Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography
Vol. 35, No. 5, 329-338, 2017
https://doi.org/10.7848/ksgpc.2017.35.5.329

ISSN 1598-4850(Print)
ISSN 2288-260X(Online)
Original article

Classifying Alley Markets through Cluster Analysis Using Dynamic Time Warping and Analyzing Possibility of Opening New Stores*

Kang, Hyun Mo[1] · Lee, Sang-Kyeong[2]

Abstract

This study attempts to classify 1008 alley markets in Seoul through cluster analysis using Dynamic Time Warping, one of the methods used to analyze the similarity of time series, and evaluate the possibility of opening new stores. The sequence of the gross sales of an alley market and that of gross sales per store stand for the potential of growth and profitability of the market, respectively and are used as variables for cluster analysis. Five clusters are obtained for the gross sales and four clusters for the gross sales per store. These two types of clusters are again classified as rising and falling trends, respectively, and the combination of these trends produces four categories. These categories are used to evaluate the possibility of opening new stores in alley markets. The results show that the southeast which is relatively wealthy inferior to other regions in opening new stores. Alley markets in the northeast and the southwest are better than other regions such that opening a new store is justified. In the northwest, there are many markets with trend of gross sales and that of gross sales per store moving in opposite directions, and new store openings in these markets should be postponed.

Keywords : Cluster Analysis, Dynamic Time Warping, Time Series, Alley Market

1. Introduction

Currently, obtaining meaningful information from vast amounts of data is gaining importance. Many government agencies worldwide have been allowing the public access to their data, which generates new value. The Seoul Metropolitan Government in Korea is also pursuing this policy and operating a website called “Seoul Open Data Plaza”, which contains a large amount of data collected over the years. This website also provides public data on alley markets about the number of stores, sales, floating population, and so on. The Seoul Metropolitan Government classifies the alley market into residential, semi-residential, and non-residential areas according to hinterland. This classification mainly takes into account locational characteristics but has limitations because it does not diagnose the change in market status over time. Taking flow of time into consideration when classifying alley markets is significant in two aspects. First, all alley markets have their own lifecycle. It is important to figure out the cycle by analyzing the inauguration, growth, stabilization, and decline of the market. For example, stores that have opened during the beginning or growth period of the market will have a high rate of return, but those opened during the stabilization period or the decline period may not be so successful. In fact, it may be counterproductive for them. Second, we need to find the change in the seasonal characteristics of each market. Some markets have a high rate of return in summer while the same is true for others during winter. For markets near colleges, sales decrease in the holiday season, while the markets for the younger generation,

Received 2017. 08 01, Revised 2017. 08. 24, Accepted 2017. 10. 26
*This paper is an enriched and revised version of a paper presented in Korean at the 2017 Spring Congress of Korea Planning Association.
1) Department of Urban Planning, Gachon University, Korea (E-mail: khm2963@naver.com)
2) Corresponding Author, Member, Department of Urban Planning, Gachon University, Korea (E-mail: skylee@gachon.ac.kr)

such as Gangnam, Shinchon, and Itaewon, experience an increase in sales during the vacation season. Therefore, it is important to take the flow of time into consideration while characterizing each market.

Cluster analysis can be one of the effective methods to classify alley markets. Distances such as the Euclidean, the Manhattan, and the Minkowski, are generally used to measure similarity in cluster analysis. These distances cause distortion when measuring the similarity of time series data. DTW (Dynamic Time Warping) method is advantageous as it minimizes the distortion of distance. DTW is widely used in speech recognition, bioinformatics, and online handwriting recognition (Lee and Oh, 2011). However, we cannot find studies applying DTW method when clustering alley markets. This paper, therefore, aims to classify alley markets through cluster analysis using DTW method and analyze the possibility of opening new stores. Specifically, the research subject is alley markets in Seoul, and their gross sales and gross sales per store are used as variables for the cluster analysis.

2. Literature Review

We examine earlier research on trade area analysis and application of DTW to geospatial analysis. Tae and Rhim (2010) attempted to create a model to determine appropriate retail store locations. They calculated occupation population for each store by applying Huff's Probability Theory for hypermarkets in Seoul. They categorized each unit of a district into four types of market according to the prevailing competition. The most appropriate place to open a new store is one with the largest share of the occupation population, and such a place is far from a closely competing market and as less not competing market as possible at the same time. Lee (2003) extracted the domain of trade areas through spatial data mining methods. The DENCLUE method showed excellent results when compared with the K-medoids technique or the SMTIN technique. Kwon and Yu (2015) applied the AMOEBA method to official land value and floating population data to identify the boundaries of trade areas. Lee and Lee (2014) found location decision factor for the marts and optimum range for trade areas in cities. Optimum range for trade area was larger in large cities than regional small-to-medium cities, and location decision factor is different both large cities and regional small-medium cities. Shin *et al*. (2002) identified leading industries that moved to the center of retail trade area and found that the centers shifted over time. While there are several studies on segmentation, range, and choice of location, it is difficult to find classifying trade area with indexes over time.

There are a few studies which have applied cluster analysis to the similarity of time series data. Park *et al*. (2011) conducted cluster analysis using land transaction data in Goyang City. This study could be a good case as it adopted DTW method. However, it is basically different from this paper because neither did it apply DTW method to alley markets nor suggest an effective way to classify them.

3. Methodology and Data

3.1 Analysis procedure

Gross sales of an alley market and gross sales per store are measured by DTW for calculating similarity. After building a proximity matrix by using DTW method, clusters are divided on the basis of hierarchical agglomerative clustering using Ward's method (Ward, 1963) for recursively updating the proximity matrix at each step. We recognize the features of clusters, and then we obtain the rising trend and falling trend of both clusters of gross sales of an alley market and clusters of the gross sales per store of alley markets. These clusters are reclassified into two groups based on the trends. We suggest the evaluation method for possibility of opening a new store according to the reclassified clusters.

DTW measures the similarity between two time series. More specifically, DTW matches two time signals by computing a temporal transformation causing the signals to be aligned (Akl and Valaee, 2010). Assume two time series of alley markets, X and Y, of same length n, are $X=(x_1, x_2, \cdots, x_n)$ and $Y=(y_1, y_2, \cdots, y_n)$, respectively. DTW($X$, Y) is computed based on dynamic programming using the following formulation:

$$D(i,j) = d(x_i, y_j) + \min \begin{cases} D(i, j-1) \\ D(i-1, j) \\ D(i-1, j-1) \end{cases} \quad (1)$$

where the distance function $d(\cdot,\cdot)$ varies with the application. In this study, $d(x_i, y_j)$ is defined as,

$$d(x_i, y_j) = |x_i - x_j| \quad (2)$$

For further explanation of DTW, the reader is referred to Sardá-Espinosa (2017).

The proximity matrix is created by comparing the similarity in gross sales of alley markets using DTW method. The other proximity matrix is generated as above for gross sales per store, which means gross sales of an alley market divided by the number of stores. Cluster analysis is performed based on the proximity matrix. Hierarchical cluster analysis is a method of grouping the nearest object stage by stage and forming one cluster at the end (Song and Chang, 2010). This study adopts Ward's method which is not sensitive to noise or outlier. The procedure for dividing appropriate clusters is as follows. First, a dendrogram is made through Ward's method. Second, the number of clusters increases and a representative value for each cluster is extracted. Third, the representative value of each cluster is compared with the previous one. Fourth, clusters are selected when different characteristics are not observed between them.

We examine then the features of both clusters of gross sales of an alley market and clusters of gross sales per store of alley markets. The HP filter is applied to the prototype of a cluster to produce a trend. HP filter is a method of extracting long-term trend changes after decomposing time series into trend and cyclic fluctuations based on the assumption that time series data consist of trend fluctuations and cyclic fluctuations (Loh and Sung, 2015). We categorize both clusters of gross sales of alley markets and clusters of gross sales per store as rising trend and falling trend. The possibility of opening a new store is evaluated by four classifications of gross sales of an alley market and gross sales per store on the basis of trend.

3.2 Data

This study is conducted on the alley market profiling data obtained from the "Seoul Open Data Plaza". The definition of alley market is as given by "Seoul Open Data Plaza". Since the alley market consists of alley stores, it is necessary to examine the definition of alley store. The alley store is defined as follows: ① the 43 kinds of stores which are closely related to everyday life, ② the store which is adjacent to a narrower road than a four-lane road ③ the store which is not included in the developed market, ④ the store which is not near large retailers, and ⑤ the store of which hinterland has a high density of residences. The alley market is defined as follows: ① the alley market contains a certain number of stores, and ② the density of alley stores is high. This study proceeds with 1008 alley markets data provided by "Seoul Open Data Plaza". Fig. 1 shows the distribution of alley markets in Seoul.

This study uses gross sales of an alley market and gross sales per store of an alley market for cluster analysis. The gross sales of an alley market are a metric for the overall sales of a market. The monthly gross sales of an alley market are divided by monthly numbers of stores to obtain the gross sales per store. The period is from January 2013 to October 2016 for a total of 46 months of data. The unit of analysis is polygon data provided by "Seoul Open Data Plaza". The gross sales figure of an alley market is estimated on the basis of sales information provided by the BC card, Shin Han card, and Niceinfo; the estimation of the proportion of card sales and cash sales is also used. The number of stores of an alley market was constructed using licensing store information in Seoul.

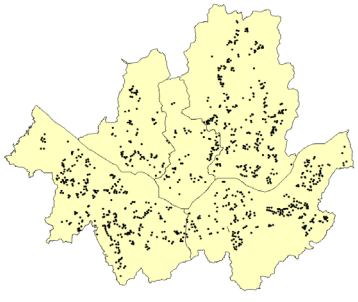
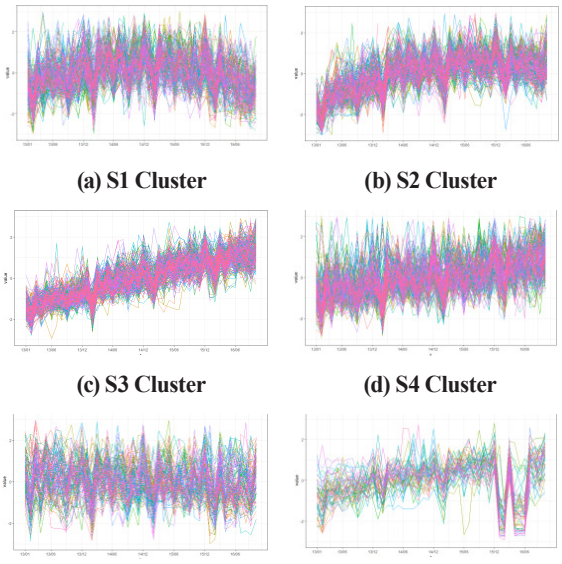


Fig. 1. Distribution of alley markets in Seoul

4. Application and Evaluation

4.1 Cluster analysis of the gross sales of alley markets

Alley markets are classified into six clusters by cluster analysis applying DTW to gross sales of an alley market. Fig. 2 shows all the time sequences at each cluster plotted for showing the features. The S6 Cluster in Fig. 2 is ruled out due to a drastic drop in data.

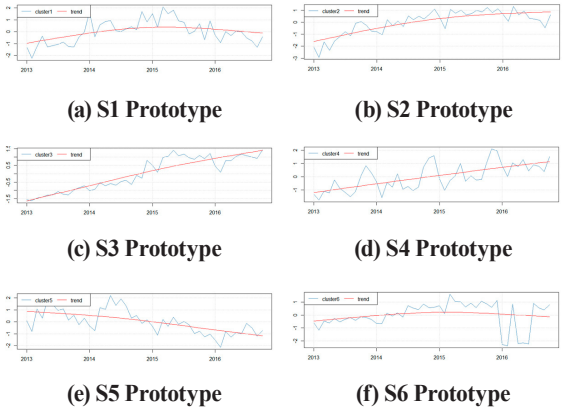


(a) S1 Cluster (b) S2 Cluster

(c) S3 Cluster (d) S4 Cluster

(e) S5 Cluster (f) S6 Cluster

Fig. 2. Results of cluster analysis by gross sales of alley market



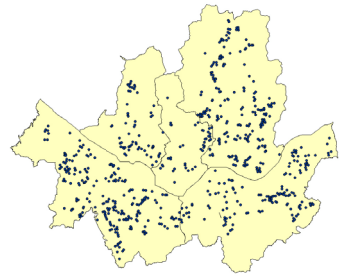
(a) S1 Prototype (b) S2 Prototype

(c) S3 Prototype (d) S4 Prototype

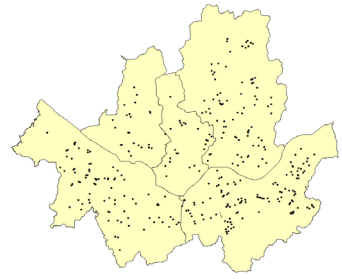
(e) S5 Prototype (f) S6 Prototype

Fig. 3. Prototype and HP filter trend of gross sales by alley market cluster

Fig. 3 shows prototype of gross sales by alley market cluster and trend produced by HP filter. S1 Cluster represents the recently fall in gross sales of an alley market, although it rose in the past. The S2 Cluster shows the rise in gross sales are slowing and S3 Cluster a rapidly rise in gross sales. The S4 Cluster and S5 Cluster have seasonal characteristics, and the former is rising in gross sales, while the latter is moving in opposite directions. As a result of trend analysis using HP filter, S2, S3, and S5 Clusters can be categorized as a rising trend, and S1 and S5 Clusters as a falling trend.



(a) Alley markets of rising trend in gross sales



(b) Alley markets of falling trend in gross sales

Fig 4. Distribution of alley markets by trend of gross sales

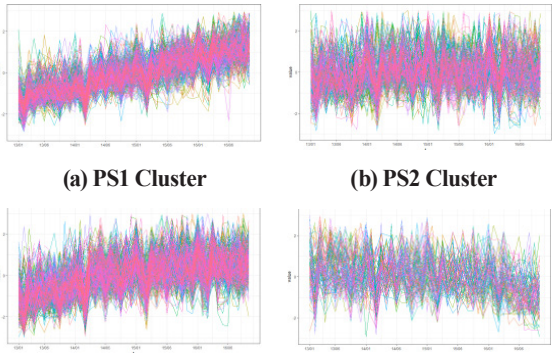
Figs. 4(a) and 4(b) present the spatial distribution of alley markets of rising trend and falling trend, respectively. As shown in Figs, alley markets in rising trend are much more than falling trend and this is also shown in Table 1. The percentage of alley markets which shows a rising trend is 66.7% and that a falling trend is 27.7%. Seoul is generally divided into five regions: center, southeast, northeast, southwest, and northwest region. The northeast, the northwest and the southwest regions show a rising alley markets percentage of more than 70%, but the center and the southeast show relatively lower 62.5% and 53.9%, respectively. Specifically, the percentage of falling alley markets of the southeast region is a relatively high 42.3%. Although the southeast region is wealthier than other regions, in terms of growth of an alley market, it is inferior to other regions.

Table 1. Distribution of clusters reclassified by trend of gross sales

Region		Rising Trend				Falling Trend			Unidentified	Total
		S2	S3	S4	Subtotal	S1	S5	Subtotal	S6	
Center	Count	12	15	13	40	9	7	16	8	64
	% of Total	18.8	23.4	20.3	62.5	14.1	10.9	25.0	12.5	100.0
Southeast	Count	44	47	49	140	75	35	110	10	260
	% of Total	16.9	18.1	18.9	53.9	28.9	13.5	42.3	3.9	100.0
Northeast	Count	70	76	61	207	41	25	66	17	290
	% of Total	24.1	26.2	21.0	71.4	14.1	8.6	22.8	5.9	100.0
Southwest	Count	63	97	64	224	40	27	67	18	309
	% of Total	20.4	31.4	20.7	72.5	12.9	8.7	21.7	5.8	100.0
Northwest	Count	11	28	22	61	12	8	20	4	85
	% of Total	12.9	32.9	25.9	71.8	14.1	9.4	23.5	4.7	100.0
Total	Count	200	263	209	672	177	102	279	57	1008
	% of Total	19.8	26.1	20.7	66.7	17.6	10.1	27.7	5.7	100.0

4.2 Cluster analysis of gross sales per store of alley markets

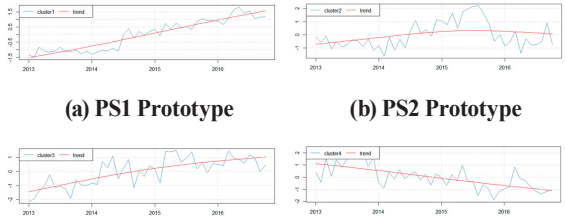
When considering opening a new store, the time series of the gross sales of an alley market is a significant index which informs the growth potential of the market and that of the gross sales per store is another important index in that it stands for the profitability potential of the market. For this reason, the cluster analysis using DTW is carried out on time sequences of gross sales per store. As a result of the cluster analysis, alley markets are classified four clusters. Fig. 5 shows all the time sequences at each cluster plotted for showing the features.



(a) PS1 Cluster (b) PS2 Cluster

(c) PS3 Cluster (d) PS4 Cluster

Fig. 5. Results of cluster analysis using gross sales per store of alley markets



(a) PS1 Prototype (b) PS2 Prototype

(c) PS3 Prototype (d) PS4 Prototype

Fig. 6. Prototype and HP filter trend of gross sales per store by alley market cluster

Fig. 6 shows prototype of gross sales by alley market cluster and trend produced by HP filter. PS1 Cluster shows a rapidly rise in gross sales per store. PS2 Cluster represents the recently fall in gross sales per store, although it rose in the past. PS3 and PS4 Cluster have seasonal characteristics, and the former is rising in gross sales per store, while the latter is moving in opposite directions. As a result of trend analysis using HP filter of gross sales per store, PS1 and PS3 Cluster can be categorized as rising trend and PS2 and PS4 as falling trend.

Figs. 7(a) and 7(b) shows the spatial distribution of alley markets of rising trend and falling trend, respectively. As seen in the trend analysis of gross sales, alley markets in rising trend are also much more than falling trend and this is shown in Table 2. The percentage of alley markets which

show rising trend is 66.2% and that of alley markets which show falling trend is 33.8%. By region, the northeast and southwest show the percentage of rising alley markets as more than 70%, but the center, southeast, and northwest regions show relatively lower 62.5%, 53.5% and 63.5% respectively.

(a) Alley markets of rising trend in gross sales per store

(b) Alley markets of falling trend in gross sales per store

Fig. 7. Alley market distribution by trend of gross sales per store

Table 2. Distribution of clusters reclassified by trend of gross sales per store

Region		Rising Trend			Falling Trend			Total
		PS1	PS3	Subtotal	PS2	PS4	Subtotal	
Center	Count	16	24	40	19	5	24	64
	% of Total	25.0	37.5	62.5	29.7	7.8	37.5	100.0
Southeast	Count	64	75	139	86	35	121	260
	% of Total	24.6	28.9	53.5	33.1	13.5	46.5	100.0
Northeast	Count	112	105	217	47	26	73	290
	% of Total	38.6	36.2	74.8	16.2	9.0	25.2	100.0
Southwest	Count	131	86	217	70	22	92	309
	% of Total	42.4	27.8	70.2	22.7	7.1	29.8	100.0
Northwest	Count	23	31	54	22	9	31	85
	% of Total	27.1	36.5	63.5	25.9	10.6	36.5	100.0
Total	Count	346	321	667	244	97	341	1008
	% of Total	34.3	31.8	66.2	24.2	9.6	33.8	100.0

As with the analysis of sales, the percentage of falling alley markets of the southeast region is relatively higher at 46.5%. Considering the profitability potential of a store, it can be inferred that the southeast is inferior to other regions.

4.3 Analyzing the possibility of opening a new store

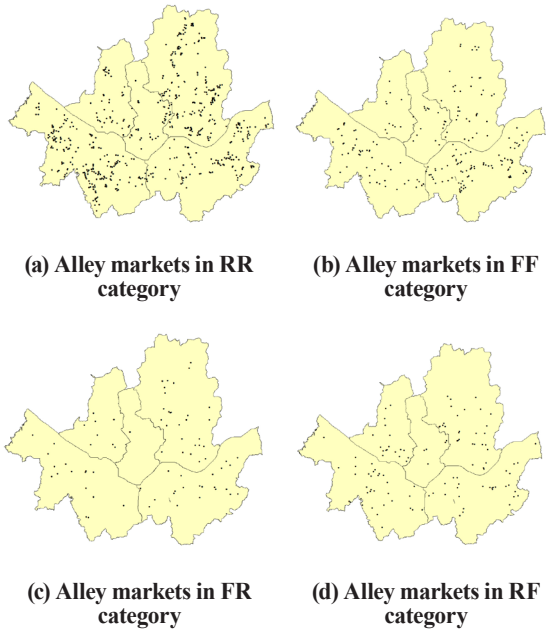
Table 3 shows the result of four categories reclassified by using the combination of the rising and falling trends. The RR category, which indicates the alley market most advantageous to open a new store, has a rising trend in both gross sales of an alley market and gross sales per store of an alley market. The FF category, which indicates the alley market disadvantageous to open a new store, has a falling trend in both gross sales of an alley market and gross sales per store of an alley market. The FR category in which gross sales per store of an alley market are rising due to easing competition has falling trend in gross sales of an alley market. The RF category in which gross sales per store of an alley market are falling due to fierce competition has a rising trend in gross sales of an alley market. FR and RF categories

indicate that opening a new store should be postponed here. Alley market of FR and RF categories are inferior to RR for opening a new store but superior to the FF category.

Table 3. Four categories produced by the combination of trends

		Gross sales per store	
		Rising Trend (PS1, PS3)	Falling Trend (PS2, PS4)
Gross sales	Rising Trend (S1, S3, S4)	RR	RF
	Falling Trend (S1, S5)	FR	FF

Fig. 8 show the spatial distribution of alley markets in 4 categories which are produced by the combination of the trend of gross sales and the trend of gross sales per store. As you can see from the Figs, the number of alley markets is in the order of RR, FF, RF, and FR. Table 4 shows the distribution of 4 categories by region in Seoul. RR and FR categories represent highest and lowest percentage in all regions respectively. By region, the southeast shows relatively lower 42.7% in the RR category and relatively higher 33.9% in FF category. This means that the southeast markets are disadvantageous for opening a new store. On the contrary, alley markets of the northeast and southwest, which show a higher percentage of RR category and a lower percentage of FF category, are advantageous in opening a new store. Opening new stores should be postponed in alley markets of the northwest, which shows a higher percentage of RF category.



(a) Alley markets in RR category

(b) Alley markets in FF category

(c) Alley markets in FR category

(d) Alley markets in RF category

Fig. 8. Alley market distribution by category

Table 4. Alley markets in four categories produced by the combination of trends

Region		RR	FF	FR	RF	Unidentified	Total
Center	Count	35	15	1	5	8	64
	% of Total	54.7	23.4	1.6	7.8	12.5	100.0
Southeast	Count	111	88	22	29	10	260
	% of Total	42.7	33.9	8.5	11.2	3.9	100.0
Northeast	Count	177	38	28	30	17	290
	% of Total	61.0	13.1	9.7	10.3	5.9	100.0
Southwest	Count	191	51	16	33	18	309
	% of Total	61.8	16.5	5.2	10.7	5.8	100.0
Northwest	Count	43	12	8	18	4	85
	% of Total	50.6	14.1	9.4	21.2	4.7	100.0
Total	Count	557	204	75	115	57	1008
	% of Total	55.3	20.2	7.4	11.4	5.7	100.0

5. Conclusion

In this study, we classify alley markets in Seoul through cluster analysis using Dynamic Time Warping and evaluate the possibility of opening new stores. The gross sales of an alley market and the gross sales per store of an alley market are used as variables for clustering. The time series of the gross sales of an alley market stands for the growth potential of the market and that of gross sales per store the profitability potential of the market.

The analysis process is summarized as follows. The cluster analysis using DTW method is applied to two kinds of time series of 1008 alley markets. Five clusters are obtained for the gross sales of an alley market, and four clusters for gross sales per store of an alley market. These clusters are categorized into rising and falling trends through trend analysis using HP filter. The combination of these trends produces four categories, which are used to evaluate the possibility of opening new stores in alley markets. From a methodological point of view, a series of analysis processes are evaluated to have been performed well. We can confirm the usefulness of cluster analysis using DTW and the applicability of this method for various market indexes of time series form.

The results of the analysis are summarized as follows. The southeast is inferior to other regions when opening new stores. Alley markets in the northeast and the southwest are better than other regions such that opening a new store is justified. In the northwest, there are many markets with trend of gross sales and that of gross sales per store moving in opposite directions, and new store openings in these markets should be postponed. From these results, we can realize the common sense that a wealthy neighborhood would be good for opening a store is wrong; the southeast region is relatively wealthy in Seoul but it is not good place to open new stores.

This study is limited in that it only analyzes the gross sales index of the alley market. Deciding whether to open a new store requires more understanding the various characteristics of alley markets. We expect further studies to analyze more alley market characteristics.

Acknowledgments

This work was supported by the Gachon University research fund of 2015. (GCU-2015-0088)

References

Akl, A. and Valaee, S. (2010), Accelerometer-based gesture recognition via dynamic time warping, affinity propagation, & compressive sensing, *2010 IEEE International Conference on Acoustics, Speech and Signal Processing, Dallas, TX, 2010*, pp. 2270-2273.

Kwon, P. and Yu, K. (2015), Trade area delimitation using AMOEBA technique, *Journal of the Korean Society for Geospatial Information Science*, Vol. 23, No. 2, pp. 11-16.

Lee, B. (2003), Defining of trade area using spatial data mining technique in business GIS, *The Journal of GIS Association of Korea*, Vol. 11, No. 2, pp. 171-184.

Lee, S. and Lee, Y. (2014), A study on characteristics of location decision factors of mart by using GIS – comparative analysis of large and small-medium cities, *Journal of the Korean Cadaster Information Association*, Vol. 30, No. 3, pp. 47-57.

Lee, S. and Oh, K. (2011), Finding the optimal frequency for trade and development of system trading strategies in futures market using dynamic time warping, *Journal of the Korean Data & Information Science Society*, Vol. 22, No. 2, pp. 255-267.

Loh, J. and Sung, J. (2015), A study on the business cycle and causal relationship in housing market, *Korea Real Estate Academy Review*, Vol. 61, pp. 208-221.

Park, M., Park, K., and Ahn, J. (2011), A study on the regional classification method using dynamic time warping, *The Geographical Journal of Korea*, Vol. 45, No. 3, pp. 387-395.

Sardá-Espinosa, A. (2017), Comparing time-series clustering algorithms in R using the dtwclust package, https://cran.r-project.org/web/packages/dtwclust/vignettes/dtwclust.pdf (last date accessed: 31 May 2017).

Shin, W., Jung, C., and Lee, S. (2002), A study on the changing patterns of the commercial power based on retail trade area analysis, *Journal of the Korea Real Estate Analysts Association*, Vol. 8, No. 1, pp. 1-11.

Song, M. and Chang, H. (2010), Characterization of cities in Seoul metropolitan area by cluster analysis, *Journal of the Korean Society for Geo-Spatial Information System*, Vol. 18, No. 1, pp. 83-88.

Tae, K. and Rhim, B. (2010), A study for locating of a new store considering competition for trading area: focusing on the case of hypermarket in Seoul metropolitan area, *Journal of the Economic Geographical Society of Korea*, Vol. 45, No. 5, pp. 609-627.

Ward, J. (1963), Hierarchical grouping to optimize an objective function, *Journal of the American Statistical Association*, Vol. 58, No. 301, pp. 236-244.