Resilience Evaluation of Automakers after 2008 Financial Crisis by UMAP

Seiji Matsuhashi and Yukari Shirota*

***Abstract*—Under various risks, companies are required to be resilient, so that they can flexibly respond to changes. In this paper, we evaluate the resilience of the automobile manufacturing industry after the 2008 financial crisis, using machine learning methods. The approach consists of time series data clustering named "Amplitude-based clustering" and dimensional reduction method named "UMAP". In the analysis, companies' indexed market capitalization data are used. In order to investigate the recovery power of companies, the growth rate is important and the amplitude-based clustering is required. In the UMAP result, we found that one of the principal components can be used as the resilience measurement. Our approach in this paper is widely applicable to measure industries' resilience after the drastic decline of stock prices.**

***Index Terms*—Time series data clustering Amplitude-based clustering, dimensionality reduction, UMAP, market capitalization, automakers, stock prices, resilience, 2008 financial crisis**

I. Introduction

There are various environmental changes that affect global supply chains, such as the spread of COVID-19 infection, geopolitical risks, trade friction between major powers, and frequently occurred natural disasters. Under such risks, companies are required to be resilient, so that they can flexibly respond to changes in order to ensure stable supply continuity and business continuity.

In this paper, we evaluate the **resilience of the automobile manufacturing industry after the 2008 financial crisis**, using machine learning methods. The movement of market capitalization data of 135 companies are indexed, and the fluctuation patterns are extracted by time series data clustering. In this clustering, amplitude/variance information such as how many times the market capitalization after one year, is important. For example, even if the market capital price recovers, it is significantly different which is 1.1 times or 3.6 times. Therefore, clustering methods that require data standardization in advance cannot be used [1]. For example, because k-shape with SBD [2–4] and DTW [5, 6] methods require the data standardization, the methods are not suitable. Therefore, we used an amplitude-based clustering method developed by our team as the clustering method [1, 7].

Next, to extract the characteristics of the recovery patterns, the dimensional reduction method "UMAP" (Uniform Manifold Approximation and Projection) was conducted, using the amplitude-based distances among the companies. Before the UMAP development, t-SNE (t-Distributed Stochastic Neighbor Embedding) [8–11] was the most widely used dimensional reduction tool. The UMAP algorithm is competitive with t-SNE for visualization quality, and arguably preserves more of the global structure with superior run time performance and can model the manifold with a fuzzy topological structure and can search for a low dimensional projection of the data that has the closest possible equivalent fuzzy topological structure [12].

In this paper, using the UMAP, the 135-dimensional space was reduced to 2-dimensional data. As a result, we finally found that the resultant first principal component can be regarded as a measure of **resilience strength** of the individual industry. Since the financial crisis had globally damaged automakers, we had expected whether the resultant principal component would be (1) the magnitude of the decline or (2) the strength of resilience. As we expected, the above-mentioned "strength of resilience" measurement was extracted as the first principal component.

The combined approach by clustering and dimensional reduction proposed in this paper is widely applicable to evaluate the industry resilience strength at global market capital declines.

Section II describes the market capitalization data used for this analysis. In Section III, we describe the algorithm of the Amplitude-based method and the resultant clusters of this case study. In Section IV, the UMAP result is shown as a two-dimensional scatterplot and in Section V interpretation of the extracted principal component axis is described; the principal component measures the strength of industry resilience. The last section is a summary section.

II. Market Capital Data

In the section, we shall explain the data we used and effects of the 2008 financial crisis that occurred during this period 2007 to 2011. In addition, our research standpoint will be described, in the researches in AI in finance.

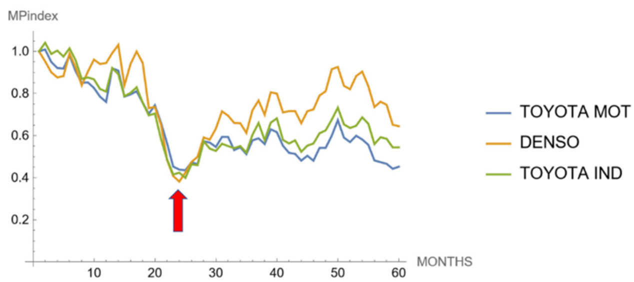

Fig. 1. Indexed market capitalization movement.

A. Market Capitalization Data

The data we used are the monthly automakers' market capital data. The market capital is found by multiplying the current stock price with the number of shares. The larger the

Manuscript received August 31, 2022, revised September 10, 2022; accepted January 15, 2023.

Both authors are with the Faculty of Economics, Gakushuin University, Tokyo, Japan.

*Correspondence: yukari.shirota@gakushuin.ac.jp (Y.S.)

doi: 10.18178/ijml.2023.13.3.1139

market capital, the higher the value of the company as a whole. Therefore, market capital is an important factor to measure the company size and status.

The data period we investigated are **5 years from Jan. 2007 to Dec. 2011.** The data consists of the global 135 top automobile manufacturers in 2007; then, TESLA was not yet grown and not included in the data. Only the companies which have all data through the period are selected and used in this analysis. The data was retrieved from the web-based data base ORBIS by Bureau Van Dijk. The programs for the clustering and data visualization were written in Mathematica by Wolfram, version 13. UMAP library with Python is umap-learn 0.5.3 (https://pypi.org/project/umap-learn/#description) [13].

Because the individual company's data scales are different, the data are indexed with the initial value as 1 (see Fig. 1). In Fig. 1, the movement patterns of TOYOTA MORTORS and its alliance companies DENSO and TOYOTA INDUSTORIES are illustrated. The three patterns of such group companies have high similarity on the patterns. Japanese representative automobile industry is likely to construct a pyramid structure. There are on the top vehicle manufacturers such as TOYOTA, parts manufacturers under them, small manufacturers that provide parts for the parts, and subcontractor manufacturers at the bottom [14]. The sales status of the top manufacturers will greatly influence the whole affiliates.

In Fig. 1, the lowest point is at Oct. 2008 owing to Lehman shock (see the red arrow point). The 2008 financial crisis triggered by the subprime mortgage problem caused great damages on automakers only in the United States but also in Japan, resulting in a large decline in stock prices and the market capitalization. In addition, in Japan the soaring yen problem seriously affected Japanese automakers. The global automakers, reducing inventory, cancelled expansion plans, closed the factories, reduced the number of employees, and conducted the production system restructure, and many automobile manufacturers became involved this global industry restructuring, although some companies went bankrupt [14].

B. AI Researches in Finance

Many researches on AI in finance have been conducted. Over the next few decades, machine learning and data science will transform the finance industry [15]. Many concrete programming books for AI in finance have already been published [16]-[20]. The main theme of the field is stock price prediction.

There are two approaches to market forecasting; they are technical and fundamental forecasting. Murphy said "While technical analysis concentrates on the study of market action which has to do with the study of human psychology, the fundamental analysis focuses on the economic forces of supply and demand that cause process to move higher, lower, or stay the same." [21]. Our interests are the fundamentals and not forecasting the stock prices. Our research theme is to find the intrinsic value of the company with soaring stock prices.

For example, suppose that company X has a higher supply chain management skill. That causes a higher profitability, and then finally the stock prices will increase. Therefore, we could say that the intrinsic strength of the company X is its supply chain management skills.

In our studies, although our aim is not the market prediction, we utilize both AI techniques of the technical analysis and those of the fundamental analysis. The time series data clustering algorithms as described in this paper are techniques used in the technical analysis. In another paper, we use regression and its Shapley value-based approach to find the important variables to increase the target values [22]-[24]. These regression analysis techniques are related to both, but they are more related to fundamental analysis techniques. Like the two wheels of a car, using both clustering and regression, we are conducting an integrated analysis. In the next step, we would like to conduct the regression analysis, using the clustering result in this paper

III. Amplitude-Based Clustering

In the section, first we shall simply explain the Amplitude-based clustering. Then, the result of the clustering in this case analysis will be illustrated by a dendrogram, a heatmap of the distance, and market capital movement graphs.

A. Amplitude-Based Clustering Method

Here our previously developed Amplitude-based clustering method is explained in short. The detailed are described in [1], [7]. For the distance definition, we adopted the Euclidean distance which is given below.

$$ED_{i,j} = \sqrt{\sum_{k=1}^{T} \left(\mathrm{G}_{\mathrm{i,k}} - \mathrm{G}_{\mathrm{j,k}}\right)^2}$$

where $\boldsymbol{G}_{\boldsymbol{j},\mathrm{k}}$ is the index data of j-th company on k-th day and T is the number of sales days. Then the distance-distance is defined as the Euclidean distance as follows:

$$\widetilde{\mathbf{ED}}_{\mathbf{i,j}} = \sqrt{\sum_{\boldsymbol{n}=\mathbf{1}}^{\boldsymbol{N}} \left(\boldsymbol{ED}_{\boldsymbol{n,i}} - \boldsymbol{ED}_{\boldsymbol{n,j}}\right)^2}$$

where N is the number of companies.

Then a hierarchical clustering [4, 25–27] is conducted with input as the distance of distance matrix $\{\widetilde{\mathbf{ED}}_{\mathbf{i,j}}\}$. Then, the matrix serialization or the quasi-diagonalization [28, 29] on the distance matrix is executed, so that the similar cells can be laid out as much as diagonally.

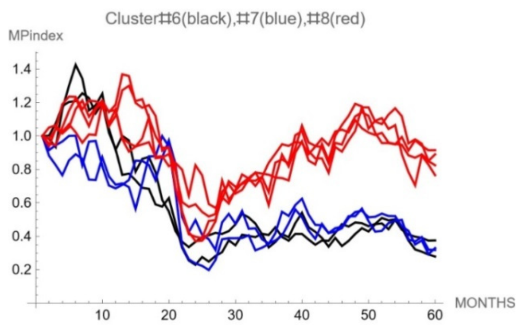

Fig. 2. Three clusters of indexed market capitalization movement.

The aim of Amplitude-based clustering is to make similarity pattern groups, keeping the variance/amplitude information. Fig. 2 illustrates a sample result of the Amplitude-based clustering. There are three clusters colored in red, black, and blue. Cluster#8 in red is significantly different from other two clusters; after the 28th month companies in Cluster#8 increased the market cap., compared to the others. When we compare Cluster#6 in black with Cluster#7 in blue, we can find that Cluster#6 has a large hike/decline in the initial period from 1 to 15 month.

In an economics analysis such as a GDP growth rate analysis, this amplitude-based clustering is required. Therefore, we have developed this Amplitude-based clustering algorithm. Shirota et al. analyzed Indian IT service companies' rapid growth by using Amplitude-based clustering [13].

As the existing time series data clustering methods, there are k-shape and Dynamic Time Warping (hereafter DTW) methods. However, they require the data standardization, so the data scale is arranged, in advance. The standardization deletes the variance/amplitude information. In Fig. 3, the standardized patterns of the data in Fig. 2 are illustrated. Because the variance is arranged to be 1 and the average is 0 after the standardization, we cannot observe the amplitude of the growth. A small fluctuation will become a large wave movement by standardization. As shown in Fig. 3, the recovery movement of Cluster#8 in red was removed and we cannot identify the superiority of Cluster#8 there.

This variance information missing problem was firstly discussed by Matsuhashi [30]. In conclusion, k-shape method cannot meet requirements of many economic growth ratio comparison analysis.

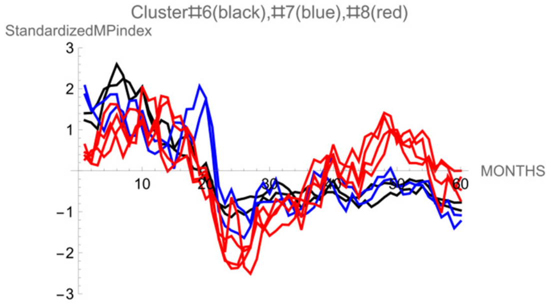

Fig. 3. Standardization removes the variance information.

B. Clustering Results

In the section, we shall show the results of the automakers' market capital analysis by the Amplitude-based clustering.

The results are illustrated by a dendrogram and its heatmap (see Fig. 4 and Fig. 5). The dendrogram and heatmap are ones after the matrix serialization; the company list order is there arranged by the serialization. In the heatmap in Fig. 5, because the distance 0 is colored in yellow, yellow-colored cells are diagonally placed.

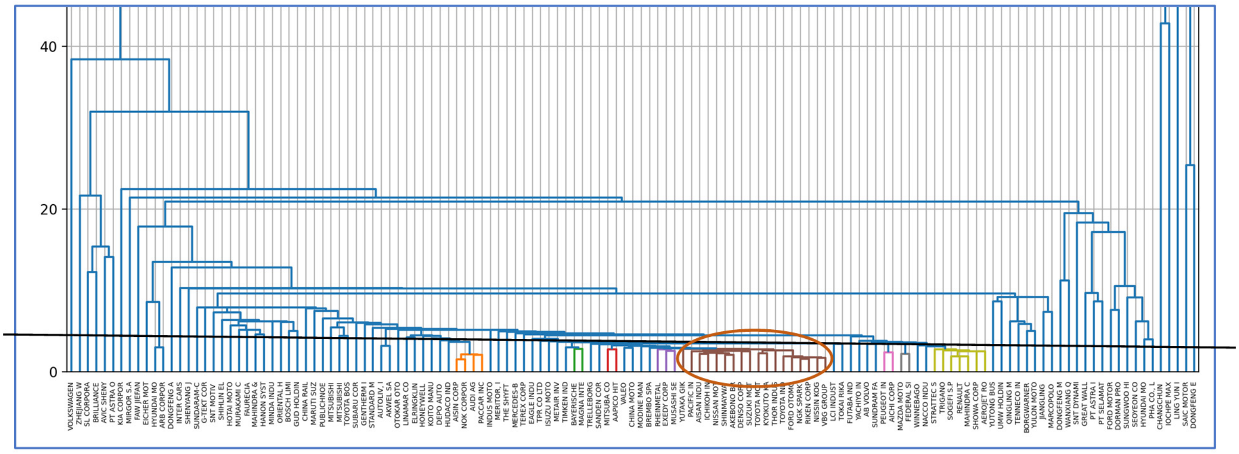

Fig. 4. Resultant dendrogram.

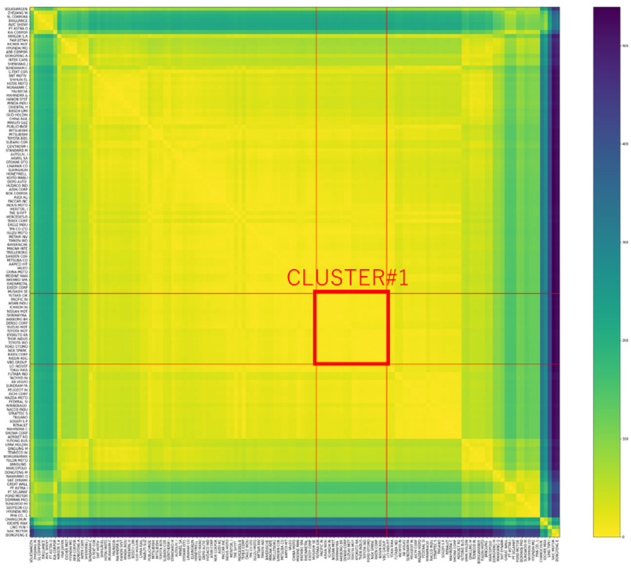

Fig. 5. Resultant heatmap of the distance matrix.

We are interested in the movement of the TOYOTA group. From the heatmap, we shall extract a cluster which includes TOYOTA MORTORS (hereafter TOYOTA). We title this cluster CLUSTER#1. The cluster size is set to be 20 here (see the brown-colored cluster in Fig. 4). The vertical height in the dendrogram means the distance between companies.

In Fig. 4, the boundary value of the CLSUTER#1 is represented by the horizontal black line. If the boundary value is increased, the CLUSTER#1 size can also be increased. We can identify the increase operation's result visually in both the dendrogram and the heatmap. This can be done because of the matrix serialization.

When we use time series data clustering such as k-shape method, the elbow method [31–33] or so is used to select an appropriate number of clusters. However, in the Amplitude-based clustering, there is no concept of an appropriate number of clusters and then a certain cluster area of our interests should be extracted. For example, the size of the above-mentioned cluster with TOYOTA is set to be 20. The size can be changed. This is because the volatility (variance)/distance range is very wide in the amplitude-based clustering.

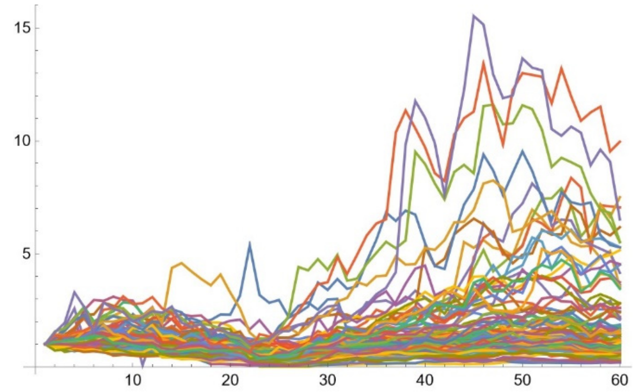

Fig. 6. All companies' indexed movement of market capital.

In Fig. 6, the whole companies' indexed market capital data are illustrated. The highest index value records about 15 (see Fig. 6). As shown here, the variance range is very wide. Therefore, in the Amplitude-based clustering, we are supposed to focus on a cluster with a company of our interests.

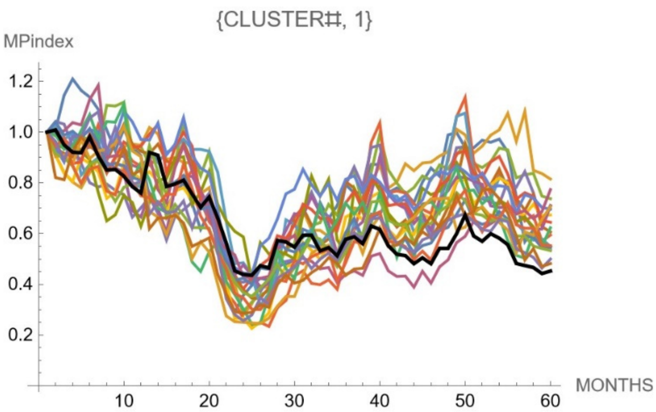

Fig. 7. Movement of Cluster#1 including TOYOTA of which size is 20.

The indexed market capital movement of the above-mentioned CLUSTER#1 is shown in Fig. 7. There TOYOTA's movement is illustrated in black. In CLUSTER#1, there exist at least 4 TOYOTA group companies such as DENSO and TOYOTA INDUSTRIES. In general, since group companies hold mutually stocks and invest to each other in the group, the stock fluctuations are likely to have a high similarity [34]. In this case, the group companies' high similarity can be seen. The pattern similarity among the group companies is our future research theme. We will continue to analyze that using a time series data analysis.

IV. Result by UMAP

In the section, the result by the UMAP dimension reduction is illustrated.

In this analysis, the input data to the UMAP program is the 135 by 135 distance matrix because the number of companies is 135. The distance is the above-mentioned Euclidean distance of the distance of the Amplitude-based clustering (see Section III.A).

In Fig. 8, the UMAP result is shown. There the dimensional reduction from 135-dimentional world to 2-dimentional world is conducted. The point is that all companies' data are plotted on a curve. We found that on the curve, companies with a high similarity are plotted nearer.

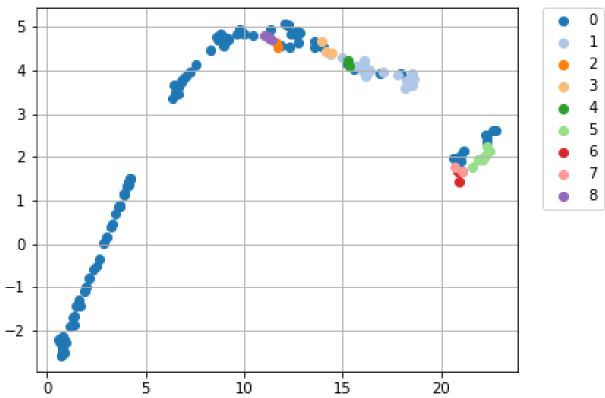

Fig. 8. Two-dimensional mapping by UMAP.

In Fig. 8, the above-mentioned CLUSTER#1 companies are plotted in pale-blue. In Fig. 4, the threshold level of CLUSTER#1 is illustrated as the horizontal black line. By the threshold level, we can extract other 7 clusters, from CLUSTER#2 to CLUSTER#8, which of distances are lower than the threshold level. In the UMAP result, these 7 clusters are plotted nearer on the resultant curve. We can identify these 7 clusters visually using different 7 colors in Fig. 8.

We title the other companies other than (CLUSTER#1 to CLUSTER#8) "GROUP#0" which is colored in blue. GROUP#0 is not a cluster. As shown in Fig. 4 (the dendrogram), Group#0 is the rest of the whole except CLUSTER#1 to CLUSTERR#8. If we would like observe the CLUSTER#0, a further clustering is needed to investigate its details.

V. Evaluation

In the section, we describe the meanings of X and Y axes on UMAP in this case.

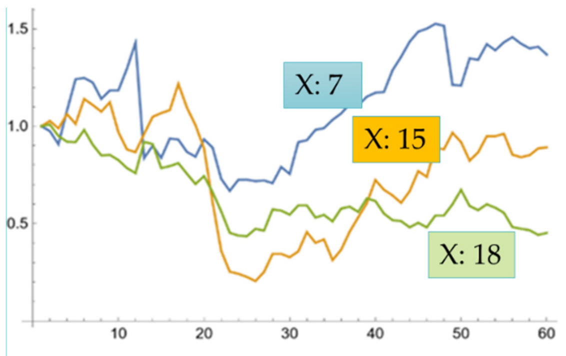

Fig. 9. Three companies' movements with different X values s.

First, the meaning of the X axis on the UMAP is explained as the resilience from the financial crisis. In Fig. 9, three companies' market capital movement with different the X values on the UMAP. The X values are approximately 7, 15, and 19. The less X value is, the larger the resilience is, as shown in Fig. 9.

Then, we interpret the UMAP X-axis as the weakness of the resilience.

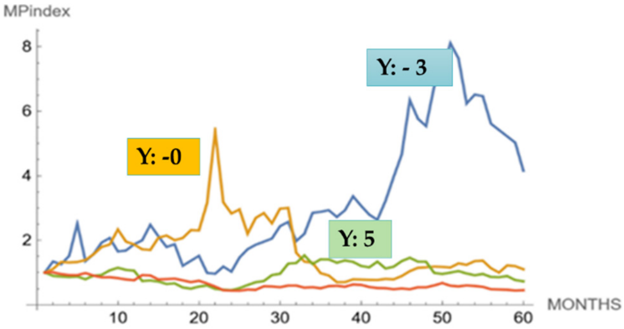

Fig. 10. Five companies'movements with different Y values.

Next, the meaning of the Y axis is interpreted as a variance width of the market capitalization. **The less Y value is, the larger the variance scale is**.

The company in blue in Fig. 10 has the lowest Y value, about -3 on the UMAP and the growth range was the largest; The company temporarily recorded the capitalization by up to 8 times that of January 2007. Then, the yellow line companies' movement has the second lowest Y value, about 0, among the above the four companies. The company's capitalization had the next largest range of its changing around 21st month, although the company had the decline after the 30th month. The green line company has the large Y value, about 5; the movement is very flat, comparing the above two companies.

From this result, we have learnt the following result: In the analysis of time series changing market capitalization, the dual approach by Amplitude-based clustering and dimensional compression UMAP are very effective. Amplitude-based can extract the similar movement company clusters and UMAP can extract the principal factors as the axis. There, the meaning of the axis can be clearly visualized.

VI. Conclusion

In this paper, we conducted time series data clustering by using market capitalization data. The data is that of the top 135 automobile manufacturing companies from January 2007 to December 2011. The analysis method is the Amplitude-based clustering. The results were visualized by the dimensional compression method UMAP which is the dimensional reduction tool. In this UMAP result, the principal components could be extracted and all the companies are plotted on the almost one curve. This is because the distance range is very large in this clustering. The one of the extracted UMAP axes represents the resilience of the company from the 2008 financial crisis.

As a result, we were able to discover two interesting findings. First, companies with affiliated relationships such as TOYOTA and DENSO were clustered over a short distance on UMAP. We could find the similar movement of the group companies.

The second point is that we could clarify the meaning of the two axes of the UMAP mapping. One axis shows the resilience and another does the variance width of the movement. We can say that this clustering and dimensional reduction approach is effective to evaluate the resilience of industries.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

For S.M. and Y.S. substantially contributed to the research conceptualization. S.M. significantly contributed to data analysis and interpretation. Y.S. substantially contributed to the manuscript drafting. All authors critically reviewed and revised the manuscript draft and approved the final version for submission.

Funding

The research was partly supported by Japanese JSPS KAKENHI Grant Number 20H01537.

References

[1] Y. Shirota and B. Chakraborty, "Amplitude-based time series data clustering method," Gakushuin Economics Papers, vol. 59, no. 2, 2022.

[2] J. Paparrizos and L. Gravano, "k-shape: Efficient and accurate clustering of time series," in *Proc. the 2015 ACM SIGMOD International Conference on Management of Data*, 2015, pp. 1855–1870.

[3] J. Paparrizos and L. Gravano, "Fast and accurate time-series clustering," *ACM Transactions on Database Systems*, vol. 42, no. 2, pp. 1–49, 2017.

[4] S. Aghabozorgi, A. S. Shirkhorshidi, and T. Y. Wah, "Time-series clustering — a decade review," *Information Systems*, vol. 53, pp. 16–38, 2015.

[5] F. Petitjean, A. Ketterlin, and P. Gan ҫarski, "A global averaging method for dynamic time warping, with applications to clustering," *Pattern Recognition*, vol. 44, no. 3, pp. 678–693, 2011.

[6] H. Ito, A. Murakami, N. Dutta, Y. Shirota, and B. Chakraborty, "Clustering of ETF data for portfolio selection during early period of corona virus outbreak," *Gakushuin Journal of Economics*, vol. 58, no. 1, pp. 99–114, 2021.

[7] Y. Shirota, B. Chakraborty, and H. Yoshiura, "An analysis of Indian companies' profit growth pattern by amplitude-based clustering," presented at the 5th IEEE Intl. Conf. on Knowledge Innovation and Invention (ICKII) 2022 Hualien, Taiwan and online, July 23, 2022.

[8] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, no. 11, 2008.

[9] D. Kobak and P. Berens, "The art of using t-SNE for single-cell transcriptomics," *Nature Communications*, vol. 10, no. 1, pp. 1–14, 2019.

[10] D. Kobak and G. C. Linderman, "Initialization is critical for preserving global data structure in both t-SNE and UMAP," *Nature Biotechnology*, vol. 39, no. 2, pp. 156–157, 2021.

[11] L. Van Der Maaten, "Accelerating t-SNE using tree-based algorithms," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 3221–3245, 2014.

[12] L. McInnes, J. Healy, and J. Melville, "Umap: Uniform manifold approximation and projection for dimension reduction," *arXiv preprint arXiv:1802.03426*, 2018.

[13] L. McInnes. (2022). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. [Online]. Available: https://umap-learn.readthedocs.io/en/latest/#

[14] F. Kurokawa, *Trend and Structure of Automotive Industry*, 4th ed. Tokyo: Shuwa System, 2022.

[15] H. Tatsat, S. Puri, and B. Lookabaugh, *Machine Learning and Data Science Blueprints for Finance*, O'Reilly Media, 2020.

[16] Y. Hilpisch, *Artificial Intelligence in Finance*, O'Reilly Media, 2020.

[17] S. Jansen, "Machine Learning for Algorithmic Trading: Predictive models to extract signals from market and alternative data for systematic trading strategies with Python," Packt Publishing Ltd, 2020.

[18] I. Aldridge and M. Avellaneda, *Big Data Science in Finance*, John Wiley & Sons, 2021.

[19] F. Chollet, *Deep Learning with Python*, Simon and Schuster, 2021.

[20] T. A. Borges and R. Neves, "Financial data resampling for machine learning based trading: Application to cryptocurrency markets," *Springer Nature*, 2021.

[21] J. J. Murphy, *Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications*, Penguin, 1999.

[22] Y. Shirota, K. Kuno, and H. Yoshiura, "Time series analysis of SHAP values by automobile manufacturers recovery rates," in *Proc. the 2022 6th International Conference on Deep Learning Technologies*, July 2022, pp. 135–141.
[23] T. Hashimoto, Y. Shirota, and B. Chakraborty, "SDGs India index analysis using SHAP," presented at the International Electronics Symposium (IES) 2022, Surabaya, Indonesia and online, August 9–11, 2022.
[24] M. Fujimaki, E. Tsujiura, and Y. Shirota, "Automobile manufacturers stock price recovery analysis at covid-19 outbreak," in *Proc. 6th World Conference on Production and Operations Management P&OM 2022*, 2022, pp. 431–439.
[25] Y. Shirota, K. Yamaguchi, A. Murakami, and M. Morita, "An analysis of political turmoil effects on stock prices: A case study of US-China trade friction," in *Proc. the First ACM International Conference on AI in Finance*, 2020, pp. 1–7.
[26] C. M. Bishop and N. M. Nasrabadi, *Pattern Recognition and Machine Learning*, no. 4, Springer, 2006.
[27] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, Elsevier, 2006.
[28] M. L. De Prado, *Advances in Financial Machine Learning*, John Wiley & Sons, 2018.
[29] M. M. L. de Prado, *Machine Learning for Asset Managers*, Cambridge University Press, 2020.
[30] S. Matsuhashi and Y. Shirota, "Note for shape-based clustering in stock prices," in *Proc. 2022 12th International Congress on Advanced Applied Informatics (IIAI-AAI)*, 2022, pp. 373–377.
[31] D. Marutho, S. H. Handaka, and E. Wijaya, "The determination of cluster number at k-mean using elbow method and purity evaluation on headline news," in Proc. *2018 International Seminar on Application for Technology of Information and Communication*, 2018: IEEE, pp. 533–538.
[32] P. Bholowalia and A. Kumar, "EBK-means: A clustering technique based on elbow method and k-means in WSN," *International Journal of Computer Applications*, vol. 105, no. 9, 2014.
[33] M. Syakur, B. Khotimah, E. Rochman, and B. D. Satoto, "Integration k-means clustering method and elbow method for identification of the best customer profile cluster," in *Proc. IOP Conference Series: Materials Science and Engineering*, 2018, vol. 336, no. 1: IOP Publishing, p. 012017.
[34] Solo TOYOTA's approaching trials, *Weekly _TOYOKEIZAI*, no. 2022/8/6, pp. 40–65, 2022.

Copyright © 2023 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (CC BY 4.0).