belong to the 23 industries among the middle industrial classification of the MIC. We excluded companies which belong to the manufacture of leather tanning, leather products and fur skins. The number of companies is 694.

The variables and their definitions used for performing overlapping cluster analysis and evaluating validity are in shown in Table 1 below.

Table 1. Descriptive statistics

| Variable name (Variable definition) | Mean | Median | Max | Min | Number of Corp |
|---|-------|--------|-------|--------|----------------|
| Operating margin (Operating income ÷ Sales) | 0.070 | 0.063 | 0.534 | -0.376 | 1210 |
| Capital adequacy ratio (Netassets ÷ total assets) | 0.567 | 0.569 | 0.946 | 0.124 | 1210 |
| Total asset turnover rats (Sales ÷ total assets) | 0.904 | 0.848 | 2.705 | 0.264 | 1210 |
| Sales growth rate (Change rate from the previous term) | 0.020 | 0.015 | 0.992 | -0.353 | 1210 |

We compared our new industrial classification technique by overlapping cluster analysis with the existing industrial classification. At that time, we used the Japan Standard Industry Classification (JSIC) as target industries. We obtained the data of the JSIC from Nikkei NEEDS.

4 Method

4.1 Cluster Analysis

In this paper, we adopted Fuzzy C Means (FCM) proposed by Bezdek [15] as in Eq. (1), as a way of classifying some companies. Normal cluster analyses demands that all data belongs to only one cluster. However, in FCM, by introducing a fuzzy set, it is possible to allow learning vectors to belong to two or more clusters. The centered algorithm is as follows:

$$J = \sum_{i=1}^N \sum_{k=1}^K (g_{ik})^m \|x_i - c_k\|^2 \quad (1)$$

In this paper, we set the initial value for cluster centers (K) to 17, which is larger than Budayan, Dikmen and Birgonul [14]. This number is the same as the MIC, except for three industries (financial or insurance industry, the complex service business, public service industry, and industry not classifiable). The distance ($\|x_i - c_k\|$) adopts the Euclidean distance. The degree of fuzzification (m) is 2. For the data (x_i), we selected four data; (1) operating margin which represents profitability, (2) the capital adequacy ratio which represents safety, (3) the total asset turnover which represents activity, and (4) the growth rate which represents growth. Definitions of these variables and descriptive statistics of the entire sample are shown in Table 1. As the result of the analysis, each company has a membership value (g_{ik}) for each of the 17 clusters¹. We rearrange the clusters in descending order of membership value. We set the highest

¹ In this paper, FCM analysis was repeated 50 times. We used a representative one after confirming each results.

membership value as the 1st industry. So the 17th industry has the lowest membership value. We conducted standardization before analysis. (see Fig. 1).

| Company Name | Rank | Highest membership value | | | Lowest membership value | | |
|--------------|------|--------------------------|--------------|-----|-------------------------|---------------|--|
| | | 1st | 2nd | ... | ... | 17th | |
| Company 1 | | Cluster 1 | Cluster 2 | ... | ... | Cluster 10 | |
| Company 2 | | Cluster 3 | Cluster 6 | ... | ... | Cluster 2 | |
| Company 3 | | Cluster 2 | Cluster 10 | ... | ... | Cluster 6 | |
| ⋮ | | ⋮ | ⋮ | ... | ... | ⋮ | |
| Company 1210 | | Cluster 17 | Cluster 5 | ... | ... | Cluster 15 | |
| | | ↑ | ↑ | | | ↑ | |
| | | 1st Industry | 2nd Industry | | | 17th Industry | |

Fig. 1. Image after work

4.2 Verification Method

In Sects. 4.2 and 4.3, we propose two verification methods to confirm the validity of our proposed FCM classification. We compared the industrial classification newly created by FCM in the Sect. 4.1 with the existing industry classification. One is composite variance (Sect. 4.2), and the other is absolute prediction error (Sect. 4.3).

At first, we verified the reliability of industrial classification using the composite variance proposed in Amit and Livnat [16]. By using the composite variance, it is possible to compare the homogeneity of the group of companies.

The composite variance value (S_a) of the industry category (a) for a certain evaluation index (x) is calculated in Eq. (2).

$$S_a = \frac{\sum_{i=1}^{N_a} (n_{ai} - 1) V_{x_{ai}}}{\sum_{i=1}^{N_a} (n_{ai} - 1)} \quad (2)$$

$$\bar{X}_{ij} = \frac{1}{n_{ai}} \sum_{k=1}^{n_{ai}} X_{aik} \quad (3)$$

$$V_{x_{ai}} = \frac{1}{(n_{ai} - 1)} \sum_{k=1}^{n_{ai}} (X_{aik} - \bar{X}_{ai})^2 \quad (4)$$

Here, S_a is the composite variance value in the industry category (a), N_a is the number of business group in the industry category (a), $V_{x_{ai}}$ is the variance of the evaluation index x in the industry group i when classification of (a) is used, and n_{ai} is the number of firms in group (i) when classification (a) is used.

We calculated the ratio with the composite variance value S_b of another industrial classification (b). The dispersion ratio between S_a and S_b as shown in Eq. (5) follows the F distribution.

$$S(a, b) = \frac{S_a}{S_b} \quad (5)$$

If this ratio is statistically significantly different from 1 in the F test, we can judge that there is a difference in homogeneity in both industrial classifications. If the variance ratio is statistically significantly greater than 1, which means that the denominator S_b is statistically significantly smaller, the industry classification S_b is evaluated to be more reliable than the industry category S_a . We should remember that comparison by composite variance is not absolute but relative with respect to reliability.

In this paper, we evaluated by using the operating margin, the capital adequacy ratio, the total asset turnover, and the sales growth rate. These variables are used in FCM. We compare five of the first, second, third, fourth, and fifth industries of clusters created by FCM respectively with the JSIC. When we carried out the composite variance, we used values before standardization.

On calculating the composite variance value (S) of each evaluation index, it is greatly affected by outliers. Therefore, we calculated the composite variance value (S) by excluding outliers that are 1% above or below for each evaluation index.

4.3 Absolute Prediction Error

As the second verification method, we compared industrial classification using absolute prediction error (APE) proposed by Weiner [2]. Through the result of Sect. 4.1, we can use our industrial classification to select similar companies. We then compare the APE of these selected companies with the APE of companies selected under traditional classification. We use the APE of the enterprise value calculated using a multiple approach.

A multiple approach estimates the enterprise value of the firm by multiplying earnings with an enterprise value to EBIT multiple determined from a set of comparable companies.

The estimation for firm i 's enterprise value \widehat{EV}_i is given by

$$\widehat{EV}_i = \text{median}_{j \in C_i} \left(\frac{EV_j}{EBIT_j} \right) \times EBIT_i, \quad (6)$$

where C_i is the set of comparable firms based on FCM, EV_j is firm j 's enterprise value, and $EBIT_j$ is the firm j 's EBIT.

The valuation accuracy is calculated by the deviation between the estimated firm value and the real firm value. Therefore, we can calculate the APE_i for firm i as Eq. (7).

$$APE_i = \left| \frac{\widehat{EV}_i - EV_i}{EV_i} \right| \quad (7)$$

\widehat{EV}_i is the estimated enterprise value for firm and EV_i is the observed market value for firm i . We statistically test the results by performing a Wilcoxon rank sum test on the differences between our new industrial classification system and the existing method of industrial classification. If APE_i of our industrial classification was statistically significantly smaller than that of the existing industrial classification, it would indicate that the selection of similar companies works well and is responsive to a multiple approach. This result suggests that our industrial classification shows more homogeneous.

5 Result

5.1 Analysis of Fuzzy C Means

Figure 2 displays the results of cluster analysis, plotting operating profit ratio on the horizontal axis and capital adequacy ratio on the vertical axis. Since we show the result with two axes despite analyzing with four variables, it may be difficult to understand the result. However, it turns out that cluster number 8, 13, and 14 are clearly classified.

Table 2 displays the industry to which each company belongs, with regard to using the first industry named in the FCM. We list three companies in each industry, in descending order of market capitalization in FY2016. We can observe that industry number 8 is a group of companies famous for their high capital adequacy ratio and industry number 17 is a high operating margin group. As in the above example, we specify the characteristics of some groups.

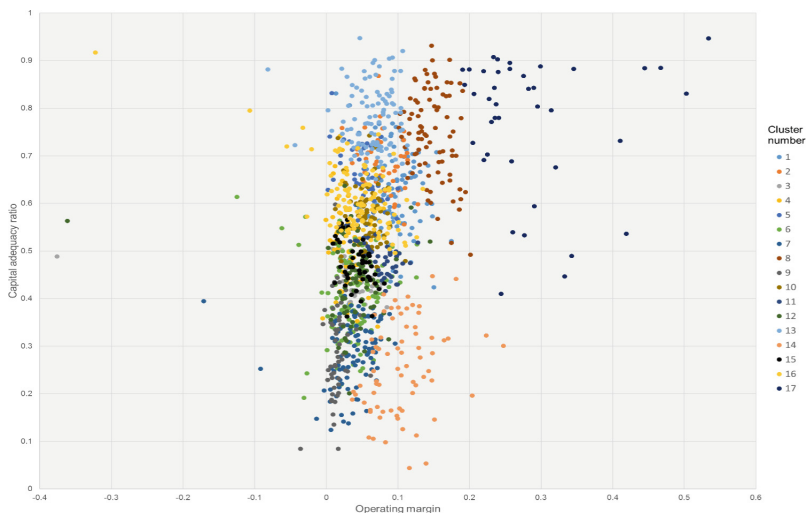


Fig. 2. Clustering result of Fuzzy C Means

Table 2. Clustering result of Fuzzy C Means

| Cluster number | Corporate name | Cluster number | Corporate name |
|----------------|---|----------------|--|
| 1 | Nippon Telegraph and Telephone Corp Subaru Corp Nidec Corp | 10 | AISIN SEIKI Co., Ltd. Sumitomo Electric Industries, Ltd. TOTO LTD. |
| 2 | DENSO Corp Eisai Co., Ltd. Uncharm Corp | 11 | Mitsubishi Electric Corp DAIKIN INDUSTRIES, Ltd. Kao Corp |
| 3 | Asahi Group HD NTT DATA Corp Shiseido Company | 12 | Dentsu Inc. Fujitsu Limited MAZDA Motor Corp |
| 4 | Daito Trust Construction Co., Ltd. NH Foods Ltd. Sundrug Co., Ltd. | 13 | Nintendo Co., Ltd. FUJIFILM Corp KYOCERA Corp |
| 5 | OMRON Corp Sekisui Chemical Company, Limited Hitachi High-Technologies Corp | 14 | Toyota Motor Corp SoftBank Corp Sony Corp |
| 6 | Mitsubishi Corp MITSUI & CO., LTD. ITOCHU Corp | 15 | SUZUKI MOTOR CORP YAMATO HOLDINGS CO., LTD. Nippon Express Co., Ltd. |
| 7 | Honda Motor Co., Ltd. Nissan Motor Co., Ltd. Hitachi, Ltd. | 16 | Canon Inc. Takeda Pharmaceutical Company Limited Bridgestone Corp |
| 8 | NTTDOCOMO, INC. KDDI CORP Shin-Etsu Chemical Co., Ltd. | 17 | JAPAN TOBACCO INC. KEYENCE CORP FANUC Corp |
| 9 | Marubeni Corp TOYOTA TSUSHO CORP TOYOTA BOSHOKU CORP | | |

(Note 1) In this paper, we treat the cluster number as industry type.

(Note 2) Three companies are listed in descending order of market capitalization in each industry.

5.2 Composite Variance Analysis

The results of comparing the reliability by the composite variance were as shown in Table 3. As a result of the analysis, all variables from the first industry S_{M1} to the third industry S_{M3} are statistically significant². From Table 3, we may observe that the denominator of the classification of the newly constructed industrial classification is statistically significantly less than 1 in all four financial indicators from the first industry to the third industry. Therefore, it indicates that our new industrial classification method forms a more homogeneous group than the JSIC. However, value of the fourth industry S_{M4} and the fifth industry S_{M5} are smaller than 1 at the total asset

² $S_{M1}, S_{M2}, S_{M3}, S_{M4}$ and S_{M5} represent first industry, second industry, third industry, fourth industry and fifth industry respectively, which are newly obtained by the FCM.

turnover and the sales growth rate. It means that the composite variance at the denominator is greater than that at the numerator. In other words, homogeneity is lower than the existing industrial classification with regard to total asset turnover rate and sales growth rate.

In addition to these observations, looking at the values of the composite variance ratios in Table 3, the dispersion ratio decreases from the first industry to the fifth industry. This means that the variance of the industrial classification created by FCM, which is the denominator of the composite variance ratio, is large. Less homogeneity is observed as the number of industrial classification increases.

Table 3. The result of composite variance

| | |
|---------------------------|----------------------|
| First industry | $\frac{S_N}{S_{M1}}$ |
| Operating margin | 2.518*** |
| Capital adequacy ratio | 4.771*** |
| Total asset turnover rate | 2.733*** |
| Sales growth rate | 1.570*** |
| Second industry | $\frac{S_N}{S_{M2}}$ |
| Operating margin | 1.351*** |
| Capital adequacy ratio | 3.074*** |
| Total asset turnover rate | 1.927*** |
| Sales growth rate | 1.220*** |
| Third industry | $\frac{S_N}{S_{M3}}$ |
| Operating margin | 1.181*** |
| Capital adequacy ratio | 2.388*** |
| Total asset turnover rate | 1.303*** |
| Sales growth rate | 1.090* |
| Fourth industry | $\frac{S_N}{S_{M4}}$ |
| Operating margin | 1.110** |
| Capital adequacy ratio | 2.132*** |
| Total asset turnover rate | 0.911 |
| Sales growth rate | 0.857 |
| Fifth industry | $\frac{S_N}{S_{M5}}$ |
| Operating margin | 1.095** |
| Capital adequacy ratio | 1.570*** |
| Total asset turnover rate | 0.911 |
| Sales growth rate | 0.857 |

(Note 1) The value indicates the composite variance value. The numerator S_N indicates the JSIC. Additionally, the denominator S_{M1} , S_{M2} , S_{M3} , S_{M4} and S_{M5} represent the first industry, second industry, third industry, fourth industry and fifth industry respectively, which are newly constructed by the FCM.

(Note 2) Statistical significance is denoted with ***, ** and * for 1%, 5% and 10% rejection levels.

5.3 Absolute Prediction Error Analysis

We performed Wilcoxon rank sum test. Table 4 displays the result of APE. The number in the APE row shows the APE by the newly constructed industrial classification system and the APE by the JSIC. The number in the difference row means the difference between APE in each industrial classification system. A negative value indicates that our new industrial classification system has higher accuracy of corporate valuation than the JSIC. From the results, we found that the difference between the APE of our new industrial classification system and the APE of the JSIC was statistically significantly negative at the 10% level for the first industry and second industry. The APE of the first industry and second industry is smaller than that of the JSIC. Therefore, we perceived that the first industry and the second industry which we newly constructed by FCM has higher homogeneousness than industries based on the JSIC. However, it is not statistically significant after the third industry. From Table 4, we see that the *APE* of the newly constructed industrial classification increases as the number of classifications increases. In other words, this shows that homogeneity is lost as we proceed down the industry ranking.

Table 4. Result of absolute prediction error

| | | |
|-----------------|-------|---------|
| First industry | M1 | N |
| APE | 0.279 | 0.311 |
| Difference | | -0.032* |
| Second industry | M2 | N |
| APE | 0.295 | 0.311 |
| Difference | | -0.016* |
| Third industry | M3 | N |
| APE | 0.306 | 0.311 |
| Difference | | -0.005 |
| Fourth industry | M4 | N |
| APE | 0.308 | 0.311 |
| Difference | | -0.003 |
| Fifth industry | M5 | N |
| APE | 0.326 | 0.311 |
| Difference | | 0.015 |

(Note 1) The value in the row of *APE* are the median of the absolute prediction error of the first industry, the second industry, third industry, fourth industry and fifth industry of the newly constructed industrial classification and the median of the absolute prediction error of the JSIC.

(Note 2) Statistical significance is denoted with ***, ** and * for 1%, 5% and 10% rejection levels.

6 Manufacturing Industry

In Sect. 5, we constructed our new industrial classification and verified its validity for all industries. Next, in this chapter, we focused on the manufacturing industry in large classification of the JSIC, classified them through our FCM method, and compared the result with the existing industrial classification. In FCM, we set 23 to the number of clusters, since this number is the same number of manufacturing industry. The other settings for the parameters are the same as in Sect. 4.1. We also apply the same verification methods; composite variance method (Sect. 4.2) and absolute prediction error (Sect. 4.3).

6.1 Absolute Prediction Error Analysis

Table 5 shows the result of composite variance for manufacturing industry. From Table 5, we found that the denominator of the classification of the newly constructed industrial classification system is significantly less than 1 in all four financial indicators from the first industry to the third industry constructed by FCM. Therefore, it suggests us that the newly constructed industrial classification system provides more homogeneous groupings of firms than the existing industrial classification. However, according to sales growth rate of the fourth industry, total asset turnover rate and sales growth rate of the fifth industry, the value of them are less statistically significantly larger than 1. It means that the composite variance at the denominator is equal to the numerator. As a result, our newly constructed method is comparable to those existing classification methods.

Table 5. The result of composite variance

| | |
|---------------------------|-----------------------|
| First industry | $\frac{S_N}{S_{MM1}}$ |
| Operating margin | 3.476*** |
| Capital adequacy ratio | 6.111*** |
| Total asset turnover rate | 3.404*** |
| Sales growth rate | 1.917*** |
| Second industry | $\frac{S_N}{S_{MM2}}$ |
| Operating margin | 1.934*** |
| Capital adequacy ratio | 3.966*** |
| Total asset turnover rate | 1.987*** |
| Sales growth rate | 1.215*** |
| Third industry | $\frac{S_N}{S_{MM3}}$ |
| Operating margin | 2.029*** |
| Capital adequacy ratio | 3.045*** |
| Total asset turnover rate | 1.649*** |
| Sales growth rate | 1.115* |

(continued)

Table 5. (continued)

| | |
|---------------------------|-----------------------|
| Fourth industry | $\frac{S_N}{S_{MM4}}$ |
| Fourth industry | $\frac{S_N}{S_{MM4}}$ |
| Operating margin | 1.257*** |
| Capital adequacy ratio | 2.732*** |
| Total asset turnover rate | 1.374*** |
| Sales growth rate | 1.091 |
| Fifth industry | $\frac{S_N}{S_{MM5}}$ |
| Operating margin | 1.191** |
| Capital adequacy ratio | 2.112*** |
| Total asset turnover rate | 1.034 |
| Sales growth rate | 1.038 |

(Note 1) The value indicates the composite variance value. The numerator S_N indicates the JSIC. Additionally, the denominator S_{MM1} , S_{MM2} , S_{MM3} , S_{MM4} and S_{MM5} represent the first industry, second industry, third industry, fourth industry and fifth industry respectively, which are newly constructed for manufacturing industry by the FCM.

(Note 2) Statistical significance is denoted with ***, ** and * for 1%, 5% and 10% rejection levels.

6.2 Absolute Prediction Error Analysis

We performed Wilcoxon rank sum test. Table 6 shows the result of APE for manufacturing industry. The number in the APE row shows the APE by the newly constructed industrial classification system and the APE by the JSIC. The number in the difference row means the difference between APE in each industrial classification systems. A negative value indicates that our new industrial classification system has higher accuracy of corporate valuation than the JSIC. From the result, we found that the difference between MMI and N is statistically significantly negative in the first industry. The APE of the first industry is smaller than that of the JSIC. Therefore, we perceived that our newly industrial classification system is gathering similar companies than JSIC's companies in the first industry. However, from the result of second industry to that of fifth industry, our classification system is statistically significantly positive. We confirmed that the JSIC has higher similarity than our classification system from the second to fifth industries.

Compared to the result of Sect. 5.3, the number of industries constructed by FCM, which are more homogeneous than the JSIC, has decreased from 2 to 1. However, looking at the value of APE in the first industry our newly constructed, the APE of this result is smaller than the APE of Sect. 5.3. This result suggests us that the exacter we classify the companies, the smaller the error between the actual enterprise value and the enterprise value based on multiple approach would be.

Table 6. Result of Absolute Prediction Error for manufacturing industry

| | | |
|-----------------|-------|-----------|
| First industry | MM1 | N |
| APE | 0.257 | 0.281 |
| Difference | | -0.024*** |
| Second industry | MM2 | N |
| APE | 0.296 | 0.281 |
| Difference | | 0.016*** |
| Third industry | MM3 | N |
| APE | 0.291 | 0.281 |
| Difference | | 0.010*** |
| Fourth industry | MM4 | N |
| APE | 0.312 | 0.281 |
| Difference | | 0.032*** |
| Fifth industry | MM5 | N |
| APE | 0.306 | 0.281 |
| Difference | | 0.025*** |

(Note 1) The value in the row of *APE* are the median of the absolute prediction error of the first industry, the second industry, third industry, fourth industry and fifth industry of the newly constructed industrial classification for manufacturing industry and the median of the absolute prediction error of the JSIC.

(Note 2) Statistical significance is denoted with ***, ** and * for 1%, 5% and 10% rejection levels.

7 Conclusion and Further Discussion

In this paper, we analyzed whether it is possible to construct a new industrial classification system by using FCM. The result of composite variance and APE for all industries in the JSIC indicated that FCM is effective in making homogeneous industrial clusters. The result of analysis for the manufacturing industry in the JSIC shows us that there is a possibility of classifying companies more homogeneously by dividing them exactly. FCM is effective in constructing the new industrial classification system. Moreover, Nikkei NEEDS shows up to three industries for each company. Our three industries made by FCM are consistent with Nikkei NEEDS through composite variance analysis. As a conclusion, FCM is one of the tools to assign multiple industry to one company. Through the industrial classification with our new proposed FCM methods, it may be possible to express the company's situation and it may be useful for us to correctly perceived company's reality.

Finally, we would like to point out three limitations in this paper. Firstly, it is necessary to conduct further analysis that might enable our FCM to be more effective in wide range of clusters. Since this paper followed the methods used in previous research, the validation method is not necessarily suitable for FCM. Secondly, with regard to the cluster analysis, it is not always possible to obtain the same result each time when classifying using FCM. Therefore, it is necessary to verify the robustness of classification result by repeating it multiple times. Finally, in this paper, we classified all companies into 17 groups based on financial data. However, it is difficult to find financial features individually for each cluster, based only on the four types of financial indicators. We plan to apply other indicators to construct our classification method for a deeper understanding of economic features of each cluster.

Acknowledgements. This research was supported by a grant-in-aid from the Kayamori Foundation of Informational Science Advancement.

References

1. Elton, E.J., Gruber, M.J.: Homogeneous groups and the testing of economic hypotheses. *J. Financ. Quant. Anal.* **4**(5), 581–602 (1970)
2. Hrazdil, K., Trottier, K., Zhang, R.: An intra-and inter-industry evaluation of three classification schemes common in capital market research. *Appl. Econ.* **46**(17), 2021–2033 (2014)
3. Weiner, C.: The impact of industry classification schemes on financial research. SFB 649 discussion paper, No. 2005,062, SFB 649, Economic Risk, Berlin (2005)
4. Kimura, F.: A comparison of reliability of industrial classification in Japan. *Contemp. Discl. Res.* **9**, 3–42 (2009)
5. Shintani, O.: Sector classification systems: comparison of reliability between the TSE 33 (stock price index by industry <33 sectors>) and Global Industry Classification Standard (GICS) taxonomies. *Secur. Anal. J.* **48**(4), 77–88 (2010)
6. Nakaoka, T.: Statistical analysis on industrial classification: a comprehensive survey and a suggestion for new methods of composing homogeneous groups. *J. Bus. Econ. Shokeigakuso* **61**(1), 151–180 (2014)
7. Sasaki, M., Shinno H.: Industrial classification from corporate website by using document classification method. In: *The Association for Natural Language Processing 12th Annual Convention Proceedings*, pp. 352–355 (2006)
8. Ando, K., Shirai, K.: Extracting and classification of industry information from the corporate web pages. In: *The Association for Natural Language Processing 24th Annual Convention Proceedings*, pp. 1015–1018 (2018)
9. Isogai, T., Dam, H.: Building classification trees on Japanese stock groups partitioned by network clustering. *IEEJ Trans. Electron. Inf. Syst.* **137**(10), 1387–1392 (2017)
10. Peneder, M.: Creating industry classifications by statistical cluster analysis. *Estudios de economía aplicada* **23**(2), 451–464 (2005)
11. Lee, C.M., Ma, P., Wang, C.C.: Search-based peer firms: aggregating investor perceptions through internet co-searches. *J. Financ. Econ.* **116**(2), 410–431 (2015)
12. Lewellen S.: Firm-specific industries. Working paper (2012)
13. Sook, L.Y.: A latent class cluster analysis study of financial ratios and industry characteristics. *Aust. J. Basic Appl. Sci.* **7**(11), 46–53 (2013)

14. Budayan, C., Dikmen, I., Birgonul, M.T.: Comparing the performance of traditional cluster analysis, self-organizing maps and fuzzy C-means method for strategic grouping. *Expert Syst. Appl.* **36**(9), 11772–11781 (2009)
15. Bezdek, C., Robert, E., William, F.: FCM: the fuzzy c-means clustering algorithm. *Comput. Geosci.* **10**(2–3), 91–203 (1984)
16. Amit, P., Livnat, J.: Grouping of conglomerates by their segments' economic attributes: towards a more meaningful ratio analysis. *J. Bus. Finance Account.* **17**, 85–99 (1990)