Sequential Rule Mining for Analysis of Customer Movements in Different Visits

Pelin Yildirim Taser
Computer Engineering Department
Izmir Bakircay University
Izmir, Turkey
pelin.taser@bakircay.edu.tr

Onur Dogan
Industrial Engineering Department
Izmir Bakircay University
Izmir, Turkey
onur.dogan@bakircay.edu.tr

Abstract

Developing technologies in customer analytics provide several opportunities to retailers. Analyzing customer movements in the stores as a part of customer analytics can reveal various shopping behaviors. It facilitates to understand better customers' visit purposes. This study applies four sequential mining algorithms, CMRules, CMDeo, ERMiner, and RuleGrowth, to analyze the visit purposes of customers in a supermarket. Moreover, it compares variations among different visits belonging to the same customers. This study concludes three main results. First, it indicates that customers prefer visiting the supermarket not only for their specific needs but also for all their needs at every visit. Second, the ERMiner algorithm is faster than the other algorithms. Third, customers who visit {Construction, Kitchen} and {Sanitary ware, Garden} bought at least one product with a high probability. Moreover, this study describes the concept of interesting rule, which has a lower support value and higher confidence value. Customers can visit the supermarket for various purposes resulting in different interesting rules. As an interesting rule in the second visit, purchased customers visited Construction, Garden and Kitchen aisles before leaving the supermarket whereas this rule did not appear in the first visits. Customers visited the Construction aisle more after they visited the Entrance and Ironmongery aisles in their second visit.

Keywords
Customer movement analysis, Sequential rule mining, Interesting rules, Customer analytics, Retail sector.

1. Introduction

Advanced technologies have been rapidly developed in the ever-shifting global retail domain (Adapa et al. 2020; Daunt and Harris 2017; Ferracuti et al. 2019). Some retailer companies are confused by technology-based possibilities and use technologies without a precise perception of customer needs (Inman and Nikolova 2017). Customer analytics has obtained speedup in the last decades (Bonacchi and Perego 2019; Kitchens et al. 2018) thanks to the development of recent high-level technologies, including new channels and tools (Herhausen et al. 2019). There is a greater demand than ever before to serve shoppers, leveraging a vast volume of customer data using customer analytics technologies (Sun et al. 2014; Wedel and Kannan 2016). Customer analytics serves to gain advantages for retailer companies regarding various perspectives such as sales forecasting, marketing, segmentation, dynamic pricing and churn prediction.

Knowing aisles visited by customers in a supermarket involves closely examining customer analytics because it helps to uncover information such as customer shopping behaviors and bottlenecks. First of all, collecting customers' location data is necessary to understand customer visit purposes. Data can be collected in several ways, such as WiFi, RFID and Bluetooth. Dogan and Oztaysi (2018) showed that Bluetooth-based devices are the most appropriate data collection technologies for indoor customer tracking. Data used in this study were collected by iBeacon devices which work based on Bluetooth principles in a supermarket in Turkey.

In the literature, historical customer data such as shopping transactions instead of visited aisles were mainly used to understand customer behaviors in the retail domain. This kind of data provides decision-makers which customers purchased what. Some studies applied various data mining techniques to manage vast amounts of data to clarify this point (Dogan et al. 2019a; Dogan et al. 2019b). In these studies, however, only purchasing transactions were recorded. The answers to the questions are missing: Which aisles were visited by customers in the time window, and they purchased or not. However, knowing both purchased and non-purchased customer aisles can give the advantage to understand customer visit purposes.

Tracing customers' movements in a retail store enable to understand better customers' visit purposes than solely regarding the product purchases, as was the situation with former investigations. There have been few types of research on customer movements in retail stores. The reason was that collecting customers' indoor location data was challenging. Hence, customer location data obtained using Bluetooth will be a springboard for new retailing studies.

In the retail business, those targeting customers for a given marketing policy require to comprehend some characteristics and understand purchasing behavior. The sequential pattern mining approach discovers interesting sequential patterns in sequences. Although numerous researches were introduced to reveal some patterns in sequence databases (Mabroukeh and Ezeife 2010), sequential pattern mining (Agrawal and Ramakrishnan 1995) is probably the most common method. It comprises of detecting subsequences frequently appearing in a set of sequences. Sequential rule mining (SRM) is an alternative for the problem of prediction (Fournier-Viger et al. 2011). A sequential rule shows that if some items appear in a sequence, some other items probably disappear after certain confidence or probability.

This study's motivation is to understand customer visit purposes and reveal the differences among different visits using location data, which refers to customer movements instead of historical transaction data. It contributes to the literature from two perspectives. First, the scope of the research includes multiple visits by the same customers. It enables to see changing customer behaviors. Second, although many studies used transaction data, this study uses customer location data collected by Bluetooth-based devices. Then they are transformed into aisles to categorize products that the customer is interested in. This study is organized as follows. Section 2 presents previous works within a similar scope. Section 3 gives the details of the methods used in this research. Section 4 presents the dataset description, and then it shows experimental results by discussing some managerial implications. Finally, section 5 concludes the study and provides some future directions.

2. Literature Review

Point-of-sale data (POS data) were used traditionally to propose customer shopping behavior in earlier studies (Guadagni and Little 1983; Gupta 1988) from the marketing perspective. Transaction data were also used in many studies to investigate customer needs and visiting purposes (Chang and Tsai 2011; Dogan et al. 2020; Khajvand et al. 2011). For example, Dogan et al. (2020) collected transaction data from a supermarket in Turkey and implemented a fuzzy method to cluster customers according to their purchasing data. Some recent studies related to indoor customer behaviors utilized customer location data adapted the technological improvements to non-invasive data collection. They collected data via sensors to create movements (Dogan et al. 2019a; Dogan et al. 2019b; Yada 2011). Willeims et al. (2017) utilized WiFi technology to organize a retail inventory categorized according to the type of shopping value and the stage of the shopping cycle. Fukuzaki et al. (2015) learned the real number of customers in the shopping mall with WiFi technology. Yewatkar et al. (2016) introduced an intelligent shopping cart that records purchased products and online transactions with RFID and ZigBee. Oosterlinck et al. (2017) established Bluetooth-based tracking devices in a shopping mall and collected data of high quality at a low cost.

Clustering methods are one of the most common ways for customer behavior analysis by segmenting customers. They were also applied to analyze human movements (Landmark and Sjøbakk 2017; Larson et al. 2005). Several customer groups were created, and some hypotheses were discussed in these studies. Classification problems or abstraction of features from customer movement data are also considered another type of customer analytics research (Yada 2011). Some researchers identified human movements with process mining by considering personal paths (Dogan et al. 2019a; Nakatumba and van der Aalst 2009; Ma'arif 2017). Ma'arif (2017) applied a process mining discovery algorithm to illustrate people's everyday movements. Dogan et al. (2019a) used process mining to explore and describe the principal behavioral alterations considering the gender in a visual description. Association rule mining, also known as market basket analysis, is a popular method in the retail domain to obtain information about consumer shopping habits and preferences. For instance, Griva et al. (2018) introduced a customer analytics approach that analyzes customer visit segments from transaction data. They described a customer visit by the purchased product categories and classified the visit purpose using market basket analysis. Similar to association rule mining, sequential rule mining (SRM) is another customer analytics approach that discovers interesting patterns in the customer data. The SRM has a wide range of application areas, such as customer shopping sequences, DNA sequences, intrusion detection, web mining, and customer behavior analysis. In this study, the supermarket customers' routes in their visits at different times were analyzed using the SRM method. For example, Wu and Yu (2020) suggested a sequential rule mining method for the analysis of online customers' search habits to understand consumer needs. They explored how recommendations can assist customers in online shopping by considering their needs. As a contribution to the literature, this study also applies sequential rule mining to uncover interesting and valuable patterns from location data captured by Bluetooth-based devices in a supermarket. Moreover, it presents variations among different visits to show customer behaviors.

3. Methods

This study presents the implementation of SRM algorithms on supermarket customer data for analyzing customer visits at different times. Figure 1 shows the general structure of this study. In the first phase, customer locations are collected using iBeacon devices in the aisles of the supermarket. Second, the obtained supermarket dataset is passed through a data preprocessing step. Third, four different SRM algorithms (CMRules, CMDeo, ERMiner, and RuleGrowth) are applied to the preprocessed data. Lastly, four different experiments are performed for analyzing customer visit purposes and discussing differences among customer visits.

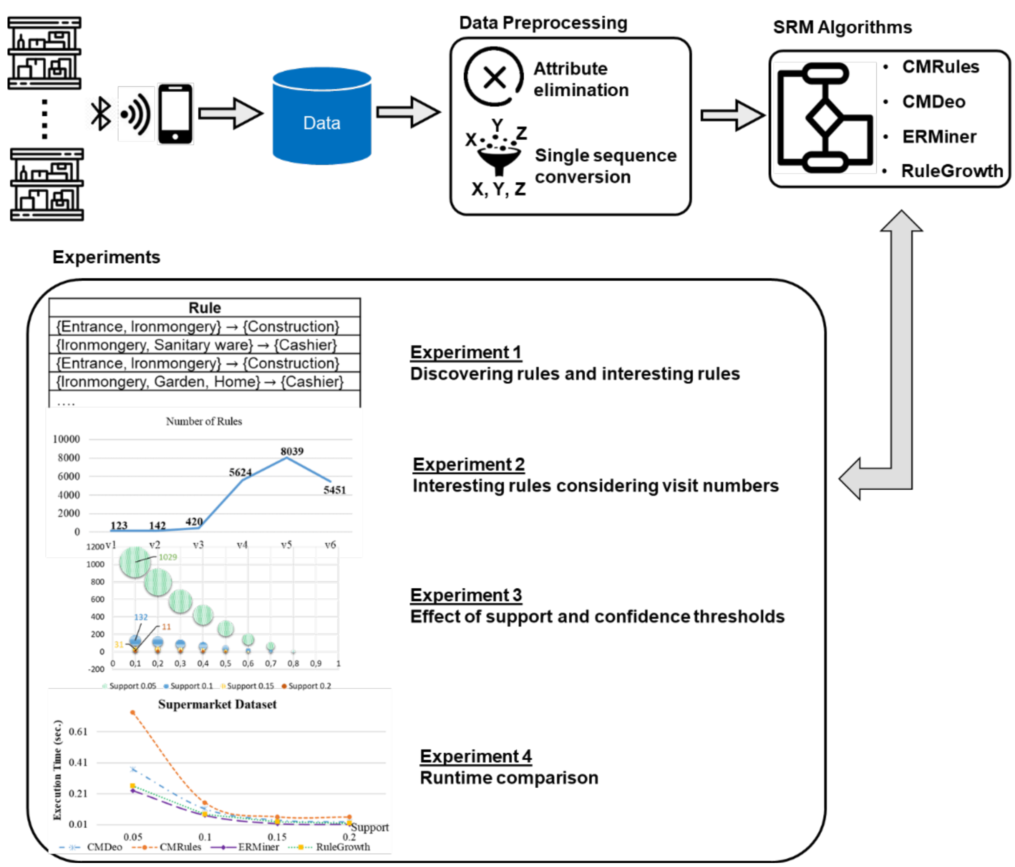

Figure 1. General structure of this study

3.1 Sequential Rule Mining (SRM)

Sequential rule mining is a data mining task that discovers useful hidden patterns from an ordered list of items called a sequence (Vu et al. 2018). Let *D* be a sequence dataset $D = \{S_1, S_2, \dots, S_n\}$ of *n* sequences and $I = \{I_1, I_2, \dots, I_m\}$ be a set of distinct items, where *m* is the number of items. Each sequence *S* consists of an ordered list of items $S = \{X_1, X_2, \dots, X_k\}$ with a unique identifier such that $X \subseteq I$. A sequential rule $X \longrightarrow Y$ presents the relation between X and Y items. The rule is interpreted as if items of X occur in any order of sequence, the items in Y will occur afterward from the same sequence. Items X and Y do not have to occur in the same transaction (itemset) of a sequence.

The most commonly utilized measures to obtain sequential rules from the transaction dataset are Support (*Supp*) and Confidence (*Conf*). The support of a sequential rule $X \longrightarrow Y$ gives the probability of sequences containing the transactions in which all the items of X are followed by all Y items (Yildirim et al. 2017). The sequential rule $X \longrightarrow Y$'s confidence value is evaluated as the support value of the sequential rule divided by the number of sequences that include item X. The support and confidence values are evaluated using the Equation (1) and Equation (2):

$$\boldsymbol{Support(X \longrightarrow Y) = \frac{freq(X \cup Y)}{N}} \tag{1}$$

$$Confidence(\mathrm{X} \rightarrow \mathrm{Y}) = \frac{freq(X \cup Y)}{freq(X)} \tag{2}$$

where *N* is the number of sequences, $freq(X \cup Y)$ and $freq(X)$ refers to the number of occurrences of the related item sets.

Table 1 illustrates an example sequential database that contains the supermarket customers' routes in their visits at different times. For example, the first sequence (ID 1) infers that a customer visits the Entrance firstly, then visits Ironmongery, Lightning and Cashier aisles, respectively, in one of the visits.

Table 1. An example customer route sequential database

ID	Sequences
1	{Entrance},{Ironmongery},{Lighting},{Cashier}
2	{Entrance},{Ironmongery},{Cashier}
3	{Entrance},{Kitchen},{Decoration},{Exit Non Purchasing}
4	{Entrance},{Exit Non Purchasing}

The sequential rules are extracted from this sample dataset by applying an SRM algorithm. The SRM algorithm takes minimum support and minimum confidence threshold values from the user to output all rules having no less than these threshold values. Considering the example dataset, minimum support and minimum confidence values were given as 0.2 by the user. According to these user-defined values, some of the obtained sequential rules were presented in Table 2. For example, the second rule {Entrance} → {Ironmongery, Lighting} has a 0.25 support and confidence value. The 0.25 value of support means that the customer visits Ironmongery and Lighting aisles after Entrance in only one out of the four sequences (ID 1 in Table 1) by purchasing at least one item from these aisles because Cashier indicates a purchased shopping visit. Also, 0.25 confidence value refers that if customers visit Entrance, they are likely to visit Ironmongery and Lighting with a confidence of 0.25 before purchasing an item.

Table 2. The obtained sequential rules

ID	Rule	Supp	Conf
1	{Entrance} → {Ironmongery}	0.5	0.5
2	{Entrance} → {Ironmongery, Lighting}	0.25	0.25
3	{Ironmongery} → {Lighting}	0.25	0.5
4	{Kitchen} → {Decoration}	0.25	0.25
…	…	…	…

This study considers CMRules, CMDeo, ERMiner, and RuleGrowth SPM algorithms, which are commonly preferred and very efficient, on a supermarket visit dataset to analyze customer visits at different times.

- **CMRules:** CMRules is one of the most popular SRM algorithms that first discover association rules from the dataset to prune the search space (Fournier-Viger et al. 2012). Then, it applies minimum support and minimum confidence threshold values by considering time ordering between items in transactions.
- **CMDeo:** CMDeo algorithm evaluates the support values of single items in the transactions (Fournier-Viger et al. 2012). Then, the candidate rules are generated using each pair of frequent items X and Y. The sequential support and confidence values are calculated as the next step, and the algorithm extracts sequential rules in a level-wise manner.
- **RuleGrowth:** RuleGrowth is an efficient SRM algorithm that applies a pattern-growth approach to discover sequential rules more scalable (Fournier-Viger et al. 2011). According to this approach, the rules between X and Y items are obtained, and then the sequential database is scanned to expand the left and right parts of the rules recursively.
- **ERMiner:** ERMiner algorithm uses the Sparse Count Matrix structure to eliminate the search space (Fournier-Viger et al. 2014). Furthermore, it uses equivalence rules' classes that have the same antecedent.

4. Experimental Study

In the experimental studies, the SRM methods were executed on real-world customer data for analyzing customer visits at different times. CMRules, CMDeo, ERMiner, and RuleGrowth algorithms were applied to discover interesting association rules in different visits and compared with each other in terms of their execution times. The implemented SRM approach was developed using the Sequential Pattern Mining Framework, an open-source data mining library in the Java programming language (Fournier-Viger et al. 2016). The experiments were executed on a personal computer with an Intel Core i7-7500U 2.90-GHz processor and 8 GB of memory.

4.1 Dataset Description

The supermarket dataset consists of customers' ID, visit information (visit number), the aisle where the customer is located, and the time of appearing and disappearing in the aisle. The location information of the customers was obtained using iBeacon devices in the aisles of the supermarket. As special locations, Cashier and Not-Purchasing show whether customers purchased (when they were seen in Cashier) or not (when they were seen in Not-Purchasing). Table 3 shows a small part of the experimental dataset including 1100 instances.

Table 3. Example instances from the supermarket dataset

Customer ID	Visit Number	Aisle	Entry Time	Exit Time
1462	v1	Entrance	18/11/2018 18:45	18/11/2018 18:46
1462	v1	Cashier	18/11/2018 18:46	18/11/2018 19:01
1462	v1	Entrance	18/11/2018 19:02	18/11/2018 19:02
1462	v4	Entrance	23/12/2018 12:49	23/12/2018 12:50
1462	v4	Cashier	23/12/2018 12:50	23/12/2018 12:51
1462	v4	Ironmongery	23/12/2018 12:52	23/12/2018 12:54
10095	v1	Entrance	30/11/2018 20:20	30/11/2018 20:21
10095	v1	Lighting	30/11/2018 20:21	30/11/2018 20:23
10095	v1	Sanitary ware	30/11/2018 20:23	30/11/2018 20:24
10095	v1	Cashier	30/11/2018 20:24	30/11/2018 20:26
…	…	…	…	…

The dataset was passed through data preprocessing steps to be ready for the input demands of the SRM algorithms. First, entrance and exit time attributes were eliminated because they were used to obtain visit numbers and aisle attributes. Then, the aisles that each customer visits at different times are combined in a single sequence separately, similar to Table 1.

4.2 Results and Discussion

Four different experiments were performed on the dataset described in the previous section to investigate the followings: (i) the generated interesting sequential rule patterns, (ii) the obtained interesting sequential rules by visit numbers, (iii) the relationship between the number of sequential rules and the minimum support and confidence values, and (iv) the execution time performances of the CMRules, CMDeo, ERMiner, and RuleGrowth algorithms. Although the SRM algorithms applied in the study, they extracted the same sequential rules from the experimental dataset but at different execution times and memory usages. Because of this reason, the CMRules algorithm was used for the analysis in the first three experiments.

In the first experiment, the interesting sequential rules were discovered from the supermarket dataset. This study describes a rule as an interesting rule if a sequential rule has lower support and higher confidence values (Dogan et al. 2018). Table 4 presents some examples of the generated interesting sequential rules selecting minimum support and confidence values as 0.05 and 0.7, respectively, in all experiments. The rule of {Entrance, Sanity ware, Home} → {Cashier, Construction} indicates that customers visited Entrance, Sanity ware, and Home aisles before Cashier and Construction in 6% of all visits. Also, this rule indicates that all customers who visited Entrance, Sanity ware, and Home aisles will likely visit Construction and Cashier aisles with a confidence of 0.81. In other words, although customers rarely visit Entrance, Sanity ware, and Home aisles before Construction and Cashier, when customers visit these aisles, they will also visit Construction and Cashier with a high probability.

Table 4. Examples of the generated interesting sequential rules from the supermarket dataset

ID	Rule	Supp	Conf
1	{Construction, Kitchen} → {Cashier}	0.08	0.73
2	{Sanitary ware, Garden} → {Cashier}	0.1	0.74
3	{Kitchen} → {Construction}	0.11	0.89
4	{Ironmongery, Lighting, Garden} → {Construction}	0.07	0.95
5	{Kitchen} → {Ironmongery}	0.07	0.72
6	{Entrance, Sanity ware, Home} → {Construction, Cashier}	0.06	0.81
7	{Construction, Lighting, Home} → {Garden}	0.07	0.8
…	…	…	…

In the second experiment, the supermarket dataset was divided into six groups according to customers' visit numbers, such as first visit (v1), second visit (v2), …, and sixth visit (v6). Then, the CMRules algorithm was applied to each dataset separately for the extraction of interesting sequential rules considering visit numbers. Figure 2 depicts the number of discovered interesting sequential rules from each dataset. The highest number of rules (8039) was obtained from customers' 5th visits to the same supermarket. Besides, it is seen from this graph that as the number of visits by customers increases, a higher number of rules are obtained.

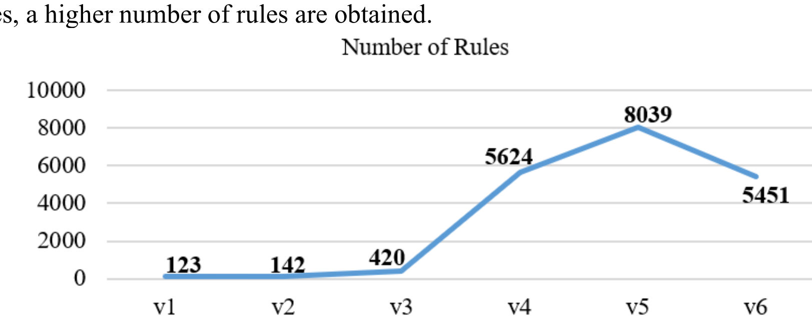

Figure 2. Number of interesting sequential rules from each visit groups

Table 5 gives some examples of the generated interesting sequential rules from the second experiment. For example, the rule {Lighting, Decoration} → {Cashier} states that the customers visited Lighting and Decoration aisles, and then they purchased with a rate of 6% of their 4th visits. This rule also investigates that all customers who visited Lighting and Decoration aisles will absolutely visit Cashier. Furthermore, when the rules ID 1 and ID 3 are examined, it is clearly observed that the customers visit the Construction aisle more after they visited the Entrance and Ironmongery aisles in their second visit to the supermarket. In addition to these, customers mainly purchased because the numbers of rules ending with Exit_Non-Purchasing and Cashier are 655 and 7273, respectively.

Table 5. Examples of the obtained interesting sequential rules by visit numbers

ID	Visit Number	Rule	Supp	Conf
1	v1	{Entrance, Ironmongery} → {Construction}	0.1	0.91
2	v1	{Ironmongery, Sanitary ware} → {Cashier}	0.09	0.72
3	v2	{Entrance, Ironmongery} → {Construction}	0.11	0.94
4	v2	{Ironmongery, Garden, Home} → {Cashier}	0.06	0.84
5	v3	{Kitchen} → {Cashier, Sanitary ware}	0.06	0.83
6	v4	{Lighting, Decoration} → {Cashier}	0.06	1
7	v4	{Entrance, Garden, Decoration} → {Exit_Non_Purchasing}	0.06	1
8	v5	{Kitchen} → {Garden, Exit_Non_Purchasing}	0.18	1
9	v6	{Entrance, Ironmongery, Lighting, Sanitary ware} → {Cashier, Construction, Home, Kitchen}	0.07	1
…	…	…	…	…

Furthermore, the obtained same interesting sequential rules from the different visit numbers were analyzed. Table 6 gives some information about these rules with their visit number and support values. For example, while the rule {Entrance, Ironmongery} →{Construction} has a 0.10 support value in the first visit, it presents a support value of 0.11 in the second visit. It implies that customers visit Entrance, Ironmongery, and Construction aisles more in their second visits. Also, the rule {Construction, Home, Kitchen} → {Cashier} signifies that the customers who make purchases after visiting Construction, Home, and Kitchen aisles, do not revisit the same aisles in the second visits. Finally, the obtained rule {Construction, Garden, Kitchen} → {Cashier} is interpreted as the customers make purchasing after they visit Construction, Garden, and Kitchen aisles in their second visit, whereas this rule did not appear in the first visits as an interesting rule.

Table 6. Examples of the obtained same interesting sequential rules from the different visit numbers

Visit Number	Rule	Supp
v1 v2	{Entrance, Ironmongery} →{Construction} {Entrance, Ironmongery} →{Construction}	0.10 0.11
v3 v4	{Entrance, Home} → {Construction} {Entrance, Home} → {Construction}	0.09 0.2
v1 v2	{Construction, Home, Kitchen} → {Cashier} -	0.05
v1 v2	{Entrance, Ironmongery, Construction, Sanitary ware} → {Cashier} -	0.06
v1 v2	- {Construction, Garden, Kitchen} → {Cashier}	 0.05
…	…	…

In the third experiment, the CMRules algorithm was executed with changing support and confidence values. Figure 3 presents the generated sequential rules number according to varied support and confidence values. The bigger bubble area indicates the larger number of sequential rules. The results indicate that when the minimum support and minimum confidence values decrease, the number of generated sequential rules increases almost exponentially. Rule numbers in Figure 3 are given for 0.05 confidence value as an example to clarify this point. Furthermore, there is no interesting rule discovered when the minimum support value is selected higher than 0.2. Therefore, the support value affects the number of obtained sequential rules substantially.

In the last experiment, the execution time comparisons of the applied SRM algorithms were performed. Figure 4 gives the results obtained from each algorithm with a constant 0.1 minimum confidence value and varying minimum support values from 0.05 to 0.2 in increments of 0.05. This graph shows that the ERMiner algorithm provides the best execution time performance among the other applied algorithms for all support thresholds from 0.05 to 0.2. Also, it is possible to say that the RuleGrowth algorithm presents a closer execution time to the ERMiner algorithm. However, EMRules ans CMDeo algorithms have a gap for a support value of 0.05.

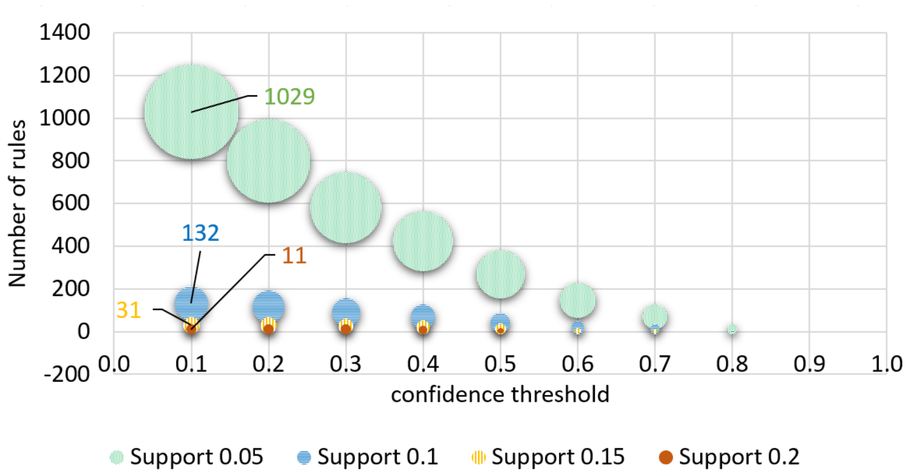

Figure 3. Number of sequential rules with different support and confidence thresholds

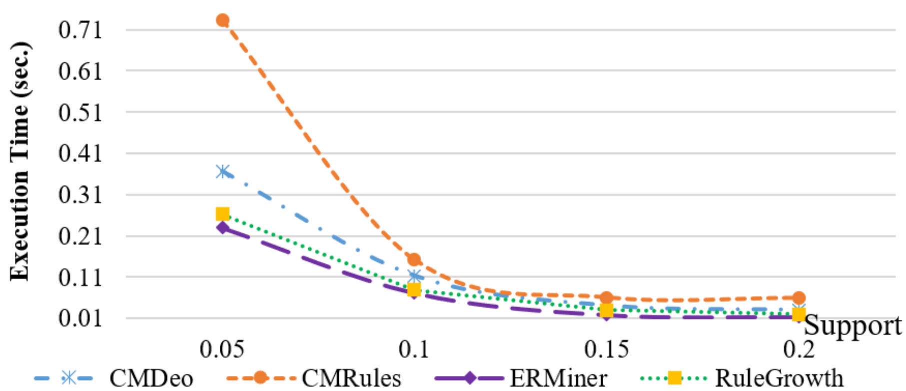

Figure 4. Execution times of the applied SRM algorithms with 0.1 minimum confidence value

6. Conclusion

Analyzing customers' visited aisles in a supermarket plays an important role in discovering customer shopping behaviors and understanding their visit purposes. This study applies some SRM algorithms (CMRules, CMDeo, ERMiner, and RuleGrowth) to real-world customer location data obtained using iBeacon devices in supermarket aisles for investigating customer visit purposes at different times. Four different experiments are carried out within the scope of the study.

Experiment 1 and Experiment 2 indicated that as the number of visits increases, the number of interesting rules increases, in general. This shows that customers prefer this supermarket not only for their specific needs but also for all their needs at every visit. Otherwise, it would be concluded that there are few interesting rules and customers only visit the supermarket for certain needs. The highest number of rules (8039) was obtained from customers' 5th visits to the same supermarket. Furthermore, when the obtained same interesting sequential rules from the different visit numbers are analyzed, it is seen that the customers visit some aisles more in their next visit. Also, the results specify that some customers can visit the aisles, which are not visited in their first, at their next visit and make purchases. Besides, the CMRules algorithm was executed with changing support and confidence values in Experiment 3. The results presented that when the minimum support and minimum confidence values decrease, the number of obtained sequential rules increases exponentially. In addition to these, Experiment 4 concluded when the applied SRM algorithms are compared with each other in terms of execution times they provide within the experiments, the results state that the ERMiner runs faster than the other three algorithms.

In most cases, determining the minimum support value is problematic because it is generally selected by the trial-and-error method. The top-k sequential rules can be discovered from the same supermarket dataset as future work to handle this challenge. Furthermore, to avoid redundant rules, which are a variation of other rules with the same support and confidence values, the top-k non-redundant sequential rules method can be used for this study.

References

Adapa, S., Fazal-e-Hasan, S.M., Makam, S.B., Azeem, M.M., and Mortimer, G., Examining the antecedents and consequences of perceived shopping value through smart retail technology, *Journal of Retailing and Consumer Services*, vol. 52, pp. 101901, 2020.

Agrawal, R., and Ramakrishnan, S., Mining sequential patterns, *Proceedings of 11th International Conference on Data Engineering,* Taipei, Taiwan, March 6 – 10, pp. 3–14, 1995.

Bonacchi, M., and Perego, P., *Customer Accounting: Creating Value with Customer Analytics*, 1st Edition Springer, 2019.

Chang, H. C., and Tsai, H. P., Group RFM analysis as a novel framework to discover better customer consumption behavior, *Expert Systems with Applications*, vol. 38, no. 12, pp. 14499-14513, 2011.

Daunt, K.L., and Harris, L.C., Consumer showrooming: value co-destruction, *Journal of Retailing and Consumer Services*, vol. 38, pp. 166–176, 2017.

Dogan, O., Bayo-Monton, J. L., Fernandez-Llatas, C., and Oztaysi, B., Analyzing of gender behaviors from paths using process mining: A shopping mall application, *Sensors*, vol. 19 no. 3, pp. 557, 2019a.

Dogan, O., Fernandez-Llatas, C., and Oztaysi, B., Process mining application for analysis of customer's different visits in a shopping mall. *Proceedings of the 2nd International Conference on Intelligent and Fuzzy Systems*, Istanbul, Turkey, July 23 – 25, pp. 151-159, 2019b.

Dogan, O., Gurcan, O. F., Oztaysi, B. and Gokdere, U., Analysis of frequent visitor patterns in a shopping mall. *Proceedings of the Global Joint Conference on Industrial Engineering and Its Application Areas*, June 21 – 22, pp. 217 – 227, 2018.

Dogan, O., Hiziroglu, A., and Seymen, O. F., Segmentation of Retail Consumers with Soft Clustering Approach, *Proceedings of the 2nd International Conference on Intelligent and Fuzzy Systems*, July 21 – 23, pp. 39 – 46, 2020.

Dogan, O., and Oztaysi, B., In-store behavioral analytics technology selection using fuzzy decision making, *Journal of Enterprise Information Management*, vol. 31, no. 4, pp. 612 – 630, 2018.

Ferracuti, N., Norscini, C., Frontoni, E., Gabellini, P., Paolanti, M., and Placidi, V., A business application of RTLS technology in Intelligent Retail Environment: defining the shopper's preferred path and its segmentation, *Journal of Retailing and Consumer Services*, vol. 47, pp. 184–194, 2019.

Fournier-Viger, P., Faghihi, U., Nkambou, R., and Mephu Nguifo, E., CMRules: Mining Sequential Rules Common to Several Sequences, *Knowledge-based Systems*, vol. 25, no.1, pp. 63-76, 2012.

Fournier-Viger, P., Gueniche, T., Zida, S., and Tseng, V. S., ERMiner: Sequential Rule Mining using Equivalence Classes, *Proceedings of the 13th International Symposium on Intelligent Data Analysis*, Leuven, Belgium, October 30 -November 1, pp. 108-119, 2014,.

Fournier-Viger, P., Lin, C.W., Gomariz, A., Gueniche, T., Soltani, A., Deng, Z., and Lam, H.T., The SPMF open-source data mining library version 2. *Proceedings of the 19th European Conference on Principles of Data Mining and Knowledge Discovery*, Riva del Garda, Italy, September 19 – 23, pp. 36–40, 2014.

Fournier-Viger, P., Nkambou, R., and Tseng, V.S., RuleGrowth: Mining Sequential Rules Common to Several Sequences by Pattern-Growth, *Proceedings of the 26th Symposium on Applied Computing*, TaiChung, Taiwan, March 22 – 26, pp. 954–959, 2011.

Fukuzaki, Y., Mochizuki, M., Murao, K., and Nishio, N., Statistical analysis of actual number of pedestrians for WiFi packet-based pedestrian flow sensing, *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*, New York, United States, September 7 – 11, pp. 1519-1526, 2015.

Griva, A., Bardaki, C., Pramatari, K., and Papakiriakopoulos, D., Retail business analytics: Customer visit segmentation using market basket data, *Expert Systems with Applications*, vol. 100, pp. 1-16, 2018.

Guadagni, P. M., and Little, J.D.C., A logit model of brand choice, calibrated on scanner data, *Marketing Science*, vol. 2, pp. 203–238, 1983.

Gupta, S., Impact of sales promotions on when, what, and how much to buy, *Journal of Marketing Research*, vol. 25, pp. 342–355, 1988.

Herhausen, D., Kleinlercher, K., Verhoef, P.C., Emrich, O., and Rudolph, T., Loyalty formation for different customer journey segments, *Journal of Retailing*, vol. 95, no. 3, pp. 9–29, 2019.

Inman, J.J., and Nikolova, H., Shopper-facing retail technology: a retailer adoption decision framework incorporating shopper attitudes and privacy concerns *Journal of Retailing*, vol. 93, no. 1, pp. 7–28, 2017.

Khajvand, M., Zolfaghar, K., Ashoori, S., and Alizadeh, S., Estimating customer lifetime value based on RFM analysis of customer purchase behavior: Case study, *Procedia Computer Science*, vol. 3, pp. 57-63, 2011.

Kitchens, B., Dobolyi, D., Li, J., and Abbasi, A., Advanced customer analytics: strategic value through integration of relationship-oriented big data, *Journal of Management Information Systems*, vol. 35, no. 2, pp. 540–574, 2018.

Landmark, A. D., and Sjøbakk, B., Tracking customer behaviour in fashion retail using RFID, *International journal of retail & distribution management*, vol. 45, vo. 7/8, pp. 844-858, 2017.

Larson, J. S., Bradlow, E. T., and Fader, P. S., An exploratory look at supermarket shopping paths, *International Journal of research in Marketing*, vol. 22, no. 4, pp. 395-414, 2005.

Ma'arif, M.R., Revealing daily human activity pattern using process mining approach, *Proceedings of 4th International Conference on Electrical Engineering, Computer Science and Informatics*, Yogyakarta, Indonesia, September 19 – 21, pp. 1-5, 2017.

Mabroukeh, N.R., and Ezeife, C.I., A taxonomy of sequential pattern mining algorithms. *ACM Computing Surveys* vol. 43, no. 1, pp. 1–41, 2010.

Oosterlinck, D., Benoit, D. F., Baecke, P., and Van de Weghe, N., Bluetooth tracking of humans in an indoor environment: An application to shopping mall visits, *Applied Geography*, vol. 78, pp. 55-65, 2017.

Sun, N., Morris, J.G., Xu, J., Zhu, X., and Xie, M., iCARE: a framework for big databased banking customer analytics, *IBM Journal of Research and Development*, vol. 58, no. 5/6, pp. 1-4, 2014.

Wedel, M., and Kannan, P., Marketing analytics for data-rich environments, *Journal of Marketing*, vol. 80, no. 6, pp. 97–121, 2016.

Willems, K., Smolders, A., Brengman, M., Luyten, K., and Schöning, J., The path-to-purchase is paved with digital opportunities: An inventory of shopper-oriented retail technologies, *Technological Forecasting and Social Change*, vol. 124, pp. 228-242, 2017.

Wu, I.C., and Yu, H.K., Sequential analysis and clustering to investigate users' online shopping behaviors based on need-states, *Information Processing & Management*, vol. 57, no. 6, pp. 102323, 2020.

Yada, K., String analysis technique for shopping path in a supermarket, *Journal of Intelligent Information Systems*, vol. 36, no. 3, pp. 385-402, 2011.

Yewatkar, A., Inamdar, F., Singh, R., and Bandal, A., Smart cart with automatic billing, product information, product recommendation using rfid & zigbee with anti-theft, *Procedia Computer Science*, vol. 79, pp. 793-800, 2016.

Yildirim, P., Birant, D., and Alpyildiz, T., Discovering the relationships between yarn and fabric properties using association rule mining, *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 25, pp. 4788-4804, 2017.

Biographies

Pelin Yildirim Taser is an assistant professor in Computer Engineering Department of İzmir Bakırçay University. She received her B.Sc, M.Sc., and Ph.D. degrees in Computer Engineering from Dokuz Eylül University in 2013, 2015, and 2019, respectively. After her B.Sc. graduation, she worked as a project assistant at the Computer Engineering Department of Dokuz Eylül University for six months. Between 2014 – 2019, she worked as a research assistant at the Software Engineering Department of Manisa Celal Bayar University, Turkey. Her research interests include data mining, machine learning, deep learning, and intelligent systems. She referees for the assessment of the industrial R&D project proposals. Dr. Yildirim Taser has graduation awards, including certificates of high achievement from her M.S. and Ph.D. educations and certificate of publication honor from her Ph.D.

Onur Dogan is an assistant professor at Izmir Bakircay University. He graduated from the Sakarya University with a Bachelor's Degree in Industrial Engineering in 2010 and received the Master's Degree and Doctor of Philosophy in Industrial Engineering from the Istanbul Technical University in 2013 and 2019, respectively. He studied on intelligent decision support system, lean manufacturing and quality approaches such as QFD, FMEA or DOE during the master thesis. Process mining is the primary focus his Ph.D. study. He is one of the first academicians who have a Ph.D. title on process mining in Turkey. During his doctorate study, he had been at Universitat Politecnica de Valencia as a visitor researcher for a 12-month research project. He is one of the members of a newly established research center (Research Center for Data Analytics and Spatial Data Modeling). Dr. Dogan is currently project coordinator at one national projects on process mining and researcher at two projects on smart campus analytics and big data analytics.

His projects are supported by Turkish Scientific and Research Council (TUBITAK) and Izmir Development Agency. His current research areas are process mining, machine learning, (fuzzy) data mining, fuzzy decision making.