Segmentation of indoor customer paths using intuitionistic fuzzy clustering: Process mining visualization

Article *in* Journal of Intelligent & Fuzzy Systems · October 2019

DOI: 10.3233/JIFS-179440

Onur Dogan
İzmir Bakırçay University

Başar Öztayşi
Istanbul Technical University

Carlos Fernandez-Llatas
Polytechnic University of Valencia

Journal of Intelligent & Fuzzy Systems 38 (2020) 675–684
DOI:10.3233/JIFS-179440
IOS Press

Segmentation of indoor customer paths using intuitionistic fuzzy clustering: Process mining visualization

Onur Dogan[a,*], Basar Oztaysi[b] and Carlos Fernandez-Llatas[c]
[a]*Izmir Bakircay University, Department of Industrial Engineering, Gazi Mustafa Kemal Mahallesi, Kaynaklar Caddesi, Izmir, Turkey*
[b]*Istanbul Technical University, Department of Industrial Engineering, Istanbul, Turkey*
[c]*Instituto Universitario de Investigación de Aplicaciones de las Tecnologías de la Información y de las Comunicaciones Avanzadas (ITACA), Universitat Politècnica de València, Camino de Vera S/N, Valencia, Spain*

Abstract. There are some studies and methods in the literature to understand customer needs and behaviors from the path. However, path analysis has a complex structure because the many customers can follow many different paths. Therefore, clustering methods facilitate the analysis of the customer location data to evaluate customer behaviors. Therefore, we aim to understand customer behavior by clustering their paths. We use an intuitionistic fuzzy c-means clustering (IFCM) algorithm for two-dimensional indoor customer data; case durations and the number of visited locations. Customer location data was collected by Bluetooth-based technology devices from one of the major shopping malls in Istanbul. Firstly, we create customer paths from customer location data by using process mining that is a technique that can be used to increase the understandability of the IFCM results. Moreover, we show with this study that fuzzy methods and process mining technique can be used together to analyze customer paths and gives more understandable results. We also present behavioral changes of some customers who have a different visit by inspecting their clustered paths.

Keywords: Fuzzy c-means clustering, intuitionistic fuzzy sets, process mining, customer behaviors, indoor locations

1. Introduction

Understanding and meeting customer needs are two critical activities for both online and offline (physical) stores in a competitive business environment [1, 2]. For these activities, the first thing that should be done is to collect customer data. While this is so easy in online stores by recording each click or cursor movement of the customer, in physical stores, data collection needs some additional preparations such as the installation of various data collection technologies [3]. One of the methods that can be used to understand customer behaviors in physical stores like indoor locations is to analyze their paths [4–7].

Analyzing customer behaviors with respect to the paths is an interesting study area [8]. Additionally, analyzing the followed path, which can be created from customer location data, is a hard problem since the customer paths cannot be structured [2]. Therefore, we aim to make easy the analysis of the customer location data by using an advanced clustering method. At the same time, Mou and colleagues [9] consider employees in the retail industry are under pressure because of different reasons, such as a lot of operations, inadequate professional training, and poor customer-oriented service. More efficient and effective use of resources is necessary for the improvement of the retail system and for strength-

*Corresponding author. Onur Dogan, Izmir Bakircay University, Department of Industrial Engineering, Gazi Mustafa Kemal Mahallesi, Kaynaklar Caddesi, Izmir, 35665, Turkey. E-mail: onur.dogan@bakircay.edu.tr.

ISSN 1064-1246/20/$35.00

ening customer relations. Some of the studies used the basic statistical information to describe customer behaviors [6], but understanding customer behavior requires a global view of customer pathways [10]. For that reason, we used the process mining technique to discover customer paths in a shopping mall. Both the discovery of pathways for a general view and generating understandable results are one of the goals of process mining.

Process mining (PM) aims to extract information from event logs by process models [11]. Event logs in real-life processes are less structured and more flexible [12]. Therefore, conventional PM discovery algorithms can cope with unstructured processes by generating spaghetti process models that are more complex and difficult to interpret. Since structured models represent more qualified process variables, PM usually uses clustering algorithms to produce more understandable results [12–15]. Each cluster corresponds to a well-structured group that includes event logs that have similar features. Bose and van der Aalst [11] showed in their study clustering reduce the spaghettiness and increase the understandability.

Fuzzy clustering is a clustering methodology assigning elements into groups according to their similarities with a membership function. In many real-life cases, finding a relationship among data, which is often not clear, gives advantages to manage data [16]. We create groups to assign data into these groups to effectively divide the information. The fuzzy logic approach can consider the uncertainty in the problems and provides solutions that are more accurate [17]. In fuzzy clustering, one element can be assigned to more than one group, in contrast to traditional clustering. Therefore, fuzzy clustering is called soft clustering. Since fuzzy clustering methods provide that data points can belong to more than one cluster, it increases the accuracy of clustering. At the same time, the fuzzy approach enables to increase to correctly assignment ratio of the data objects [18]. It is especially significant for boundary data, which usually cause wrong clustering. In some cases, the clustering algorithm may assign the boundary data objects into the first cluster. Fuzzy clustering methods can avoid these misclassifications.

Similar to the uncertainty, the hesitation in defining the membership function of data objects arises. Since the membership values are not certain, a hesitation that stems from the lack of precise knowledge in identifying the membership function. This view leads to an initiation of a higher order fuzzy set named intuitionistic fuzzy sets [19]. The intuitionistic fuzzy sets define not only membership values but also non-membership values. Whereas the non-membership value is the complement of the membership value in the ordinary fuzzy sets, it is less than or equal to the complement of the membership degree thanks to the hesitation degree in the intuitionistic fuzzy sets.

In this study, we cluster indoor customer path data using intuitionistic fuzzy c-means clustering (IFCM) algorithm concerning case durations and the number of visited locations. We compare the results of three different algorithms; IFCM, fuzzy c-means (FCM) and K-means (KM). And then, we apply PM to clustered customer paths to show differences among them.

2. Related works

Customer behavior analysis is a popular research area. Museums [6], hospitals [4, 20, 21], shopping mall or stores [2, 5, 22], manufacturing [23], airport [24] and exhibition [25] are some of the implementation areas of customer behavior analysis. Yoshimura et al. [6] aimed to determine some parameters (average number of places visited by people, the average time spent by them, and most/least visited places) with descriptive statistics. Yoshimura et al. [6] and Wu et al. [26] discovered paths followed by customers. Yim et al. [27] estimated the next visiting location of customers. Oosterlinck et al. [28] and Chen et al. [29] determined where the customer is located at any time. Dogan et al. [30] analyzed frequent customer visited stores in a shopping mall to better understand customer behaviors.

On the other hand, recent studies on path clustering show different results. Dogan and Oztaysi [2] predicted gender of customers with levenshtein-based fuzzy kNN method by classifying their indoor paths. Shaw and Gopalan [31] found and clustered frequently followed paths. Marchetti and Zhou [32] developed a clustering methodology introducing the idea of a regularization path. Everman et al. [33] developed a path clustering method based on the local alignment of sequences using k-means clustering. Hompes et al. [15] clustered paths to detect changes in processes and showed behavior changes and their reasons trace clustering. All of these studies have been used a crisp boundary clustering techniques. Few studies focus on fuzzy clustering for customer paths. D'Urso and Massari [34] adopted fuzzy c-medoids clustering for human activities in the shopping, web usage, and tourist travel areas. Table 1

Table 1
Recent works on CBA from the path

Study	CBA	Clustering	Algorithm	Area
[3]	✓			Hospital
[4]	✓			Shopping mall/Store
[5]	✓			Museum
[19]	✓			Hospital
[20]	✓			Hospital
[21]	✓			Shopping mall/Store
[22]	✓			Manufacturing
[23]	✓			Airport
[24]	✓			Exhibition
[25]	✓			Shopping mall/Store
[26]	✓			Shopping mall/Store
[27]	✓			Shopping mall/Store
[28]	✓			Sport
[29]	✓			Shopping mall/Store
[30]		✓	MCL	Hospital and Municipality
[31]		✓	MM	Biology/Gene
[33]	✓	✓	FCMd	Shopping, Web usage, and Tourist travel areas
[34]	✓	✓	K-Means	Tourist travel areas
[35]	✓	✓	SVM	Laboratory

MCL: Markov Cluster Algorithm. MM: Maximization-Minimization Algorithm. FCMd: Fuzzy C-medoids. SVM: Support Vector Machine.

indicates the summary of a literature review about customer behavior analysis (CBA) from the customer paths. Intuitionistic fuzzy c-means approach for customer path clustering is a fundamental aspect of this study. According to Table 1, an IFCM algorithm for customer behaviors and visualization with process mining enrich the literature.

Process mining is relatively new area. Therefore, it is not combine with fuzzy set theory before. There are some studies focusing on customer behavior analysis with process mining [7, 30, 37, 38]. There are some papers that apply various clustering techniques [12–14]. To our knowledge, there has not been any study which combine process mining and fuzzy set theory to analyze the customer behaviors.

3. Preliminaries

Firstly, we give basic concepts of fuzzy sets (FS), introduced by Zadeh [39], since they facilitate to understand intuitionistic fuzzy sets. A fuzzy set is defined as follows:

Definition 1. A Fuzzy Set (FS) μ in a non-empty set X is a function [39].

$$\mu : X \rightarrow [0, 1]\, and X \rightarrow \mu\,(x) \tag{1}$$

where $\mu\,(x)$ represents the membership degree of each element $x \in X$. Besides, a fuzzy set may be defined as

$$\widetilde{A} = \left\{x, \mu_{\widetilde{A}}(x) \mid x \in X\right\},$$

According to Zadeh, in fuzzy sets, one element may belong to more than one set in contrast to crisp sets. Atanassov [19] improved this idea by handling uncertainty. One element may belong to set A (membership) or not belong to set A (non-membership) or hesitance of belonging set A. The idea is expressed in Definition 2.

Definition 2. Let X is a non-empty set. An intuitionistic fuzzy set (IFS) in X is an object A given by

$$\widetilde{A} = \left\{\langle x, \mu_{\widetilde{A}}(x), \eta_{\widetilde{A}}(x)\rangle; x \in X\right\}$$

where $\mu_{\widetilde{A}} : X \rightarrow [0, 1]$ and $\eta_{\widetilde{A}} : X \rightarrow [0, 1]$ satisfy the condition

$$0 \leq \mu_{\widetilde{A}}(x) + \eta_{\widetilde{A}}(x) \leq 1 \tag{2}$$

for every x in X. Hesitancy (or uncertainty) shown by $\pi_{\widetilde{A}}(x)$ is equal to \1 $- \mu_{\widetilde{A}}(x) - \eta_{\widetilde{A}}(x)$". When $\pi_{\widetilde{A}}(x)$=0, IFS turns to FS.

Intuitionistic fuzzy sets enable defining both membership and non-membership degrees.

Definition 3. Let A and B are two intuitionistic fuzzy sets in $X = \{x_1, x_2, \cdots, x_n\}$. The Euclidean distance is equal to [40]:

$$d(A, B) = \sqrt{\sum_{i=1}^{n} \begin{matrix} (\mu_{\widetilde{A}}(x_i) - \mu_{\widetilde{B}}(x_i))^2 \\ +\,(\eta_{\widetilde{A}}(x_i) - \eta_{\widetilde{B}}(x_i))^2 \\ +\,(\pi_{\widetilde{A}}(x_i) - \pi_{\widetilde{B}}(x_i))^2 \end{matrix}} \tag{3}$$

Definition 4. Separation validity function (S) represents compactness of the mean of total variation [41].

$$S = \frac{\sum_{k=1}^{c} \sum_{i=1}^{n} \mu_{\widetilde{A}}(x_i)^m d^2(v_k, x_i)}{n \min\limits_{i,j}\{d(v_i, v_j)\}^2} \tag{4}$$

where m is a fuzziness index that determines the fuzziness of the resulting clusters and v is the cluster centers. c is the number of clusters, n is the number of data, $\mu_{\widetilde{A}}$ is the membership matrix of x belonging to the fuzzy set $\widetilde{A}$. $d(v_k, x_i)$ shows the Euclidean distance between data point i and the center of cluster k. The more separate clusters, the smaller S, in other words, the larger distances between clusters.

4. Construction of intuitionistic fuzzy sets

We use Yager's generating function [42] to create intuitionistic fuzzy complement or fuzzy generator:

$$N(\mu_{\widetilde{A}}(x)) = g^{-1}(g(1) - g(\mu_{\widetilde{A}}(x))) \quad (5)$$

where g(.) is an increasing function and g:[0,1] → [0,1].

In Equation (5), a Yager class can be generated by using Equation (6):

$$g(x) = x^{\alpha} \quad (6)$$

Then, Yager's intuitionistic fuzzy generator can be written as:

$$N(x) = (1 - x^{\alpha})^{1/\alpha}, \alpha > 0 \quad (7)$$

when $x = 1$, $N(1) = 0$ and $x = 0$, $N(0) = 1$

Non-membership values are computed by Yager's intuitionistic fuzzy generator $N(x)$. Thus, we can represent IFS with Yager's intuitionistic fuzzy generator:

$$\widetilde{A} = \left\{ \langle x, \mu_{\widetilde{A}}(x), (1 - \mu_{\widetilde{A}}(x))^{1/\alpha} \rangle ; x \in X \right\} \quad (8)$$

And the hesitancy equals to:

$$\pi_{\widetilde{A}}(x) = 1 - (\mu_{\widetilde{A}}(x) + (1 - \mu_{\widetilde{A}}(x))^{1/\alpha}) \quad (9)$$

5. Fuzzy c-means clustering

Fuzzy c-means (FCM), put forward by Dunn [43] and developed by Bezdek [44], is one of the most popular fuzzy clustering techniques [45]. Conventional FCM aims to minimize Equation (10).

$$J(U, V : X) = \sum_{k=1}^{c} \sum_{i=1}^{N} (\mu_{ki})^m d^2(v_k, x_i) \quad (10)$$

where U is the fuzzy membership matrix; V is the cluster centers, and X is the data set to be clustered. μ_{ki} is the membership value of data x_i fuzzy cluster c_k. The fuzzifier m must be larger than 1. If m equals to 1, then the clusters are formed in crisp format. An appropriate U (membership matrix) and V must be computed like Equations (11) and (12) to minimize J_m, for all $k = 1, 2, \cdots$, c.

$$\mu_{ki} = \frac{1}{\sum_{j=1}^{c} [(d_{ki}^2)/(d_{ji}^2)]^{1/(m-1)}}, \quad (11)$$

$$v_k = \frac{\sum_{i=1}^{N} \mu_{ki}^m x_i}{\sum_{i=1}^{n} \mu_{ki}^m} \quad (12)$$

Equations (11) and (12) are used until the termination criterion is satisfied.

6. Intuitionistic fuzzy c-means clustering

In this study, we adopt the intuitionistic fuzzy c-means clustering (IFCM) algorithm proposed by Chaira [46]. In order to convert conditional FCM to intuitionistic fuzzy c-means (IFCM), the cluster centers are updated by considering hesitance values like in Equation (13).

$$\mu_{ki}^* = \mu_{ki} + \pi_{ki} \quad (13)$$

μ_{ki}^* and μ_{ki} are the intuitionistic and conventional fuzzy membership of the data i in cluster k, respectively.

Then, new cluster centers are computed as:

$$v_k^* = \frac{\sum_{i=1}^{N} \mu_{ki}^* x_i}{\sum_{i=1}^{N} \mu_{ki}^*} \quad \forall k, k = 1, 2, \cdots, c \quad (14)$$

Using Equations 13 and 14, the membership values and cluster centers are updated in each iteration of the c-means algorithm. The algorithm stops when the difference between the updated cluster centers and previous cluster centers is less than ϵ, which is a predetermined parameter.

We also consider the intuitionistic fuzzy entropy (IFE) as the second dimension of the objective function. IFE aims to maximize good data in the cluster. The second dimension can be written as [46]:

$$J_2 = \sum_{k=1}^{c} \pi_k^* e^{1-\pi_k^*} \quad (15)$$

where $\pi_k^* = \frac{1}{N} \sum_{i=1}^{N} \pi_{ki}$, $k \in [1, N]$. π_{ki} shows the hesitation value of data i in cluster k. As a result, the final objective function of the algorithm contains two terms that should be minimized:

$$J = \sum_{k=1}^{c} \sum_{i=1}^{N} (\mu_{ki}^*)^m d^2(x_i, v_k) + \sum_{k=1}^{c} \pi_k^* e^{1-\pi_k^*} \quad (16)$$

The integrated IFCM algorithm is depicted in Figure 1. It does not require to know the number of clusters. Before clustering algorithm, Xie-Beni Index (XBI) is computed as a performance indicator to define the number of clusters. At the end, we calculate the ratio of between sum of squared errors (SSB) and the total sum of squared errors (SST) to examine how much of the total variance in the dataset is explained by the algorithm.

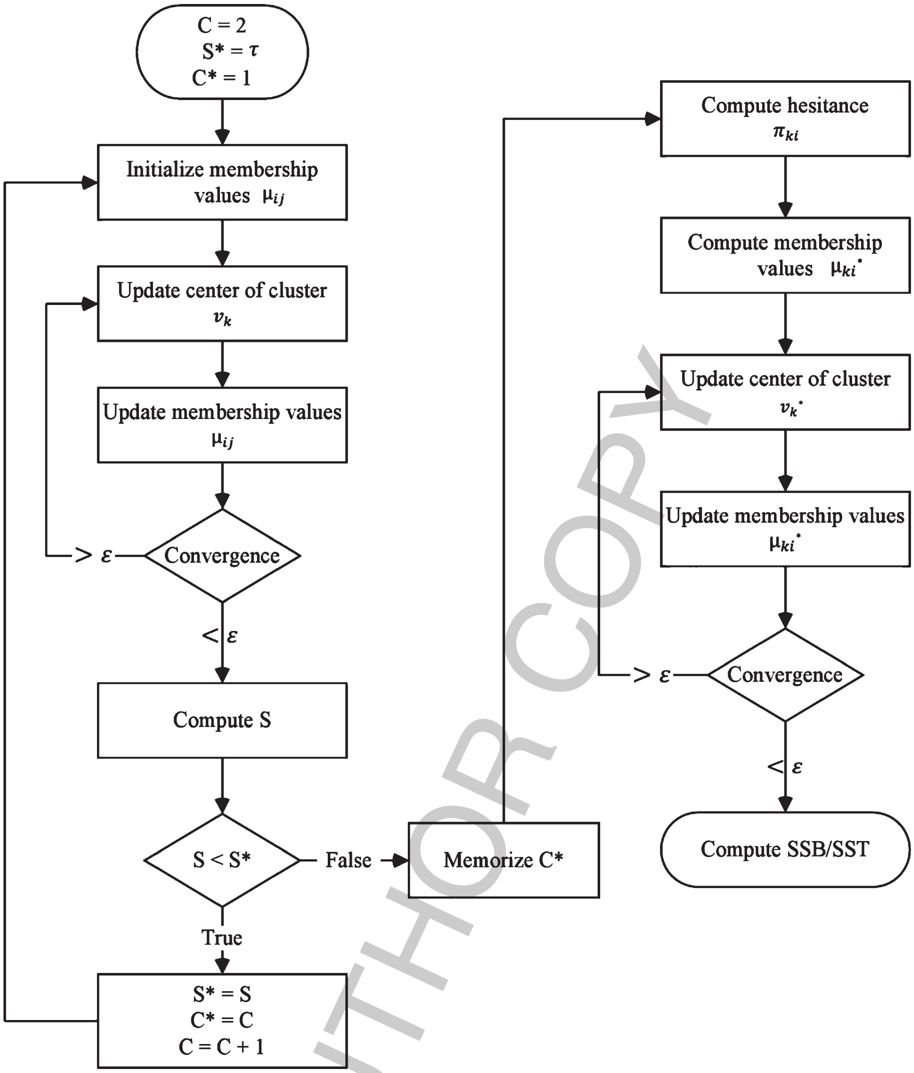

Fig. 1. The integrated IFCM algorithm.

7. Illustrative example

First of all, since the range of data attributes has a difference, they are normalized with any normalization method. Table 2 presents the values of data attributes and their normalized values. Although we have two-dimensional data, we used monodimensional data to illustrate an example.

Before computing the number of clusters, we plot a scatter chart to test it. Figure 2 gives the possible clusters.

Then XBI is calculated to find the optimum number of cluster. Table 3 represents the computed separation values using Equation (4). The algorithm continues the extra four steps after finding any smaller value than the previous iteration to avoid to catch a local minimum. Therefore, the next four iterations are computed after found the lower value for a certain number of cluster.

Now, any clustering algorithm can be applied using the optimum number of cluster, four. Figure 3 demonstrates a visual representation of the results of three different clustering algorithms.

The set of cluster centers are $\{61.33,73.75,85.67,105.00\}$ for KM, $\{61.40,74.79,90.23,107.75\}$ for FCM and $\{63.73,75.74,89.83,105.85\}$ for IFCM. According to the calculated center of clusters, the total variance in the dataset is explained by each algorithm is 0.9506, 0.9667 and 0.9777 respectively KM, FCM, and IFCM. The objective function of FCM has

Table 2
Sample data

Sample ID	Duration	Normalized duration
296945_visit1	61	0.199
256953_visit1	80	0.262
4475525_visit2	87	0.284
5190356_visit1	106	0.346
7620624_visit1	75	0.245
8510415_visit1	112	0.366
10482024_visit1	96	0.314
11245129_visit1	90	0.294
11906039_visit1	60	0.196
11906039_visit2	63	0.206
12009451_visit1	73	0.239
12084705_visit1	106	0.346
12147054_visit1	71	0.232
12731769_visit1	76	0.248

Fig. 2. Expected number of clusters.

Table 3
The optimum number of cluster

Number of clusters	S
2	0.09851
3	0.1489
4	**0.08831**
5	0.1308
6	0.14241
7	0.53297
8	0.54769

a value of 98.19 without IFE, the objective function of IFCM has a value $86.53 + 1.07 = 87.60$. The second addend shows the IFE calculated by Equation (15). These results show that IFCM gives better clusters than FCM and KM.

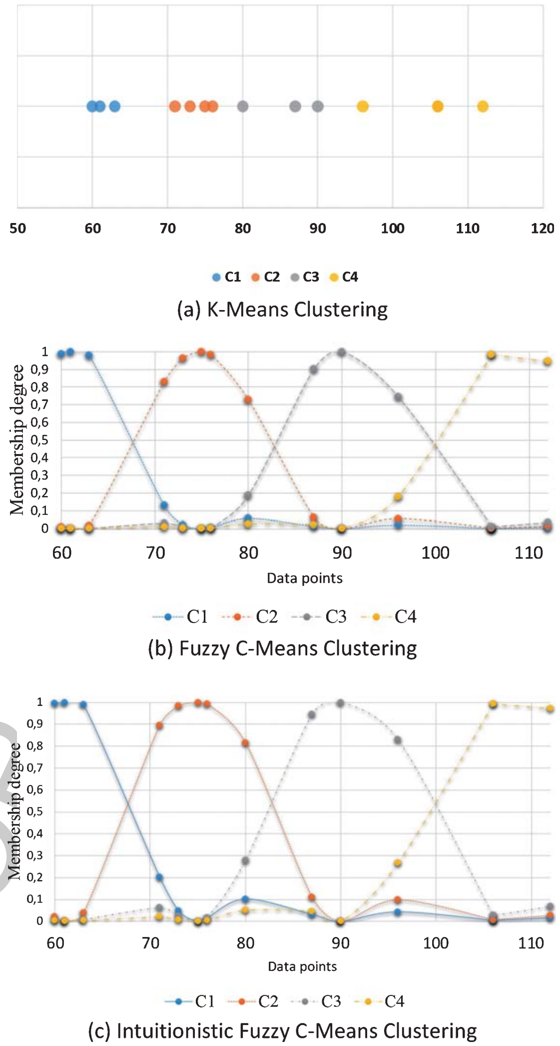

Fig. 3. Results of algorithms.

8. Application

The data were collected by iBeacon devices, a Bluetooth based technology, in one of the major shopping malls, which is located at a strategic point in Turkey's metropolis Istanbul. The data belongs to 446 visits that executed by 399 customers who came to the shopping mall during 1st December 2017. After some preprocessing steps, data has gained two dimensions; duration and number of locations. We eliminated 9 of them because they were outliers. Remaining data were used for KM, FCM and IFCM to compare the results and then visualized by process mining.

We found the optimum cluster number as seven by XBI. Therefore, when we ordered the clusters concerning duration, we get the time interval $\{C_2 : 2 -$

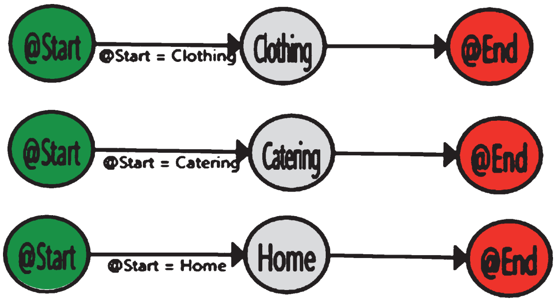

Fig. 4. Mostly followed paths in Cluster C_2.

Table 4
Cluster details

Clusters	Durations	No. of customers
C_2	2–14	129
C_1	15–29	69
C_6	30–47	66
C_4	48–70	58
C_7	71–95	26
C_3	97–131	25
C_5	133–232	18

14, $C_1 : 15-29$, $C_6 : 30-47$, $C_4 : 48-70$, $C_7 : 71-95$, $C_3 : 97-131$, $C_5 : 133-232\}$. When we apply PM to the clusters, we obtain the customer paths in the shopping mall. Since we created many paths, we depicted only mostly followed by customers in each cluster. For example, Cluster C_2 has 129 customer data. 64% of customers visited only one store group, Clothing, Catering and Home store groups visited 45, 27 and 11 times (Figure 4). Since the spent time by customers for cluster C_2 is between 2 and 14 minutes, we can conclude that customers usually visit only one store group in C_2.

When we look at other clusters in Table 4, almost half of the customers left the shopping mall in 30 minutes during 1st December. From Figure 5 to Figure 7 represent mostly followed paths in different clustered as an example of process mining visualization. In the figures showing customer flows, nodes and arrows are colored from green to red. The red nodes refer the store group that has higher case duration. On the other hand, the red arrows present the mostly executed transaction between two nodes concerning the number of execution.

Customers in C_1, have a duration of 15–29 minutes, mainly visited the shopping mall for Clothing stores (Figure 5). Catering has fewer visits, but a bit more spent time than Clothing.

In Figure 6, 32 customers go directly to Catering, and 17 customers go directly to Clothing. The average

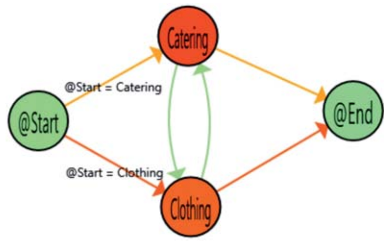

Fig. 5. The path in Cluster C_1 by 99 customers.

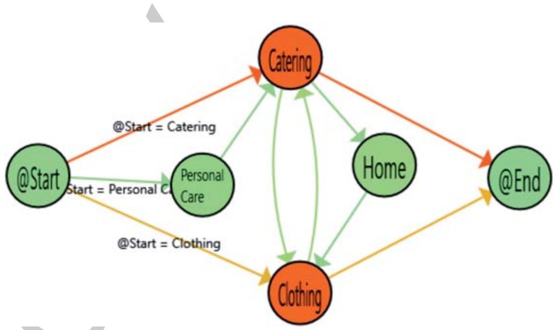

Fig. 6. The path in Cluster C_6 by 50 customers.

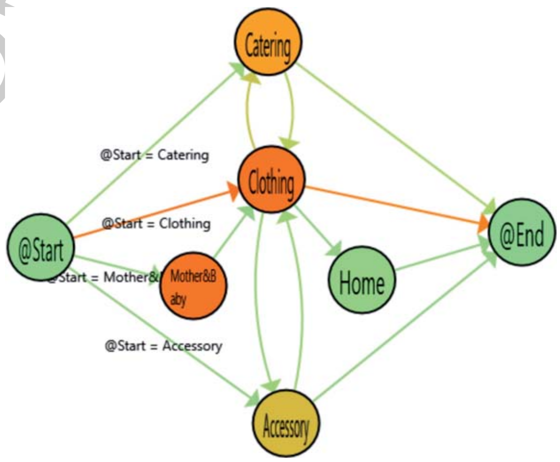

Fig. 7. The path in Cluster C_3 by 16 customers.

durations are almost equal, both of them are near to 51 minutes. A similar evaluation can be made for cluster C_3 in Figure 7.

Although one customer directly goes Mother& Baby stores and 12 of them directly go Clothing, both store groups have the nearly same duration. In this cluster, Catering and Clothing are two store groups mostly visited together. 74% of customers visiting catering also visited any clothing stores.

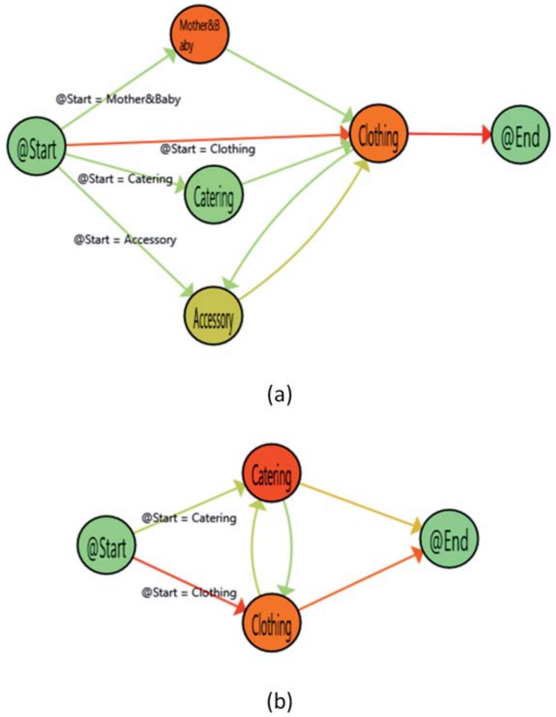

Fig. 8. Paths for the same customers in different visits.

16 customers visited the shopping mall two times in cluster C_6. To illustrate the behavioral change for the same customers in different visits, we focus on paths belonging to these customers. While Figure 8a represents the paths in the first visit, Figure 8b shows the second visit path. There is a remarkable change in the visiting purpose. As a result of clustering, Catering has importance in general. On the other hand, Accessory and Mother&Baby were not visited during the second-time visits for customers who have two visits. Also, two-visit customers visited more stores on their first visits. In their second visit, only Catering and Clothing stores were visited. It may be inferred that customers who spent 30 - 47 minutes go to more store on the first visit. It means that they spend less time in each store. On the other hand, the same customers visited fewer stores but spend more time on the second visit.

9. Conclusion and future directions

In this study, we suggest an intuitionistic fuzzy c-means clustering algorithm for customer paths. Because traditional k-means and fuzzy c-means have a drawback with the number of clusters, we modified the algorithm by computing Xie-Beni Index, is a performance indicator to define the number of clusters. We compared three algorithms, K-means, fuzzy c-means, and intuitionistic fuzzy c-means, to evaluate the results. The experimental results have shown that the intuitionistic fuzzy c-means algorithm is effective in clustering of customer paths with two-dimensional location data, duration and number of visited locations. Moreover, the fuzzy partition matrix provides more information to help the decision maker to determine the final clustering and to identify the boundary data objects.

In contrast to conventional K-means, fuzzy clustering has an advantage that one data point can belong to more than one cluster with membership value. However, this advantage cannot be used in process mining visualization. Intuitionistic fuzzy c-means clustering is used to group data points with less error. Then we discover the paths in each cluster with process mining. Duration-based clustering is not a very effective way to understand customer behaviors. It cannot consider infrequent behavior with statistical information about the activities frequency and state changes. As a future study, we will extend the study to understand customer behavior by using different clustering methods such as trace clustering and sequence clustering.

After an illustrative example for clustering problem, we present the implementation details. The study starts with fuzzy clustering and continues with process mining to visualize the clustered results. Therefore, the study has some conclusions about clustered customer paths that used to examine the cluster behaviors. Data used in this study is collected by a Bluetooth-based technology in one of the major shopping malls in Istanbul. We use process mining technique to increase the understandability of the clustering results. There is a strong relationship between Catering and Clothing store groups in some clusters. 74% of customers who visited Catering also visited Clothing. As an interesting result, customers sometimes visit more stores but spent less time and vice versa. After summarizing the global review of customers, we present two paths belonging to the same customer to understand visit purposes.

Fuzzy clustering also provides to see change in the visiting purpose when combined with process mining. For example, C_6 customers did not visited some store groups such as Accessory and Mother&Baby on the second visit. Almost all customer went to at least one store in Catering or Clothing in each cluster. In some cases, customers who spent over 30 minutes went to more store on the first visit. It means that they have lower average time in each store. On the other hand, the same customers visited fewer stores but spend more time on the second visit.

Some store groups such as hairdresser and cinema may have a high duration, but low visited location data. Since these are considered as outliers, we omit them. Nevertheless, this visit may be critical for specifying a behavioral change of the customer. For example, not visiting the cinema can be a withdrawal symptom of timelessness or economical break. Hence the results may have some missing information.

References

[1] 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* **19**(3) (2019), 557.

[2] O. Dogan and B. Oztaysi, Genders prediction from indoor customer paths by Levenshtein-based fuzzy kNN, *Expert Systems with Applications* **136** (2019), 42–49.

[3] O. Dogan and B. Oztaysi, In-store behavioral analytics technology selection using fuzzy decision making, *Journal of Enterprise Information Management* **31**(4) (2018), 612–630.

[4] M. De Leoni, W.M. van der Aalst and M. Dees, A general process mining framework for correlating, predicting and clustering dynamic behavior based on event logs, *Information Systems* **56** (2016), 235–257.

[5] R. Arroyo, J.J. Yebes, L.M. Bergasa, I.G. Daza and J. Almazán, Expert video-surveillance system for real-time detection of suspicious behaviors in shopping malls, *Expert systems with Applications* **42**(21) (2015), 7991–8005.

[6] Y. Yoshimura, S. Sobolevsky, C. Ratti, F. Girardin, J.P. Carrascal, J. Blat and R. Sinatra, An analysis of visitors' behavior in the louvre museum: A study using bluetooth data, *Environment and Planning B: Planning and Design* **41**(6) (2014), 1113–1131.

[7] 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* **14**(4) (2017), 1786–1792.

[8] N. Abedi, A. Bhaskar, E. Chung and M. Miska, Assessment of antenna characteristic effects on pedestrian and cyclists traveltime estimation based on bluetooth and wifi mac addresses, *Transportation Research Part C, Emerging Technologies* **60** (2015), 124–141.

[9] S. Mou, D.J. Robb and N. DeHoratius, Retail store operations: Literature review and research directions, *European Journal of Operational Research* **265**(2) (2017), 399–422.

[10] 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* **15**(12) (2015), 29821–29840.

[11] W.M. van der Aalst, *Process mining: Data science in action*, Springer, 2016.

[12] R.J.C. Bose and W.M. Van der Aalst, Context aware trace clustering: Towards improving process mining results, in: *Proceedings of the 2009 SIAM International Conference on Data Mining, SIAM*, 2009, pp. 401–412.

[13] M. Song, C.W. Günther and W.M. Van der Aalst, Trace clustering in process mining, in: *International Conference on Business Process Management*, Springer, 2008, pp. 109–120.

[14] D. Ferreira, M. Zacarias, M. Malheiros and P. Ferreira, Approaching process mining with sequence clustering: Experiments and findings, in: *International Conference on Business Process Management*, Springer, 2007, pp. 360–374.

[15] B. Hompes, J.C. Buijs, W.M. van der Aalst, P. Dixit and H. Buurman, Detecting Change in Processes Using Comparative Trace Clustering, in: *SIMPDA*, 2015, pp. 95–108.

[16] O. Dogan, Heuristic Approaches in Clustering Problems, in: *Handbook of Research on Applied Optimization Methodologies in Manufacturing Systems, IGI Global*, 2018, pp. 107–124.

[17] C. Kahraman, B. Oztaysi, S.C. Onar and O. Dogan, Intuitionistic fuzzy originated type-2 fuzzy ahp: An application to damless hydroelectric power plants, *International Journal of the Analytic Hierarchy Process* **10**(2) (2018), 266–292.

[18] Z. Huang and M. Ng, A fuzzy k-modes algorithm for clustering categorical data, *IEEE Transactions on Fuzzy Systems* **7**(4) (1999), 446–452.

[19] K.T. Atanassov, Intuitionistic fuzzy sets, *Fuzzy Sets and Systems* **20**(1) (1986), 87–96.

[20] Á. Rebuge and D.R. Ferreira, Business process analysis in healthcare environments: A methodology based on process mining, *Information systems* **37**(2) (2012), 99–116.

[21] J. Frisby, V. Smith, S. Traub and V.L. Patel, Contextual computing: A bluetooth based approach for tracking healthcare providers in the emergency room, *Journal of Biomedical Informatics* **65** (2017), 97–104.

[22] M.C. Popa, L.J. Rothkrantz, C. Shan, T. Gritti and P. Wiggers, Semantic assessment of shopping behavior using trajectories, shopping related actions, and context information, *Pattern Recognition Letters* **34**(7) (2013), 809–819.

[23] M.L. van Eck, N. Sidorova and W.M. van der Aalst, Enabling process mining on sensor data from smart products, in: *Research Challenges in Information Science (RCIS), 2016 IEEE Tenth International Conference on, IEEE*, 2016, pp. 1–12.

[24] L. Kang and M. Hansen, Behavioral analysis of airline scheduled block time adjustment, *Transportation Research Part E, Logistics and Transportation Review* **103** (2017), 56–68.

[25] M. Delafontaine, M. Versichele, T. Neutens and N. Van deWeghe, Analysing spatiotemporal sequences in bluetooth tracking data, *Applied Geography* **34** (2012), 659–668.

[26] Y.-K. Wu, H.-C. Wang, L.-C. Chang and S.-C. Chou, Customer's flow analysis in physical retail store, *Procedia Manufacturing* **3** (2015), 3506–3513.

[27] J. Yim, S. Jeong, K. Gwon and J. Joo, Improvement of kalman filters for wlan based indoor tracking, *Expert Systems with Applications* **37**(1) (2010), 426–433.

[28] 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* **78** (2017), 55–65.

[29] S. Chen, A. Fern and S. Todorovic, Multi-object tracking via constrained sequential labeling, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1130–1137.

[30] O. Dogan, O.F. Gurcan, B. Oztaysi and U. Gokdere, Analysis of Frequent Visitor Patterns in a Shopping Mall, in: *Industrial Engineering in the Big Data Era*, Springer, 2019, pp. 217–227.
[31] A.A. Shaw and N. Gopalan, Finding frequent trajectories by clustering and sequential pattern mining, *Journal of Traffic and Transportation Engineering* **1**(6) (2014), 393–403.
[32] Y. Marchetti, Q. Zhou, et al., Solution path clustering with adaptive concave penalty, *Electronic Journal of Statistics* **8**(1) (2014), 1569–1603.
[33] J. Evermann, T. Thaler and P. Fettke, Clustering traces using sequence alignment, in: *International Conference on Business Process Management*, Springer, 2016, pp. 179–190.
[34] P. D' Urso and R. Massari, Fuzzy clustering of human activity patterns, *Fuzzy Sets and Systems* **215** (2013), 29–54.
[35] S. Jiang, J. Ferreira and M.C. González, Clustering daily patterns of human activities in the city, *Data Mining and Knowledge Discovery* **25**(3) (2012), 478–510.
[36] A. Manzi, P. Dario and F. Cavallo, A human activity recognition system based on dynamic clustering of skeleton data, *Sensors* **17**(5) (2017), 1100.
[37] C. Fernández-Llatas, J.-M. Benedi, J.M. García-Gómez and V. Traver, Process mining for individualized behavior modeling using wireless tracking in nursing homes, *Sensors* **13**(11) (2013), 15434–15451.
[38] W.M. van der Aalst, How people really (like to) work, in: *International Conference on Human-Centred Software Engineering*, Springer, 2014, pp. 317–321.
[39] L. Zadeh, Fuzzy sets, *Information and Control* **8**(3) (1965), 338–353.
[40] E. Szmidt and J. Kacprzyk, Distances between intuitionistic fuzzy sets, *Fuzzy Sets and Systems* **114**(3) (2000), 505–518.
[41] X. Xie and G. Beni, A validity measure for fuzzy clustering, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **13**(8) (1991), 841–847.
[42] R.R. Yager, On some new classes of implication operators and their role in approximate reasoning, *Information Sciences* **167**(1-4) (2004), 193–216.
[43] J.C. Dunn, A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters, *Journal of Cybernetics* **3**(3) (1973), 32–57.
[44] J.C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Kluwer Academic Publishers, Norwell, MA, USA, 1981.
[45] B. Oztaysi and M. Isik, Supplier Evaluation Using Fuzzy Clustering, in: *Supply Chain Management Under Fuzziness*, Springer Berlin Heidelberg, 2014, pp. 61–79.
[46] T. Chaira, A novel intuitionistic fuzzy c means clustering algorithm and its application to medical images, *Applied Soft Computing* **11**(2) (2011), 1711–1717.