

Discovering Customer Paths from Location Data with Process Mining

Onur Dogan

Izmir Bakircay University

Abstract

Customer paths can be used for several purposes, such as understanding customer needs, defining bottlenecks, improving system performance. Two of the principal difficulties depend on discovering customer paths due to dynamic human behaviors and collecting reliable tracking data. Although machine learning methods have contributed to individual tracking, they have complex iterations and problems to produce understandable visual results. Process mining is a methodology that can rapidly create process flows and graphical representations. In this study, customer flows are created with process mining in a supermarket. The differences between the paths of customers purchased and non-purchased are discussed. The results show that both groups have almost similar visit duration, which is 87.5 minutes for purchased customers and 86.6 minutes for non-purchased customers. However, the duration of aisles is relatively small in non-purchased customer flows because customers aim to return or change the item instead of buying.

Index Terms—customer analytics, indoor locations, process mining, supermarket

I. INTRODUCTION

The visited places in an indoor location include many uncovered information such as customer behaviors, queue in the entrance or exit, and system bottlenecks. First, it is a necessity to collect the visited location data of customers to extract meaningful information. There are several ways of data collection for human tracking [1]. Wireless-based technologies such as WiFi, Bluetooth, and RFID provide non-manipulated location data. The gathered customer data can be used for both descriptive statistics to summarize customer analytics and discovering customers' followed paths to explain general overview [2]. Creating the customer path from location data is possible for various ways. Pattern recognition and

Bayesian network models, which are machine learning methods, enable to improve path discovery models [3]. However, these methods can formulate complicated mathematical rules that detect and analyze the user path. Hence, these models are mainly not understandable [4], [5]. The obvious and personalized paths provide decision-makers to examine customer behavior properly and detect undesired situations such as bottlenecks and queues [6].

Process mining is an intersection domain between process management and data science [7]. It uses an event log as input, which is shown in Table I as a sample. An event log must include at least three essential characteristics, which are Case ID, Timestamp, Activity. In this study, Case ID is shown with Subscriber_ID. The timestamp is divided into two parts as Start and End by data preparation steps. Activity is thought visited aisle. Process mining gives readable process flows by presenting important customer behaviors in the process using prepared event logs [8]. Workflows are mainly created to show process models. They easily represent processes by identifying behavioral differences and creating valuable insights [6], [9].

Table I: A SAMPLE OF EVENT LOG

| Subscriber_ID | Description | Start | End |
|---------------|--------------|---------------------|---------------------|
| 1462_v1 | Entrance | 18.11.2018 18:45:53 | 18.11.2018 18:46:05 |
| 1462_v1 | Cashier | 18.11.2018 18:46:06 | 18.11.2018 19:01:40 |
| 1462_v1 | Home | 18.11.2018 19:02:39 | 18.11.2018 19:02:59 |
| 1462_v3 | Entrance | 22.12.2018 16:59:24 | 22.12.2018 17:00:06 |
| 1462_v3 | Ironmongery | 22.12.2018 17:00:07 | 22.12.2018 17:01:31 |
| 1462_v3 | Construction | 22.12.2018 17:01:38 | 22.12.2018 17:02:28 |
| 1462_v3 | Cashier | 22.12.2018 17:02:29 | 22.12.2018 17:03:08 |

In the literature, customer pathway investigations usually generate a graphical illustration to explain the paths. However, the produced paths are intuitive without any detailed study [10]. Process mining reveals the core meaning of the followed paths. Accordingly, process mining is applied to find pathways of shoppers in a supermarket to learn their behaviors. The most crucial step is to collect customer data. Various data collection technologies gather much more location data to analyze customer analytics.

Since the actual time of customer movements cannot be recorded in traditional ways, the technological improvements are adapted to non-invasive data collection [11]. Willeims et al. [12] adopted WiFi to compile a retail inventory technology, which is categorized according to the type of shopping value and the stage of the shopping cycle. Also, Carrera et al. [13] applied real-time human tracking using WiFi technology. Fukuzaki et al. [14] attempt to determine the real number of customers in the shopping mall with WiFi technology. Oosterlinck et al. [15] established Bluetooth based scanners devices in a shopping mall. They collected data in high quality at low cost. Yewatkar et al. [16] introduced an intelligent shopping cart, which records purchased products and online transaction with RFID and ZigBee.

In various researches, human movements are detected by applying process mining by considering personal paths similar to a business process. Maarif [17] adopted process mining to display people daily activities in the visual description. Dogan et al. [2] used process mining to discover and explain the main behavioral changes in male and female paths in a shopping mall. The influence of workload on service times are investigated by recognizing the interaction between them with process mining [18]. Besides, it is implemented to examine the human movements in rooms for nine people. Data were collected with RFID technology and prepared for process mining [4].

Former studies analyzed individual activity recognition [19] and visualization [3], [17], [20]. However, none of them produced a readable and understandable display at a personal level. In the research, process mining procedures are employed to create shopping behaviors of customers. The suggested method is originated from the grammatical inference pattern recognition [21] and explains followed paths as timed parallel automaton, as experimented by Fernandez-Llatas et al. [22].

II. CASE STUDY

A. Data Preparation

Because the raw data are not suitable to apply process mining, some data preparation iterations are applied. For example, each visit made by the same customer refers to a new case, Subscriber_ID is revised, including visit numbers. For instance, 1462_v1 gives the details for the first visit of the customer 1462. Almost all data collection systems record only one timestamp, which is mainly the start time. Since a customer cannot be in more than one location at the same time, the end time can be easily calculated. Aisles are combined according to the most dominant product to decrease the number of considered locations. The start and end time were limited between 08:00:00 and 22:00:00 because of the working time of the supermarket. The visit duration for each aisle is determined as at least fifteen seconds to ignore walking people instead of shoppings.

B. Descriptive Statistics

The customer location data were gathered via iBeacons devices Bluetooth-based technology in a supermarket from August 2018 to March 2019. iBeacons have a 50-meter coverage area and transmit radio signals every 3 seconds. Although one corridor includes more than one product group, aisles are named according to the most dominant product in the shelves. Raw data, which belong to 1727 unique customers, consisted of near to two thousand rows at the beginning of the study. After data preparation steps, 1274 cases, including 11000 aisle visits, were obtained from 784 unique customers. The average duration is 31.3 minutes per case.

Fig. 1 shows the average visit duration for the top nine cases, which consists of 80% of total visits with respect to the number of different visits.

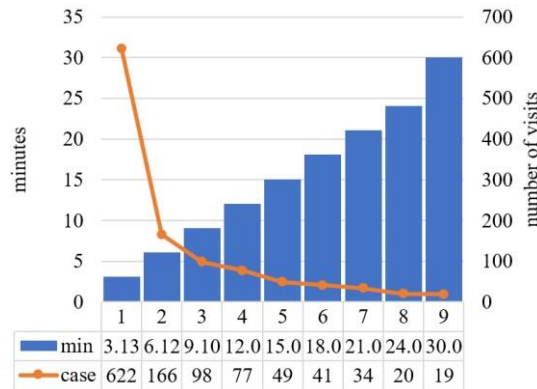


Fig. 1. Mean visit duration.

Six hundred twenty-two (622) out of 1274 visits are ended up before 3 minutes. The customers who spend less than 3.13 minutes visit only one aisle. One hundred sixty-six (166) customers, who visit at least two aisles, left the supermarket before 6.12 minutes.

Table II gives aisles statistics. Typically, the number of appearances in Entrance should be greater than or equal to other aisles. However, the data were collected by smart devices working with Bluetooth; connections sometimes can be cut due to personnel or technical issues. Therefore, the number of being active in Entrance is less than being active in Construction in the case study.

TABLE II: VISITED AISLES STATISTICS

| AISLES | Frequency | Relative frequency | Mean duration (min) |
|---------------------|-----------|--------------------|---------------------|
| Construction | 2384 | 21.67% | 5.82 |
| Entrance | 1710 | 15.55% | 10.47 |
| Cashier | 1420 | 12.91% | 7.25 |
| Garden | 1320 | 12.00% | 7.70 |
| Home | 992 | 9.02% | 4.78 |
| Sanitary ware | 827 | 7.52% | 3.32 |
| Ironmongery | 698 | 6.35% | 3.87 |
| Exit Non-purchasing | 584 | 5.31% | 10.40 |
| Lighting | 566 | 5.15% | 2.67 |
| Kitchen | 330 | 3.00% | 2.48 |
| Decoration | 169 | 1.54% | 4.35 |

One out of five shoppers visits Construction aisle with a mean duration of 5.82 minutes, whereas 15% of shoppers appeared in Entrance aisle. On the other hand, the spent times in these aisles has an inverse ratio. The mean duration in Entrance is almost two times higher than in Construction aisle.

C. Process Mining Results

Fig. 2 shows the purchased customer flows starting from Entrance. The typical property of this flow is the end with Cashier, which indicates customers wait to pay for their purchased items. The thickness of the arrows increases when the number of transitions between two nodes increases. Similarly, the darkness of the colors in the nodes increase when the number of executed activities increases.

The number on the arrows presents walking time between two aisles. For example, customers who purchase an item spend approximately 17 minutes on average, then walk about 7.5 minutes to exit. The duration in Cashier is 18.4 minutes. This is because aisles are combined according to the most dominant product. That means Cashier aisle includes some item groups or services such as customer services, snacks, and promotion items. Hence, Cashier duration does not mean the time during the payment queue.

On the other hand, Fig. 3 depicts non-purchased customer flows starting from Entrance. The typical property of this flow is the end with Exit Non-purchasing, which indicates customers do not pay. It may be because customers visit the supermarket for returns and change. Due to the supermarket layout, entering customer must walk around. Therefore, the duration shown in the rectangles is relatively small. After Entrance, customers follow several directions such as Exit without purchase, Garden, and Home. Purchased and non-purchased customers have almost similar visit duration, which are 87.5 minutes and 86.6 minutes, respectively. Interestingly, four customers return the supermarket after exit.

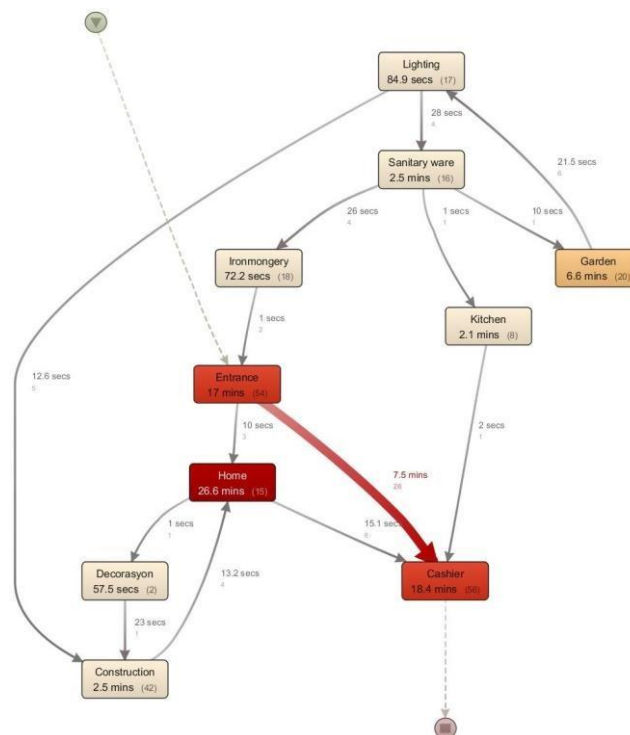


Fig. 2. Purchased Customer Flows.

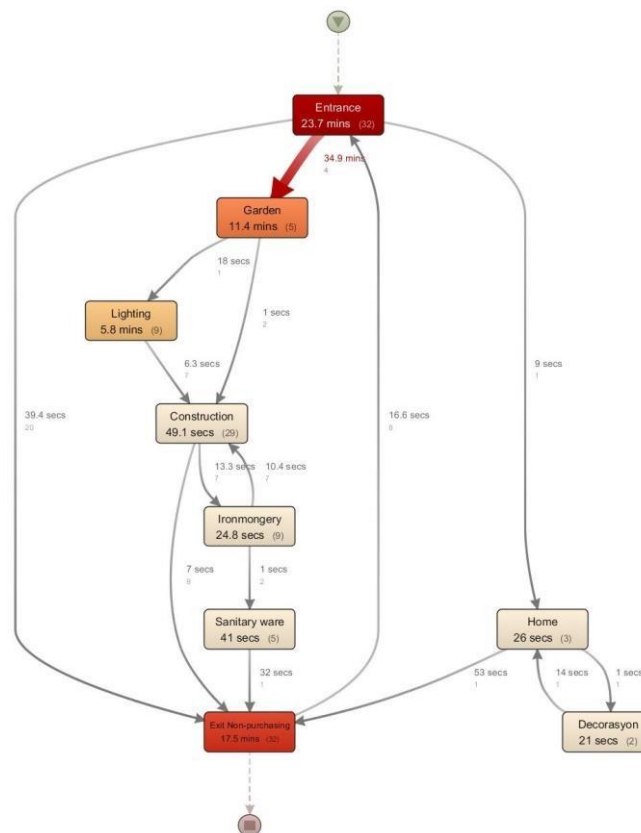


Fig. 3. Non-Purchased Customer Flows.

III. DISCUSSIONS AND CONCLUSIONS

In the investigation, a process mining implementation is presented to discover shopper paths in a supermarket. The proposed methodology has been examined with 784 unique customers using real-time tracking data from August 2018 to March 2019. The real data were gathered with iBeacon devices, which have Bluetooth-based technology. Bluetooth is a cost-effective technology that gives unbiased and impartial views.

According to the research conclusions, process mining application with an indoor location system (Bluetooth tracking) facilitates to create a clear view. The results are evaluated concerning descriptive statistics and process mining. Whereas the results of descriptive statistics is data-oriented, process mining provides process-based solutions. Because consumers visited the supermarket more than once, 1274 paths are created for 784 customers.

Because the spent time for the purchased and non-purchased customer is almost similar, developing a recommendation system may increase purchased customer rate. Therefore, we will focus on developing a personalized recommendation system in future work. Moreover, clustering followed paths may improve to understand visit purposes of shoppers. At the same time, path clustering may support the recommendation system. Hence, one of the next studies may be path clustering considering sequences of visited aisles and spent time.

REFERENCES

- [1] O. Dogan, and B. Oztaysi, "In-store behavioral analytics technology selection using fuzzy decision making," *Journal of Enterprise Information Management*, vol. 31, no.4, pp. 612-630, July 2018.
- [2] O. Dogan, J. L. Bayo-Monton, C. Fernandez-Llatas, and B. Oztaysi, "Analyzing of Gender Behaviors from Paths Using Process Mining: A Shopping Mall Application," *Sensors*, vol. 19, no. 3, pp. 557, Jan. 2019.
- [3] O. Dogan, et al., "Individual Behavior Modeling with Sensors Using Process Mining," *Electronics* vol. 8, no. 7 pp. 766, July 2019.
- [4] C. Fernandez-Llatas, J. M. Benedi, J. M. Garcia-Gomez, and V. Traver, "Process mining for individualized behavior modeling using wireless tracking in nursing homes," *Sensors* vol. 13, no. 11, pp. 15434-15451, Nov. 2013.
- [5] A. Martinez-Millana, A. Lizondo, R. Gatta, S. Vera, V. Traver, and C. Fernandez-Llatas, "Process Mining Dashboard in Operating Rooms: Analysis of Staff Expectations with Analytic Hierarchy Process," *International Journal of Environmental Research and Public Health*, vol. 16, no. 2, pp. 199, Jan. 2019.
- [6] C. Fernandez-Llatas, A. Lizondo, E. Monton, J. M. Benedi, and V. Traver, "Process mining methodology for health process tracking using real-time indoor location systems," *Sensors*, vol. 15, no. 12, pp. 29821-29840, Nov. 2015.
- [7] O. Dogan, "Process mining for check-up process analysis," *IIOABJ*, vol. 9, no. 6, pp. 56-61, Sept. 2018.
- [8] W. M. P. van der Aalst, *Process Mining: Data Science in Action*, 2nd ed. Springer, Berlin, Heidelberg, 2016, ch. 1, pp. 3-23.
- [9] W. M. P. van der Aalst, et al., "Process mining manifesto", *International Conference on Business Process Management*, pp. 169-194, 2011.
- [10] I. Hwang, and Y. J. Jang, "Process mining to discover shoppers' pathways at a fashion retail store using a WiFi-base indoor positioning system," *IEEE Transactions on Automation Science and Engineering*, vol. 14, no. 4, pp. 1786-1792, Oct. 2017.
- [11] O. Dogan, O. F. Gurcan, B. Oztaysi, and U. Gokdere, "Analysis of Frequent Visitor Patterns in a Shopping Mall," *Industrial Engineering in the Big Data Era*, pp. 217-227, 2019.
- [12] K. Willems, A. Smolders, M. Brengman, K. Luyten, and J. Schöning, "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.

- [13] J. L. Carrera, Z. Li, Z. Zhao, T. Braun, and A. Neto, "A real-time indoor tracking system in smartphones," *Proceedings of the 19th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, pp. 292-301, 2016.
- [14] Y. Fukuzaki, M. Mochizuki, K. Murao, and N. Nishio, "Statistical analysis of actual number of pedestrians for Wi-Fi 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*, pp.1519-1526, 2015.
- [15] D. Oosterlinck, D. F. Benoit, P. Baecke, and N. van de Weghe, "Bluetooth tracking of humans in an indoor environment: An application to shopping mall visits," *Applied geography* vol. 78, pp. 55-65, 2017.
- [16] A. Yewatkar, F. Inamdar, R. Singh, and A. Bandal, "Smart cart with automatic billing, product information, product recommendation using rfid & zigbee with anti-theft," *Procedia Computer Science* vol. 79, pp. 793-800, 2016.
- [17] M. R. Ma'arif, "Revealing daily human activity pattern using process mining approach," *4th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, pp. 1-5, 2017.
- [18] J. Nakatumba, and W. M. P. van der Aalst. "Analyzing resource behavior using process mining," *International Conference on Business Process Management*, pp. 69-80, 2009.
- [19] E. Kim, S. Helal, and D. Cook, "Human activity recognition and pattern discovery," *IEEE Pervasive Computing*, vol. 9, no. 1, pp. 48-53, Dec. 2009.
- [20] O. Dogan, C. Fernandez-Llatas, and B. Oztaysi, "Process Mining Application for Analysis of Customer's Different Visits in a Shopping Mall," *Intelligent and Fuzzy Techniques in Big Data Analytics and Decision Making (INFUS 2019)*, ed.s Cengiz Kahraman, Selcuk Cebi, Sezi Cevik Onar, Basar Oztaysi, Ahmet Tolga, and Irem Sari, pp. 151-159, 2019.
- [21] A. Stevenson, and J. R. Cordy, "Grammatical inference in software engineering: an overview of the state of the art," *International Conference on Software Language Engineering*, pp. 204-223, Sept. 2012.
- [22] C. Fernandez-Llatas, T. Meneu, J. M. Benedi, and V. Traver, "Activity-based process mining for clinical pathways computer aided design," *Annual International Conference of the IEEE Engineering in Medicine and Biology*, pp. 6178-6181, August 2010.