An Interval Valued Hesitant Fuzzy Clustering Approach for Location Clustering and Customer Segmentation

Sultan Ceren Öner[(✉)] and Başar Öztayşi

Industrial Engineering Department,
Istanbul Technical University, Maçka, İstanbul, Turkey
{csalkin, oztaysib}@itu.edu.tr

Abstract. Because of the irrepressible growth in information technologies and telecommunication infrastructure especially for mobile devices, people are more disposed to search proper products and find out attractive offers with lower prices. In order to reach potential customers, companies deal with offering personalized messages including special promotions and discounts. In this respect, recommender systems have begun to use as one of the essential tools for making appropriate selections considering diversified conditions and personal preferences. On the other hand, users' preferences could not be easily determined or predicted in some cases, as seen in visiting prediction of mobile users. Thus, the use of location based service applications enable the determination of users visiting patterns, except making predictions. In this study, an interval valued hesitant fuzzy clustering approach is adapted based on location similarity and fuzzy c means clustering is applied for user segmentation. After that, matching location groups and user segments is provided the representation of user visiting tendency. By using this approach, advertisers will be able to handle their advertisements considering location similarities and user groups that helps the implementation of personalized advertising recommender systems.

Keywords: Location similarity · Segmentation · Hesitant fuzzy clustering · Interval valued hesitant fuzzy clustering

1 Background

As a result of rapid changes in mobile devices and location tracking technologies, more and more people follow instant messages that present customized offers for specific locations. Location based systems or in another saying, location aware systems rely on geographical information determination of mobile users via global positioning systems (GPS) or Bluetooth (via Beacons) to pinpoint users' real time locations (Lee et al. 2015) .Today, location based systems are widely adapted in real life, especially seen for targeted advertisement. This adaptation can be seen as location based recommender systems which present an alternative channel that enables companies to send special offers based on consumers' preferences and previous location and present opportunities for personalization and increasing revenue (Xu et al. 2009). According to Junglas and Watson (2008), two main steps are necessary to conduct location based systems:

location determination and modeling to indicate the context of location, and "possible" future locations by geo spatial data processing and prediction algorithms and second, user preference modeling for specifying user properties from previously searching patterns, online profiles, previously visited places etc.

In recent years, location determination and the prediction of future visits have become important and integrated with recommender systems in targeted advertisement management, especially making personalized offers which can enhance companies' communication channels for customers with limited advertisement budgets. These systems are relied on time and preference based services including monitoring of current position of mobile user and searching appropriate offers or advertisement according to possible visiting places extracted from previously visited places (Wu et al. 2015). However, customers generally avoid to read of follow the instant messages because of the substantial number of messages, irrelevant offers or redundancy of the message context (Shin and Lin 2016). The other reasons of avoiding these push up messages are location privacy which is a significant barrier to the penetration of location based services and accidentally keystrokes (Abbas et al. 2014; Pingley et al. 2012). To overcome this problem, continuous information feedback for incoming conditions might be beneficial. Thus, advertisers are prone to search and apply more efficient approaches that provide both personalized contexts and broad alternatives to their recommender systems for individual and commonly-held advertisements (Dao et al. 2012). Additionally, user interests and needs continuously change and the differentiation of these needs and interests can be problematic to perform efficient recommender systems. Thus, location prediction and customer segmentation exist as the most important topics in location based recommender systems for reflecting user movements as a characteristic of customer preferences or needs (Fan et al. 2015).

In this respect, the main research question of this study is the matching of customer movements with respect to previously visited locations and location characteristics the determination of alternative locations according to the similarities assigned by common characteristics. For this purpose, location clustering can be applied for grouping locations and alternative location detection can be conducted of in terms of the similarities appeared from the location clusters and customer segments. Since the nature of the problem contains insufficient data, vague environment and conflicting expectations from diverse users, the clustering problem varies as a fuzzy clustering problem, in particular, fuzzy partition of clusters (Aliahmadipour et al. 2017). In addition to that, location data could be presented as interval valued fuzzy numbers to provide the hesitancy of users' movements which can enhance the recommendation of proper alternative locations. Therefore, in this study, location grouping is conducted by using interval valued hesitant fuzzy clustering. Then, the extracted location clusters are evaluated according to fuzzy c means based user segments. In this way, alternative location clusters for each characteristic are acquired.

The remainder of this paper is structured as follows: the second section explains the brief concepts of location based mobile advertising and location based clustering. The third part presents basic concepts of fuzzy clustering, hesitant fuzzy sets and interval valued hesitant fuzzy clustering. The proposed methodology is shown with a numerical application and the last section contains the conclusion and discussion part.

2   Related concepts

2.1   Location Based Mobile Advertising

Because of the widespread usage of mobile technologies such as GPS and Bluetooth 4.0, location based systems can adapt users' personalized interests in a specific location in a specific time (Fan et al. 2015). Thus, location based systems aim to provide content providers to send proper services to customers when they visit a specific location at a specific time (Tussyadiah 2012). The main concepts of these services are tagging, tracking, navigating and mobile commerce. In particular, tagging enables matching relevant information to a specific location. Tracking services provide information on the position of objects. In addition to that, navigating systems assist people to move from one place to a target destination. Mobile commerce systems are adopted as sending proper advertisements to a targeted customer in a specific location to facilitate location and event based service flow (Wu et al 2015). All these services are the results of consumer interests for the developing of targeted location based systems (Lin et al. 2016).

Location based services include both pull and push systems. Pull advertising reveals when customers search information about a specific topic or item. As distinct from pull advertising, advertisers send automatic messages to customers appeared a specific location in push advertising. All these mechanisms could be evaluated as a type of targeted advertising where vendors send customized messages to consumers' mobile devices when they are near a specific location (Unni and Harmon 2007). In this respect, permission based pull advertising orientation; real time positioning and personalization are fundamentals for location based advertising.

Although significant advantages appeared using location based systems, some challenges are also aroused. Two main challenges are related to the privacy issue and irritation of messages: customers do not want to share their location data with advertisers and do not expect to receive push messages due to the irrelevance of the message and disruption of instant messages (Xu et al. 2009). The other challenge for location based advertising is customers' location accuracy and the prediction of future routes that a large volume of both spatial and temporal data should be processed. Another research direction is the decision of the integration of different data sources especially for complicated business problems as seen in mobile marketing. In our case, selection of alternative locations requires collecting different data from diversified sources to provide conduction of the relationship between people, time, location, objects and their relationships between each other. The core business in these applications is to ensure necessary information to users, provide the automatic implementation of the services that satisfy customer expectations and collect information concerning about actual data. (Gavalas et al. 2014).

2.2   Location Based Clustering

The revealing of advancements in mobile technologies and the wide application of location based services has derived the integration of location based services with respect to user similarities. Today, advertisers are prone to deliver advertisements and promotion messages considering location information that is stated from mobile

devices (Lin et al. 2016). The demand for location based systems integration with clustering is extracted as a need of understanding the process that customer makes the purchasing decision considering location, time and actual needs (Schilke et al. 2004). The needs can be derived from consumer life styles, demographical information, consumption behavior and the reaction to previously sent messages (Shin and Lin 2016). On the other hand, these factors don't solely reflect the entire purchasing decision. Thus, researchers try to search other indicators that can reflect customer characteristics such as geographical data, digital participation in social media and search history for products for better understanding of the changes in customers purchasing tendency (Gavalas et al. 2014).

Location based clustering is an essential tool for the use of recommender systems. Recommender systems have divided into two main filters: content based filtering and collaborative filtering. Content based filtering considers the degree of similarity between offers and interests. However, collaborative filtering investigates the degree of integration of recommendations and user preferences (Wei et al. 2010). According to Zhang et al. (2007), content based filtering is conducted by grouping similar items for offering new items according to previous preferences and collaborative filtering relies on finding the appropriate option based on users' preferences to decide new users' preferences (Zhang et al. 2007). In this respect, user preference similarities could be evaluated from geo data, visited location category, price level, rating, and number of stars, as seen from Foursquare for location based recommender systems. To conduct the similarity calculation, users should be grouped as segments and the recommendations should be given according to the diversification of these segments. The divergence between users can be accomplished either by categorization or by adapting analytical techniques such as cluster analysis, heuristic methods, regression, neural network, kNN, link analysis, decision tree, association rule mining etc. (Park et al. 2012). By using these methods, gathering user comments and reflections that the essential clue of customer disposition are provided to assess the changes in customer behavior and make realistic offers not to spend huge amount of data that means a vast of time, money and manpower. Lastly, runtime of the segmentation and recommendation techniques have a significant role for the implementation of real time location based clustering (Öztayşi et al. 2016).

3   Motivation

Because of the results of mobile technologies have been widely penetrated and improved to facilitate marketing operations and advertising, mobile-wireless technologies are increasingly applied to send proper messages to customers (Ortega et al. 2013). Consequently, mobile location based applications have been gradually adapted to make individualized recommendations using customers' geo data occurred from GPS or wireless indoor positioning systems (Liu 2007). As a consequence of increasing demand on mobile technologies and improvements of the integration of smartphones with location detection systems, personal positioning techniques based advertising systems have become emitted and these improvements facilitated the tracking customers' physical location and purchasing behavior in shopping fields. For this reason,

clustering algorithms are widely conducted on the evaluation of the similarities stated from individuals' positions in shopping mall and provide the assignment of the customer groups. In other words, location-aware systems can be adapted to advertising activities by making location sensitive suggestions or promotions in accordance with customer's previous visits for focusing on the customers' future buying behavior.

The literature provides a wide range of algorithms for clustering, these algorithms are categorized according to the formation of clusters by using fuzzy data or crisp data and (Oztaysi and Onar 2014). Crisp clustering is applied for classical sets theory and includes the process of converting input data to mutually exclusive subsets. Based on this fact, crisp clustering algorithms assign input data to one certain cluster. On the other hand, fuzzy clustering algorithms assign an element to several clusters simultaneously with various degrees of membership that ranges between 0 and 1 and naturally, the belonging of the clusters are presented as degrees of each input data to allow the appearance of diversified distribution of assigned clusters (Oztaysi and Isik 2014). This property is useful for grouping customers to provide them wide range of recommendations.

Academicians and practitioners generally use clustering algorithms for dividing a dataset into different clusters. To reflect uncertainty for forming clusters, clustering tasks are handled as two points of view; i) considering uncertain data, ii) considering crisp data with uncertain clusters, in particular, fuzzy partition. (Aliahmadipour et al. 2017) Therefore, one of the most adapted clustering algorithm for fuzzy clustering is fuzzy c means clustering as agglomerative approach in which initial clusters should be determined in advance (Chen et al. 2014).

In this study, location grouping according to Foursquare ratings of each shopping mall, number of votes in Foursquare, monthly total number of visits, number of stores in shopping mall, transportation level (1 presents worst and 3 presents best), real estate index from Hurriyet Emlak and percentage of rent price variations in a year is achieved initially by using interval valued hesitant fuzzy c means clustering. Then, users are grouped according to visiting time zone (morning (1), afternoon (2), late afternoon (3) and night (4)), location, type of mobile application and location visit day using fuzzy c means clustering. Finally, user clusters are matched with location groups with some certain criteria such as number of visits per month and number of visited different places. This approach will ensure gathering location similarity and constitutes the matching of location groups and user groups and enhance various location alternatives that a specific user could visit. Thus, advertisers will be able to submit location related promotions and advertisements to relevant customer groups as a starting point of recommender systems.

4 Methodology

4.1 Preliminaries

Hesitant fuzzy sets are the extension of fuzzy sets that presents hesitancy degree when there is an uncertainty in terms of the determination of membership function. Some of the related definitions are given in the following:

Definition 1. A hesitant fuzzy set (*HFS*) on a reference set X is a function of h that returns to a subset of values in [0, 1] and h could be represented as follows:

$h: X \rightarrow \{[0,1]\}$

In this respect, a *HFS* can be expressed as the union of the membership functions.

Definition 2. Let M is a set of n number of membership function which could be represented as $M = \{\mu_1, \mu_2.., \mu_n\}$ and HFS with M could be defined as h_M: $M \rightarrow \{[0, 1]\}$ and $h_M(x) = \{\mu_1(x) \bigcup \mu_2(x) \bigcup .. \bigcup \mu_n(x)\}$.

Definition 3. Let S a linguistic term set as $\{s_0, s_1, s_2,.., s_f\}$ which has an order of terms as S: $s_i \leq s_j$ where $i \leq j$ and has a maximization and minimization operator as max $(s_i, s_j) = s_i$ and min $(s_i, s_j) = s_j$ where $i \geq j$. A hesitant fuzzy linguistic term set is represented as H_s which has an ordered finite subset of sequential linguistic terms using upper (H_{S^+}) and lower bounds (H_{S^-}) and could be also defined as $H_{S^+} = \max(s_i) = s_j$ and where $s_i \leq s_j$ and $H_{S^-} = \min(s_i) = s_j$ where $s_j \leq s_i$ for $s_i \in H_s; \forall i$

Definition 4. The HFLTS could be defined as the composition of upper and lower bounds which could be presented as H_S: $[H_{S^-}, H_{S^+}]$ where $H_{S^-} \leq H_{S^+}$.

Definition 5. Wang et al. (2014). Linguistic term s_i is involved in a linguistic term set S, if $\psi_i \in R^+$. Linguistic scale function φ is mapping as: $S \rightarrow R^+$ such that φ: $s_i \rightarrow \psi_i (i = 0, 1, \ldots, 2t)$, where $0 \leq \psi_1 < \psi_2 < \ldots < \psi_{2t} \leq 1$. In this respect, linguistic scale function used in this study is given as follows: $\varphi(s_i) =$
$$\begin{cases} \frac{t^\alpha - (t-i)^\alpha}{2t^\alpha} & i = 0, 1, \ldots, t \\ \frac{t^\beta + (i-t)^\beta}{2t^\beta} & i = t+1, t+2, \ldots, 2t \end{cases}, \text{ where } \alpha, \beta \in (0, 1].$$

Definition 6. *Interval valued hesitant fuzzy sets.* (Wang et al. 2014). Let $X = \{x_1, x_2, \ldots, x_n\}$ be a reference set and $s_{\theta(X)} \in S$. An interval valued hesitant fuzzy linguistic set is defined as $H = \{x, s_{\theta(X)}, \Gamma_H(x), x \in X\}$

Zhang et al. (2016) defined some operations on the IVHFSs. Let $\alpha = \langle s_{\theta(\alpha)}, \Gamma_\alpha \rangle$ and $\beta = \langle s_{\theta(\beta)}, \Gamma_\beta \rangle$ be two interval valued hesitant fuzzy linguistic number (IVHFN).

(1) $\text{neg}(\alpha) = \bar{\varphi}^{-1}(\bar{\varphi}(s_{2t}) - \bar{\varphi}(s_{\theta(X)})), \bigcup_{r_1 = [r_1^-, r_1^+] \in \Gamma_\alpha} \{[1 - r_1^+, 1 - r_1^-]\}$

(2) $\alpha \oplus \beta = \bar{\varphi}^{-1}(\bar{\varphi}(s_{\theta(\alpha)}) + \bar{\varphi}(s_{\theta(\beta)}) - \bar{\varphi}(s_{\theta(\alpha)}).\bar{\varphi}(s_{\theta(\beta)})),$
$\bigcup_{r_1 = [r_1^-, r_1^+] \in \Gamma_\alpha, r_2 = [r_2^-, r_2^+] \in \Gamma_\beta} \{[r_1^- + r_2^- - r_1^-.r_2^-, r_1^+ + r_2^+ - r_1^+.r_2^+]\}$

Definition 7. (Wang et al. 2014). Let $\alpha = s_{\theta(\alpha)}, \Gamma_\alpha = s_{\theta(\alpha)}, \bigcup_{r = [r^-, r^+] \in \Gamma_\alpha} \{[r^-, r^+]\}$ be an interval valued hesitant fuzzy linguistic number. Score function $S(\alpha)$ is defined for an interval valued hesitant fuzzy linguistic number (IVHFLN) as follows:

$$S(\alpha) = \varphi^-\left(s_{\theta(\alpha)}\right) \times \frac{\sum_{r=[r^-,r^+]\in\Gamma_\alpha}\left(r^- + r^+\right)}{2.\#\Gamma_\alpha}$$

where $\#\Gamma_\alpha$ is the number of the interval numbers in Γ_α. For two *IVHFNs* α and β if $S(\alpha) > S(\beta)$, then $\alpha > \beta$; if $S(\alpha) = S(\beta)$, then $\alpha = \beta$.

Definition 8. Let $\alpha = s_{\theta(\alpha)}, \Gamma_\alpha$ and $\beta = s_{\theta(\beta)}, \Gamma_\beta$ are two *IVHFLN*s. Distance between two IVHFNs is defined in the following:

$$d(\alpha,\beta) = \sqrt{\frac{1}{2\times\#\Gamma_{\tilde{h}}} \times \sum_{r_1=[r_1^-,r_1^+]\in\Gamma_\alpha, r_2=[r_2^-,r_2^+]\in\Gamma_\beta} \left(\left(\bar{\varphi}(s_{\theta(\alpha)})r_1^- - \bar{\varphi}(s_{\theta(\beta)})r_2^-\right)^2 + \left(\bar{\varphi}(s_{\theta(\alpha)})r_1^+ - \bar{\varphi}(s_{\theta(\beta)})r_2^+\right)^2\right)}$$

where $\#\Gamma_{\tilde{h}} = max\left(\#\Gamma_\alpha, \#\Gamma_\beta\right)$ that $\#\Gamma_\alpha$ and $\#\Gamma_\beta$ are the number of the interval numbers in Γ_α and Γ_β respectively.

4.2   Fuzzy c means clustering

Clustering is defined as the process of dividing a set of observations into subgroups which are entitled clusters. Various techniques can be used for clustering and the aim of these techniques is to organize input data so as to make similar objects in a cluster, and dissimilar objects in different clusters (Han et al. 2001). Clustering techniques are based on similarity term which is calculated by mathematical distance (Babuska 2009). Clustering is an unsupervised learning technique, in other words there are no predefined groups and a single correct solution (Theodoridis and Koutroumbas 2008).

In clustering problem, the input is a set of observations or objects each of which consists of different attributes. The result of cluster analysis produces the clusters and membership of each data point to these clusters. These outputs are represented by the partition matrix. Ruspini (1970) defines the conditions for a fuzzy partition matrix as follows:

$$\mu_{ik} \in [0,1], 1 \leq i \leq c, 1 \leq k \leq N \tag{1a}$$

$$\sum_{i=1}^{c} \mu_{ik} = 1, 1 \leq k \leq N, \tag{1b}$$

$$0 < \sum_{k=1}^{N} \mu_{ik} < N, 1 \leq i \leq c \tag{1c}$$

Equation (1b) constrains the sum of each column to 1, and thus the total membership of each equals one.

One of the most popular fuzzy clustering methods is fuzzy c-means (FCM) which is based on minimization of the following objective function:

$$J(Z,U,V) = \sum_{i=1}^{c}\sum_{j=1}^{N} (\mu_{ij})^m z_j - v_i^2 \tag{2}$$

where Z is the data set to be partitioned, U is the fuzzy partition matrix, V is the vector of cluster centers. In the formula, N represents the number of observations, c is the number of clusters and μ shows the membership value, m is the parameter called fuzzifier which determines the fuzziness of the resulting clusters. The fuzzifier parameter can get values 1 and more. When the fuzzifier parameter equals to one, then the clusters are formed in crisp format. In the formula, z_k - v_i shows the distance between observation k and the center of cluster i.

The minimization of the mention objective function represents a nonlinear optimization problem that can be solved by using a variety of methods such as iterative minimization, simulated annealing or genetic algorithms. Babuska (2009) presents the steps of fuzzy c-means (FCM) algorithm as follows:

1. *Initialize U=[u_{ij}] matrix, $U^{(0)}$*
2. *At k-step: calculate the centers vectors $V^{(k)}$=[v_i] with $U^{(k)}$*

$$v_i = \frac{\sum_{i=1}^{N} \mu_{ij}^m \cdot z_j}{\sum_{i=1}^{N} \mu_{ij}^m}$$

3. *Update $U^{(k)}$, $U^{(k+1)}$*

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{\left\| z_j - v_i \right\|}{\left\| z_j - v_k \right\|} \right)^{\frac{2}{m-1}}}$$

If $\left\| U^{(k+1)} - U^{(k)} \right\| < \delta$ then STOP; otherwise return to step 2.

4.3   Interval Valued Hesitant Fuzzy c means clustering

In some real life cases, data could obtain hesitancy or crisp data clustering could be presented including hesitation degree. This hesitation may be applied as considering uncertain data and hesitant fuzzy data or considering crisp data set with respect to uncertain clusters (Aliahmadipour et al. 2017). Thus, fuzzy c means clustering algorithm should be adapted to interval valued hesitant fuzzy sets. Before implementation of clustering, location value evaluation and comparison is conducted with three expert. The procedure is described in the following:

1. Initialize U = [u_{ij}] matrix and transform the data to pairwise interval valued hesitant fuzzy expression for each decision maker.
2. Aggregate pairwise interval valued hesitant fuzzy expression using IVHFN aggregation operator given in the following:

Let $\alpha_j = G_j(x) = s_{\theta(\alpha_j)}, \Gamma_{\alpha_j}$ $(j = 1, 2, \ldots, n)$ be a collection of IVHFLNs, then the IVHFLPWA operator can be defined as follows:

$$IVHFLPWA(\alpha_1, \alpha_2, \ldots, \alpha_n) = \frac{T_1}{\sum_{i=1}^n T_i} \alpha_1 \oplus \frac{T_2}{\sum_{i=1}^n T_i} \alpha_2 \oplus \ldots \oplus \frac{T_n}{\sum_{i=1}^n T_i} \alpha_n$$

$$= f^{-1}\left(\sum_{j=1}^n \frac{T_j}{\sum_{i=1}^n T_i} f\left(s_{\theta(\alpha_j)}\right)\right), \bigcup_{r_1 = [r_1^-, r_1^+] \in \Gamma_{\alpha_1}, \ldots, r_n = [r_n^-, r_n^+] \in \Gamma_{\alpha_n}} \left\{ \left[\frac{\sum_{j=1}^n f\left(s_{\theta(\alpha_j)}\right) r_j^- T_j}{\sum_{j=1}^n f\left(s_{\theta(\alpha_j)}\right) T_j}, \frac{\sum_{j=1}^n f\left(s_{\theta(\alpha_j)}\right) r_j^+ T_j}{\sum_{j=1}^n f\left(s_{\theta(\alpha_j)}\right) T_j} \right] \right\}$$

where $T_1 = 1; T_j = \prod_{k=1}^{j-1} S(\alpha_k)$ for $(j = 1, 2, \ldots, n)$ and $S(\alpha_k)$ is a score function of α_k calculated from Definition 7.

3. Determine cluster number according to Xie Beni index.
4. At k-step: calculate the centers vectors $V^{(k)} = [v_i]$ with $U^{(k)}$

$$v_i = \frac{\sum_{i=1}^N \mu_{ij}^m \cdot z_j}{\sum_{i=1}^N \mu_{ij}^m}$$

5. Calculate distance between cluster centers and points using Definition 8.
6. Same as the Step 3 in fuzzy c means clustering.

5 Application

As mentioned before, location similarity and grouping is necessary as the initial step for location prediction of moving customers. Thus, location grouping is adapted using Foursquare ratings of each shopping mall, number of votes in Foursquare, monthly total number of visits, number of stores in shopping mall, transportation level as 1 presents worst and 3 presents best, real estate index from Hurriyet Emlak and percentage of rent price variations in a year. After that, pairwise comparison matrix of each shopping mall is gathered in the form of interval valued hesitant fuzzy numbers. Due to the page restrictions, a sample of crisp data and hesitant information computed from linguistic scale function are given in Tables 1 and 2. Hesitant fuzzy information matrix of each locations are determined from comparison of two locations with each other and interval valued hesitant fuzzy numbers are gathered from this comparison using linguistic scale function description given in Definition 5. Note that the lower side of pairwise comparison matrix is calculated from Definition 6 using neg (α).

After that, interval valued hesitant fuzzy c means clustering is adapted to this pairwise comparison matrix. In order to determine the number clusters, Xie-Beni index is used. The algorithm is run with different values for c parameter as presented in Table 3. The lower values of Xie-Beni index refer to better clustering results, thus for

Table 1. A sample of crisp data of shopping malls (7 from 31 locations)

Location ID	Location name	Foursquare rating	Votes in foursquare	Total number of visits (month)	Variety of stores	Transportation level	Real estate index	Variety of rent (%)
111	City's Nişantaşı	8.8	17,447	875,342	90	3	36	−0.14
60	212 AVM	6.6	8900	543,876	115	1	16	−0.22
62	Akasya AVM	9	29,469	1,134,129	237	3	35	0.23
66	Astoria	6.6	2813	798,645	61	2	31	−0.4
68	Buyaka	8.5	17,556	654,390	106	1	14	−0.04
69	Canpark	6.6	7206	876,522	180	2	14	−0.04
71	Capitol	8	14,827	1,245,176	137	3	18	0.04

Table 2. A sample of aggregated pairwise comparison of hesitant fuzzy information of each location

	111	60	62	66	68
111	[1, 1]	{[0.6,0.8], [0.7,0.8]}	{[0.3, 0.6], [0.7, 0.9]}	{[0.3, 0.4], [0.5, 0.7]}	{[0.3, 0.4], [0.4, 0.5], [0.5, 0.6]}
60		[1,1]	{[0.3, 0.4], [0.5, 0.6]}	{[0.3, 0.5], [0.6, 0.8]}	{[0.2, 0.3], [0.4, 0.5]}
62			[1,1]	{[0.7, 0.8], [0.8, 0.9]}	{[0.5, 0.7], [0.8,0.9]}
66				[1,1]	{[0.4, 0.6]}
68					[1,1]

Table 3. Xie-Beni index values for different values of c parameter

C	2	3	4	5
Xie-Beni	1.52×10^{-06}	10.28×10^{-07}	3.45×10^{-05}	6.2×10^{-05}

this study c value is selected as three, which means in the study three clusters are formed. Second, cluster centers are randomly determined and IVHFLS based distance of each location to each cluster center is computed from using Definition 8.

In order to understand the formed groups, the best way is to analyze the centroid table. Centroid table shows the typical characteristic of each cluster by calculating the cluster centers. Table 4 shows the resulting centroid table for the study when c is equal to three. A sample centroid table for locations 111, 60, 62, 66 and 68 is presented in Table 4.

Finally, clusters and locations in each cluster are represented in Figure 1. From this graph, one could conclude that Cluster 1 and Cluster 3 have intersections that constitute alternative locations of each other. Cluster 2 and Cluster 1 have less intersection and Cluster 2 and Cluster 3 don't have any intersection. According to the characteristics of

Table 4. Centroid table of fuzzy c-means algorithm (c = 3)

Locations	Clstr1	Clstr2	Clstr3
111	2.69	0.14	1.7
60	0.23	0.01	0.99
62	3.67	0.7	6.39
66	0.35	0.01	0.83
68	1.67	0.14	1.21

These two components explain 60.44 % of the point variability.
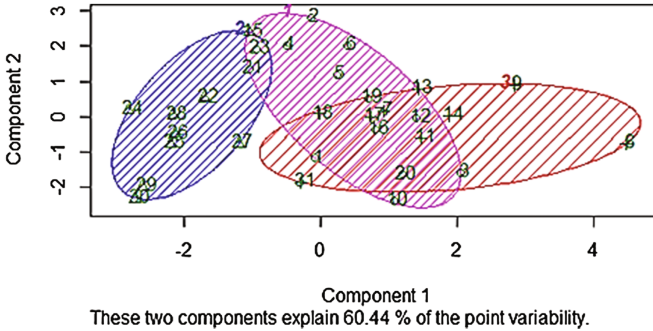
Fig. 1. Cluster plot for locations (shopping malls)

the locations, Cluster 1 is entitled as "Crowded shopping malls" due to their trans-portation availability and increasing level of sub-province development in their own right. Cluster 2 is named "Middle income focused shopping malls" as due to the variety of stores for each income level. Cluster 3 is identified as "Upstate shopping malls" because of their prestigious popularity and outstanding brand.

In the second phase, users are grouped according to time zone (morning (1), afternoon (2), late afternoon (3) and night (4)), location, type of mobile application and location visit day by implementing fuzzy c means clustering. Again, Xie-Beni index values for different values of c parameter are determined and four clusters are gathered to perform whole dataset of users. Second, cluster centers are randomly assigned and distance between user visits vector to cluster centers are calculated. Finally, customers are grouped as seen from Figure 2.

The most important result from this study is to identify the characteristics of the user segments. Based on the cluster plot given in Figure 2, the segments are entitled as "voyager visitors". The customers in cluster 1 are using airports and upstate shopping malls dramatically higher than the existing customers and generally visit in Fridays and Saturdays. In the second cluster, customers could be named as "weekday visitors" that generally involves workers that visits shopping malls in the afternoon breaks. Third cluster obtains "popular shopping mall visitors" that customers spend their leisure time in popular and overcrowded shopping malls. Final cluster is named as "vacation" focused hedonics" that generally visit prestigious shopping malls in the weekends.

The final step for location based clustering is the matching location segments and user segments. To perform the matching of visited location and user segments, location
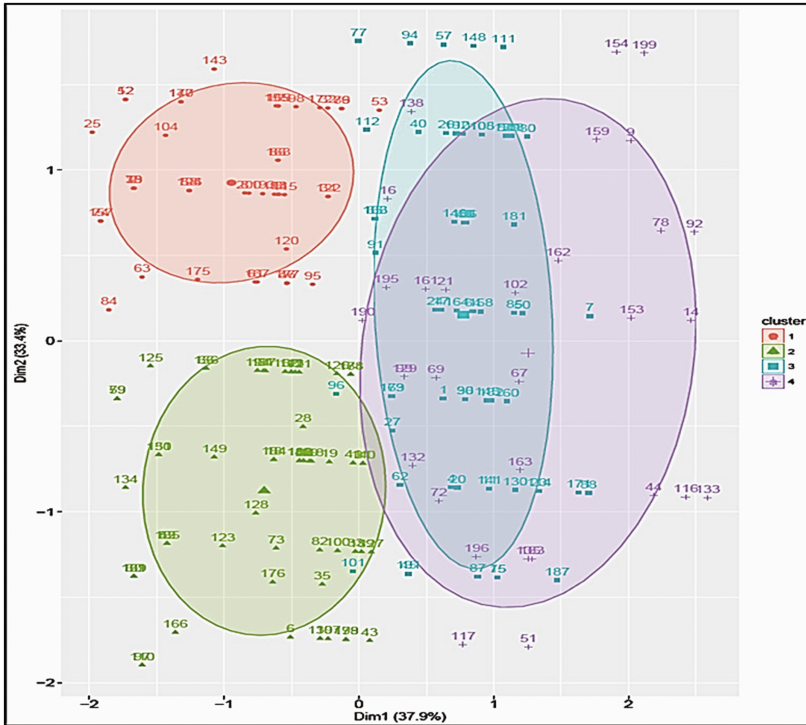
Fig. 2. Cluster plot for customer segments

Table 5. User segment-location group matching

	(1) Crowded shopping malls	(2) Middle income focused shopping malls	(3) Upstate shopping malls
Voyager visitors			x
Weekday visitors	x	x	
Popular shopping mall visitors			x
"Vacation" focused hedonics	x		

clusters and user clusters database are merged that rows present user segments and columns indicate location segmentation cluster considering some certain criteria such as number of visits per month and number of visited different places as presented in the following manner.

According to Table 5, "Vacation" focused hedonics" mainly visit crowded shopping malls to discover new shopping trends and make use of their spare time by activities. Similarly, "Weekday visitors" has the same reason but they generally use shopping malls in the afternoon break for eating lunch. Besides that, "Weekday

visitors" also visits middle income focused shopping malls after hours to follow promotions. The motivation of this preference may be they do not want to allocate their spare time for going shopping. On the other hand, "Voyager visitors" and "Popular shopping mall visitors" generally prefer to go shopping malls to spend their time in their weekends.

6   Conclusion

Because of the results that mobile technologies have been widely penetrated and improved to facilitate marketing operations, mobile-wireless technologies are increasingly applied to send proper messages to customers. Before that, customer purchasing tendency should be properly analyzed considering diversified characteristics such as previous visits, location data etc. In this respect, location clustering and user segmentation can be applied for grouping locations before the determination of alternative locations.

To detect users' location, various services can be maintained such as an advertisement or navigation to a specific location. In this study, a novel use of this data is presented and initial result from a real world case study is conducted. To this end, data from a beacon network is collected, preprocessed and clustered for user segmentation. On the other side, location segmentation is implemented using interval valued hesitant fuzzy clustering.

Results indicate that such a location data from various locations has the potential to show customers' life style and interests. In traditional marketing segmentation, segmentation is generally based on demographics or customer value. However, the proposed segmentation approach is more accurate since it is based on real location data. As a result, market segmentation based on customer location data propose a high potential for segmentation and get insight about each individual customer before implementing personalized advertising recommender systems.

References

Aliahmadipour, L., Torra, V., Eslami, E.: On hesitant fuzzy clustering and clustering of hesitant fuzzy data, fuzzy sets, rough sets, multisets and clustering. In: Madani, K. (ed.) Computational Intelligence, vol. 671, pp. 157–168. Springer International Publishing, New York (2017)

Tussyadiah, I.P.: A concept of location-based social network marketing. J. Travel Tour. Mark. **29**(3), 205–220 (2012)

Xu, H., Oh, L.-B., Teo, H.-H.: Perceived effectiveness of text vs. multimedia location-based advertising messaging. Int. J. Mobile Commun. **7**(2), 154–177 (2009)

Lee, S., Kim, K.J., Sundar, S.S.: Customization in location-based advertising: effects of tailoring source, locational congruity, and product involvement on ad attitudes. Comput. Hum. Behav. **51**, 336–343 (2015)

Shin, W., Lin, T.T.C.: Who avoids location-based advertising and why? Investigating the relationship between user perceptions and advertising avoidance. Comput. Hum. Behav. **63**(2016), 444–452 (2016)

Junglas, I.A., Watson, R.T.: Location-based services. Commun. ACM **51**(3), 65–69 (2008)

Wu, C.H., Kao, S.C., Wu, C.C., Huang, S.: Location-aware service applied to mobile short message advertising: design, development, and evaluation. Inf. Process. Manag. **51**(5), 625–642 (2015)

Dao, T.H., Jeong, S.R., Ahn, H.: A novel recommendation model of location-based advertising: context-aware collaborative filtering using GA approach. Expert Syst. Appl. **39**(3), 3731–3739 (2012)

Abbas, R., Michael, K., Michael, M.G.: The regulatory considerations and ethical dilemmas of location-based services (LBS): a literature review. Inf. Technol. People **27**(2014), 2–20 (2014)

Pingley, A., Yu, W., Zhang, N., Fu, X., Zhao, W.: A context-aware scheme for privacy-preserving location-based services. Comput. Netw. **56**(2012), 2551–2568 (2012)

Fan, S., Lau, Raymond Y.K., Leon Zhao, J.: Demystifying big data analytics for business intelligence through the lens of marketing mix. Big Data Res. **2**(1), 28–32 (2015)

Gavalas, D., Konstantopoulos, C., Mastakas, K., Pantziou, G.: Mobile recommender systems in tourism. J. Netw. Comput. Appl. **39**, 319–333 (2014)

Lin, T.T.C., Paragas, F., Goh, D., Bautista, J.R.: Developing location-based mobile advertising in Singapore: a socio-technical perspective. Technol. Forecast. Soc. Chang. **103**, 334–349 (2016)

Schilke, S.W., Bleimann, U., Furnell, S.M., Phippen, A.D.: Multi-dimensional-personalization for location and interest-based recommendation. Internet Res. **14**(5), 379–385 (2004)

Unni, R., Harmon, R.: Perceived effectiveness of push vs. pull mobile location-based advertising. J. Interact. Advert. **7**(2), 28–40 (2007)

Wei, R., Xiaoming, H., Pan, J.: Examining user behavioral response to SMS ads: Implications for the evolution of the mobile phone as a bona-fide medium. Telemat. Inform. **27**(1), 32–41 (2010)

Zhang, X., Edwards, J., Harding, J.: Personalized online sales using web usage data mining. Comput. Ind. **58**, 772–782 (2007)

Park, D.H., Kim, H.K., Choi, Y., Kim, J.K.: A literature review and classification of recommender systems research. Expert Syst. Appl. **39**, 10059–10072 (2012)

Öztayşi, B., Gokdere, U., Simsek, E., Oner, C.: A novel approach to segmentation using customer locations data and intelligent techniques handbook of research on intelligent techniques and modeling applications. In: Kumar, A., Dash, M.K., Trivedi, S.K. (eds.) Marketing Analytics, pp. 21–39. IGI Global, Hershey (2016)

Ortega, F., Bobadilla, J., Hernando, A., Gutiérrez, A.: Incorporating group recommendations to recommender systems: alternatives and performance. Inf. Process. Manag. **49**, 895–901 (2013)

Liu, H.: Survey of wireless indoor positioning techniques and systems. IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. **37**(6), 1067–1080 (2007)

Oztaysi, B., Isik, M.: Supplier evaluation using fuzzy clustering. In: Kahraman, C., Oztaysi, B. (eds.) Supply Chain Management Under Fuzziness: Recent Developments and Techniques, pp. 61–80. Springer, New York (2014)

Oztaysi, B., Onar, S.C.: User segmentation based on twitter data using fuzzy clustering: the case of sports news. In: Bhatnagar, V. (ed.) Data Mining in Dynamic Social Networks and Fuzzy Systems, pp. 316–333. IGI Global, Hershey (2014)

Ruspini, E.H.: Numerical methods for fuzzy clustering. Inf. Sci. **2**, 319–350 (1970)

Babuska, R.: Fuzzy and neural control disc course lecture notes (2009). http://www.dcsc.tudelft.nl/~disc_fnc/transp/fncontrol.pdf. Accessed 5 Oct 2012

Han, J., Kamber, M.: Data Mining Concepts and Techniques, pp. 5–33. Morgan Kauffman Publishers, Burlington (2001)

Theodoridis, S., Koutroumbas, K.: Pattern Recognition, 4th edn. Elsevier, Amsterdam (2008)

Chen, N., Xu, Z.S., Xia, M.M.: Hierarchical hesitant fuzzy K-means clustering algorithm. Appl. Math.-J. Chin. Univ. **29**, 1–17 (2014)

Zhang, W., Ju, Y., Liu, X.: Multiple criteria decision analysis based on Shapley fuzzy measures and interval-valued hesitant fuzzy linguistic numbers. Comput. Ind. Eng. Arch. **105**(C), 28–38 (2016)

Wang, J.Q., Wu, J.T., Wang, J., Zhang, H.Y., Chen, X.H.: Interval-valued hesitant fuzzy linguistic sets and their applications in multi-criteria decision making problems. Inf. Sci. **288**, 55–72 (2014)