A Recommendation System for Enhancing the Personalized Search Itineraries in the Public Transportation Domain

Aroua Essayeh and Mourad Abed
Univ. Lille Nord de France, F-59000 Lille, France
UVHC, LAMIH, F-59313 Valenciennes, France
CNRS, UMR 8201, F-59313 Valenciennes, France

Keywords: Clustering, Learning Machine, Ontology, Public Transportation.

Abstract: In traditional transport information systems, the users must explicitly provide the information related to both their profiles and travels to receive a personalized response. However, this requires, among others, an extra effort from user in term of search time. We aim to identify not only implicitly users' information, but also to anticipate their need even if some data are missing through a recommender system based on collaborative filtering technique. In this work, the information related to users is represented using the ontology which proved far more adequate model for representing semantically data.

1 INTRODUCTION

In the era of globalization, the development of the intelligent systems has seen a rapid evolution due to the emergence of new technologies. However, the information systems suffer nowadays from the proliferation of information. The information is very heterogeneous and provided from various sources. For this purpose, personalization plays a decisive role in information systems. In public transportation field, the main difficulty is to propose for a user the best itinerary that fit with his preferences and profile with reducing at the same time both search time and effort. Moreover, most of these systems do not tackle with the problem of missing information, in this case, the user is forced to fill and to complete all needed information. Furthermore, it will be more sophisticated to prognosticate their requirements instead of expressing them at every turn. Hence, this will save time when browsing for the best result, and will help novice user when finding their needs more easily.

To this end, we propose a new strategy of personalization based on reasoning on over two ontologies related to user profile and transport domain , we have to learn later from their past interaction with the system in order to reformulate the query and recommend a new personalized solution.

The paper has three main contributions. First, we propose to model users' profile by using the ontology and we include some properties that we considered important in the context of travel. Second, we use jointly, inferences rules and fuzzy clustering algorithm to anticipate user's needs implicitly even if some information seems important are missed according to the stored histories. Third, this algorithm is enhanced through a new dissimilarity measure to handle the problem of heterogeneous data.

The rest of this paper is organized as follows: We introduce first a background about the personalized information in public transport field. Second, we discuss the collaborative filtering approaches. Third, we announce our motivation in the next section. Then, we illustrate the different steps of our proposed algorithm. Finally, we show the experimental results and evaluations, we discuss some conclusions about the benefits and limitations of our approach and we outline some future works.

2 RELATED WORKS

2.1 Personalized Information System in Public Transport

As defined in (Hagen 1999), personalization is "The ability to provide content and services that are tailored to individuals based on knowledge about their preferences and behaviour". In public transportation field, the personalization has a

Essayeh, A. and Abed, M.
A Recommendation System for Enhancing the Personalized Search Itineraries in the Public Transportation Domain.
DOI: 10.5220/0006315904150423
In *Proceedings of the 19th International Conference on Enterprise Information Systems (ICEIS 2017) - Volume 1*, pages 415-423
ISBN: 978-989-758-247-9

significant consequence when searching for personalized itineraries. This itinerary must, among others, respond to users' preferences and needs. Many approaches have dealt with the personalization in this field. An MDA (Model Driven Architecture) approach is applied in (Marçal de Oliveira et al. 2013) in order to build a personalized itinerary for a user giving him at the same time some services related to his preferences. The weakness of this approach is in the manual mapping used between the domain ontology and the context model which let this system not applicable in real cases and remain a theoretic approach. In their study, (Moussa, Soui2 & Abed 2013) have introduced a multi-criteria decision making approach to personalize a system in public transportation. This method uses the ELECTRE method; it focuses on achieving a compromise between the different compensatory criteria. However, this work considers some quantitative information related to the travel and assumes implicitly that all criteria are considered fully comparable which is not always possible in complex systems. In their recent work, (Bouhana et al. 2015) have proposed a hybrid method based on CBR (Case Based Reasoning) and ontology to personalize the itinerary for stakeholders. Despite the efficiency of the adopted methodology, some drawbacks are still unresolved, among which the prediction of user's needs without involving their personal motivations.

2.2 Collaborative Filtering Approach

Besides the huge number of heterogeneous data, users have some difficulties to express clearly their needs in a significant timing. In practice, users need recommendations because they do not have enough knowledge to make an autonomous decision (Ricci , Rokach & Shapira 2015) or what is the response relative to their request.

Collaborative filtering techniques aim to perform personalized recommender system. Furthermore, several works are found in the literature and have investigated to identify recommender system that aim to provide a personalized content for a target user. As defined in (Liu et al. 2014), the recommender system's aims to "guess" the users' preferences by analysing their behaviours when interacting with the system. In other words, it reveals historic users to anticipate their needs. Nonetheless, such systems suffer from both scalability, and the cold start problem. This issue is addressed in the present paper. Moreover, these systems require the determination of the correlation between the user and the recommended service. This latter is then another challenge to mitigate. We identify the most known techniques used to deal with this issue such as the Pearson correlation technique (Resnick, Iacovou & Suchak 1994), the constrained Pearson correlation (CPC)-based similarity (Shardanand & Maes 1997), the cosine-based similarity (Sarwar et al. 2000), and the adjusted cosine based measures (Ahn 2008) to identify the similar users that rated the same services. Some approaches require aggregator model to enhance decision system. For example, a combination of the OWA (Ordered Weight Aggregator) and the LSP aggregator (Logic Scoring of Preference) used in (Moreno et al. 2013) to analyse the user's interactions with the system, and to identify the degree of interests to promote decision making. Choquet integral is also used in both (Bouhana et al. 2015) and (Bouhana et al. 2013) for making a decision. AHP and OWA aggregators are used in (Abolghasem Sadeghi & Kyehyun 2009) to investigate the decision maker.

The major key of a personalized process is to know what a user wants and also why he needs this (Lakiotaki & Matsatsinis 2011). To deal with this issue a single rating item for each item doesn't offer an appropriate understanding as it use a unique criteria to predict his needs. This issue is addressed by the use of a multi-criteria approach (Nilashi , bin Ibrahim & Ithnin 2014) (Liu, Mehandjiev & Xu 2011) (Lakiotaki & Matsatsinis 2011). These ones can provide more information about the user requirements.

To that extent, we need to build communities that gather users who shared common interests. We reason over these communities to derive similar behaviours and provide recommendations. In this light, clustering techniques have not only the goal to gather users into several groups, but also to construct communities to learn and to recommend latter similar solution. Without any prior knowledge, we aim to obtain overlapping clusters, and a single user would belong to more than one cluster. With hard classification, user is associated only to one cluster, while the soft classification allows a user to join many clusters with different degrees of membership. The most known method is the fuzzy K-means algorithm. Sometimes, it is combined with others data mining algorithms such as ANN (Artificial Neural Networks) presented in (Paireekreng & Wai Wong n.d.) in order to classify users respect to their demographics data and interests. The key drawback of this research study is that it treated with very limited resources relating to the mobile application (Paireekreng & Wai Wong n.d.). (Lazzerini & Marcelloni 2007) have applied a method gained on an unsupervised algorithm namely the Fuzzy Divisive Hierarchical Clustering (UFDHC) algorithm to classify the users of a web portal into groups with similar characteristics and interests represented by prototype. This research emphasizes only the content of web pages related to specific profiles. The Competitive Agglomeration for Relational Data (CARD) algorithm is used in (Gandy et al. 2005) to classify user's session. (Castellano et al. 2007)have focused on web personalization problem and especially on discovering what current webpage related to user profile. The clusters are generated by mining the log data of a web containing user's preferences. Moreover, (Teran & Meier 2010) have proposed a fuzzy recommender system for election field. Their work had to identify the most similar candidates according to voter's preferences and tendencies. They presented a modified fuzzy c-means algorithm. A similar work, (Jalali et al. 2010) have defined an architecture based on two phases to predict the user future requests. The first phase is turned on offline mode and it implemented the clustering model for navigation pattern mining; it consisted in computing the degree of connectivity between each pair of the Web pages, and then created an undirected graph to find the connect component. The online phase has to predict the user future intentions through mining Web server logs.

Table 1: Comparison between the fuzzy clustering approaches.

Approaches	Techniques	Data type
(Paireekreng & Wai Wong n.d.)	K-means and ANN algorithm	Categorical data
(Lazzerini & Marcelloni 2007)	Fuzzy Divisive Hierarchical Clustering (UFDHC)	characteristics and interests
(Gandy et al. 2005)	CARD algorithm	Relational Data
(Martin-Bautista et al. 2002)	fuzzy classification rules	Categorical data
(Teran & Meier 2010)	fuzzy c-means algorithm	Categorical data
(Jalali et al. 2010)	clustering model with indirect graph	Categorical and numerical data

The major challenge to consider with the use of fuzzy clustering algorithm is to distinguish, as much as possible, between the inter-clustering in order to obtain a best differentiated subset, and to minimize as much as possible the intra-class inertia, with the aim of obtaining the most homogeneous possible clusters.

2.3 Motivation

There is no doubt that personalized systems have been gaining interest in many domains and especially in the transport field. Despite the various methods and techniques presented in the literature, these approaches have weaknesses and limitations. The response must not only reply to user's request, but also it must anticipate his expectations before even he expresses them. In this setting, we intend to explore the histories of users to, one hand learns from their experiences and on the other hand, to handle with the problem of explicit or missed information. To this end, the techniques of machine leaning seem be the most appropriate to resolve this issue. We aim then to gather users who have common interests to learn from their experience. For this purpose, we opt for using the technique of fuzzy clustering and we propose a new dissimilarity measure to tackle the problem of heterogeneous data.

3 GENERAL PROPOSED FRAMEWORK

In this section, we describe the general proposed framework for enhancing personalized research. The aim of this proposal is to recommend for a current user the most similar response to his request. In this light, we split our proposal into two parts; the first concerns the modelling phase and the second part aims to respond to the user's request according to our recommender based model.

The figure 1 describes the input and the output of the proposed reasoning process.

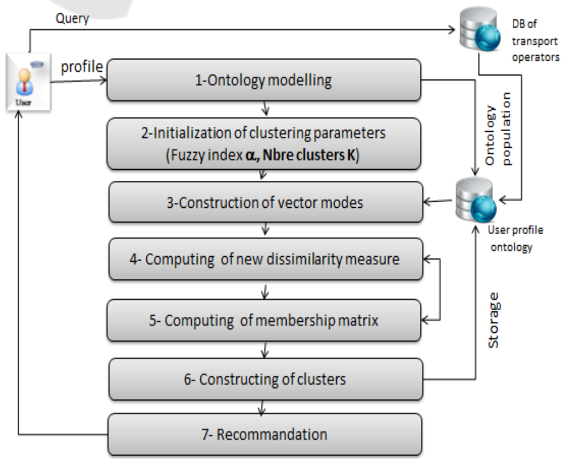

Figure 1: Reasoning process.

3.1 User Profile Modelling

This section presents the characteristics of the user profile ontology. The ontology was implemented using Protégé beta 5.0 [25] in OWL DL and its consistency tested using Jena and OWL Reasoner. The user profile ontology was created in order to facilitate the extraction of the user personal information, needs, and interests, under the context of personalization.

In this light, the user profile is defined as the union of:

$$P_u = P_S \cup P_P \cup P_C \cup P_{EX}$$

With:

P_s: represents his personal data (age, gender, address, ability, proficiency, user's state).

P_p: depicts the user's preferences for a precise service. The preferences may be related to the travel (cost, duration, walking, correspondence) or to the user's personal requirements (accommodation, administration, entertainment and healthcare).

P_{EX}: defines the users' histories (past query and validated choice).

P_c: defines contextual elements related to the user such as time and location.

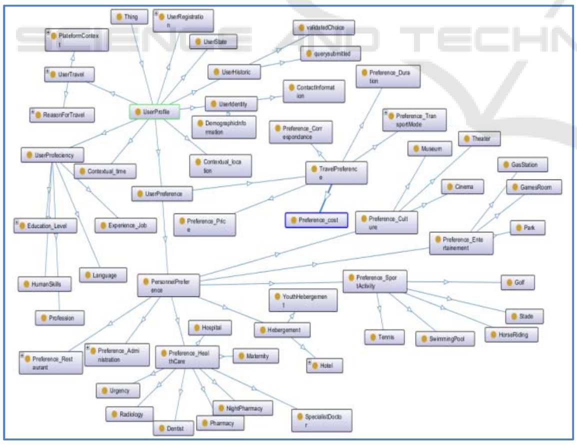

Figure 2: Top-level hierarchy of Ontology User Profile.

The figure 3 describes the main concepts of this ontology, the properties (represent the attributes) and the objects properties (used to establish relationships between classes)

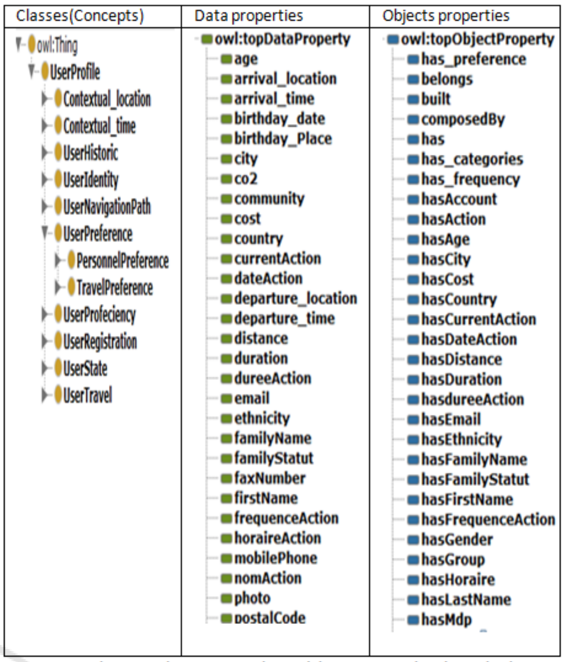

Figure 3: Classes, dataproperties, objectproperties, description.

3.2 Reasoning based Process

In this section, we have many challenges to overcome; First, unlike others approaches which used a few criteria to respond to user's request, we aim in this paper to compensate different types of criteria to obtain a personalized solution (itinerary). Further, these criteria are generally treated as numerical data, whereas it can be expressed in different forms. Even more, users' profiles are also different, and results change from a profile to another. In this view, we build our learning process. Accordingly, we propose the fuzzy k-mode algorithm (Huang & K. Ng 1999) within the inference engine to deal with the issues mentioned above.

3.2.1 Initialization of Clustering Parameters

We symbolize then the terminologies used in our clustering algorithm:

X,Y are two different users, $\xi = \{C_0, C_1, C_2, \dots, C_i\}$represent the classes related to user profile, xi={ x_1 ,x_2,..,x_N}: instances related to the first user and y_i={y_1,y_2,…,y_N} are the instances related to the second user, Cluster: is a group of users who share common characteristics (history, profile, context and preference), Reference Vector (V^{ref})represents the various concepts used in the input of our algorithm, Vector mode is specific to each cluster and composed of the instances related to the concepts of reference vector, Dissimilarity measure represents the degree of closeness between two individuals, Concept criterion(C^{crit})represents the concept topic in each cluster, Concept non-criterion (C^{Ncrit}) represents the other concepts which may influence the clustering process, Fuzzy index α, and finally, K is the number of the cluster. The fuzzy clustering algorithm is executed on 4 four step: 1) Selection modes, 2) Computing of dissimilarity measure, 3) Computing of membership matrix and 4) Upgrading clusters. In the next section, we outline the different steps of this algorithm.

3.2.2 Selection Vector Mode

The most straightforward approach is to choose arbitrarily k random modes, but it turns out that it is the weakest point of the conventional clustering algorithm; the fuzzy k-mode algorithm as an example (Choia & Chungb 2017). In addition, the correct choice of k is often ambiguous. For this purpose, several research efforts investigated the issue to more wisely choose this parameter as discussed in (Kodinariya & Makwana 2013). For the initialization process, we choose k according to the theorem of rule of thumb which is also similar to heuristic rule (Madhulatha 2012). Thereupon, k is calculated as following (n is the number of users) Eq.1

$$K \approx \sqrt{(n/2)} \tag{1}$$

First, the reference vector is composed of classes from user profile. The choice of vector mode depends on this vector, and it defines the instances related to each class. For this purpose, we calculate the frequency of instances using the following Eq2.

$$F_r(xi)=\frac{n_{xi}}{N} \tag{2}$$

With n_{xi}: is the number of users that have the same instance in the cluster and N is the total number of users.

The selection of vector modes has two main conditions; first, the instances of each vector mode must not be duplicated in the other one and second, in case of missing values, each empty instance must follow our proposition of normalization method explained in detail later with Eq.4

3.2.3 New Dissimilarity Measure

In order to deal with different type of data, we propose a new dissimilarity measure. Further, the comparison between each two instances (xi; yi) is calculated based on their similarities and the type of their related concepts (Ci (xi); Ci (yi)). We consider for two individuals X and Y, the same reference vector. Therefore, as we have different types of data (numeric, non-numeric (textual and semantic)),

We assume that the concept criteria and non-criteria are determined according to their weight in the database as discussed in the next paragraph. Forthwith, we propose four discussed cases:

- Case1: if (Ci (xi); Ci (yi)) ∈ C^{crit} and xi = yi → xi and yi are two similar instances of the same concept (non-numeric). Their related concept is considered as a concept criterion. Example: Ci (xi) = Ci (yi) = "FamilyStatus"; xi = yi = "single".
- Case2: if (Ci (xi), Ci (yi)) ∈C^{Ncrit} and xi # yi → xi and yi are two similar instances of the same concept but this concept does not belong to the concept criterion. Example: Ci (xi) = Ci (yi) = "gender"; xi = "female" and yi = "male".
- Case3: if (Ci (xi); Ci (yi)) ∈{C^{crit}; C^{Ncrit}} and (xi; yi) are numeric values → xi and yi belong to the same concept. Example: Ci (xi) = Ci (yi) = "age"; xi = "13" and yi = "20".
- Case4: if (Ci (xi); Ci (yi)) ∈ C^{crit} and → xi and yi are two different instances but belong to the same concept. This latter belongs to the concept criterion. Example: Ci (xi) = Ci (yi) = "PreferenceTravel"; xi = "fast" and yi = "cheaper".

The enhanced similarity measure is then:

$$\delta\,(x_i, y_i) = \begin{cases} d_{min} & if\ \ (x_i = y_i)\ and\ \left(C_i(x_i), C_i(y_i)\right) \in \{C^{crit}\} : case1 \\ d_1 & if\ \ \left(C_i(x_i), C_i(y_i)\right) \in \{C^{Ncrit}\}: case2 \\ d_2 & if\ \ (x_i, y_i) \in IR\ and\ \left(C_i(x_i), C_i(y_i)\right) \in \{C^{crit}, C^{Ncrit}\}: case3 \\ d_{max} & if\ \ (x_i \neq y_i)\ and\ \left(C_i(x_i), C_i(y_i)\right) \in \{C^{crit}\} : case4 \\ & 0 \le d_{min} \le d_1 \le d_{max} \end{cases} \tag{3}$$

The similarity is calculated between two users' vector. Therefore, for case1 and case 2, we opt for using Jaccard similarity (Jeff M. 2013) for computing similarities between two strings or textual attributes, each attributes is associated to a weight. For this, we use the technique of TF-IDF[1]. With this method, we can define which concept (class) is a concept criterion. Each instance is associated with a weight w_i to evaluate how important this instances in database. For case3, we compute the similarity using the method proposed in (Bouhana et al. 2013) for numerical attributes. Afterwards, *OWA* aggregator (Yager 1988) is applied to compute the average similarity. The weights are generated automatically according to the *orness measure* and dispersion measure. The case 4 discusses uncertain criteria. For this propose, a WordNet[2] is applied to search the synonyms, the hypernym, the hyponym and any

[1]http://www.tfidf.com/

[2]https://wordnet.princeton.edu/

existing relation between the attributes of each users. However, in order to deal with missing information, a normalization technique is adopted. We assign a value (ε) for each empty attribute, we have then:

$$\theta_i = \begin{cases} 0 & si \quad x_i = y_i = \varepsilon \\ 1 & si \quad x_i \neq y_i \end{cases} \tag{4}$$

As a result, the distance D is defined then as the ratio of the sum of similarities measures between the attributes related to each individual and the sum of θ_i. The new formulation is given below:

$$D_i(X,Y) = \frac{\sum_{i=1}^{m} \delta\,(x_i, y_i)}{\sum_{i=1}^{m} \theta_i} \tag{5}$$

3.2.4 Membership Matrix Upgrading

The next step of our fuzzy clustering algorithm is to calculate and update the membership matrix $\omega = [\omega_{li}]$. The membership matrix allows the degree of user closeness with its corresponding cluster to be identified. This value is to be updated as far as we do not hit the stop condition.

The formula is expressed as follows:

$$\omega_{ij} = \begin{cases} 1 & if\ x_i = c_j \\ 0 & if\ \ x_i = c_k\ with\ k \neq 1 \\ \dfrac{1}{\sum_{k=1}^{c} \left(\frac{\delta(x_i - c_j)}{\delta(x_i - c_k)}\right)^{\frac{2}{\alpha-1}}} & if\ x_i \neq c_j\ and\ x_i \neq c_k \end{cases} \tag{6}$$

Where

X_i: is the set of user's attributes, C_i: represents the current mode concepts of a cluster which a user belongs to and C_k: represents the concepts of others modes which a user does not belong to.

3.2.5 User's Clustering Upgrading

The mode related to each cluster is not static since the algorithm did not reach the stop condition. To update the mode, we calculate the following formula:

$$z_j = \frac{\sum_{i=1}^{N} {\omega_{ij}}^{\alpha} * X_i}{\sum_{i=1}^{N} {\omega_{ij}}^{\alpha}} \tag{7}$$

Where:

ω_{ij} is the membership matrix for the *ith* iteration α is the fuzziness index and X_i is the instances of user X. This step continuous and the new mode is compared with the previous one until attending the stop condition. The stop condition is obtained if and only if the following objective function is minimized:

$$min(F(W,Z)) = \sum_{j=1}^{k} \sum_{i=1}^{n} \omega_{ij}^{\alpha}\ D(Z_j, X_i) \tag{8}$$

3.2.6 Recommendation Process

Hereinafter, we describe the execution of our recommendation process; the algorithm is described as follows:

```
Input
Mod0: initial vector mode
X={x1, x2,…., xn}: instances of user X
α = fuzzifier index
N=number of clusters
t =1 // number of iterations
ω(t) : Membership matrix
Z(t) : Vector mode upgraded

Output
Clusters= {CL1^P, CL2^P, CL3^P, …, CLn^P} //set of clusters

Begin
    //Initialize the fuzzifier index α
    N=Numbers_of_concepts_criterion
  //Construction of modes
  Foreach CLi^P in clusters do
    Foreach C^crit(x)
      Begin
    1) Initialize the vector mode (see Eq.(1 and 2))
    2) Calculate new similarity measure (see Eq.(3))
    3) Calculate new dissimilarity measure Di(Xi, Mod0) (see Eq.(5))
    IfDi(Xi, Mod0) is the less one then do
       X∈Clusteri // save user profile in cluster
       End If
    4) Calculate ω(t) based on Z(t) with Eq.(6)
       t++
    5) CalculateZ(t+1)based onω(t)withEq.(7)
    6) //Verify objective function (seeEq.(8))
          If (ω(t),Z(t+1))=F (ω(t),Z(t))thenStop.
       Else
     Go step (7)
     End if
    7) Determineω(t+1)based on Z(t+1)
         If F (ω(t+1),Z(t+1))=F (ω(t),Z(t+1))thenStop.
         Else Go step (4)
         End if
    8)   Repeat Step (4) to step (7), until there is no movement between clusters (minimize the objective function)
       End
     End
    Return similar user in current cluster
   Return recommended solution
   End
```

4 EXPERIMENTAL EVALUATION

In this section, we focus on the following three questions. One is to check whether such initialled parameters (fuzziness index and number of clusters) negatively affect the quality of resulting clusters (final solution). The other is to experimentally evaluate how our clustering algorithm is in agreement with other alternative algorithms and finally, we evaluate the proposed solution by addressing the following research question. How accurately the proposed approach can provide pertinent solution? This is addressed by calculating the precision and the recall. For this setting, a case study is presented in the next section.

4.1 Case Study

Personalized information systems in intelligent public transportation domain are complex systems and dealing with a large amount of heterogeneous data from various sources. These systems must respond to user's queries based on their requirements and also profiles. In fact, in traditional systems, user should provide all needed information explicitly to get a personalized response. However, user interacts with the system in order to search a personalized itinerary even if he doesn't dispose of all the travel's information; modality, time-tables and tariffs. In this setting, we recommend for him an itinerary in a multimodal network based on the deducted community of similar users. For the instance, we are interested only on the recommendation process. We consider a community of 14 users, every user executes 3 different queries, and we have at least 42 solutions. Based on the inferred knowledge provided by this community, we will recommend for the current user the most suitable itinerary.

To this end, ontology of the domain that describes the field of transport has been exploited and enriched by RATP OPEN DATA[3] and GTFS[4]. The figure 4 describes an extract of this ontology

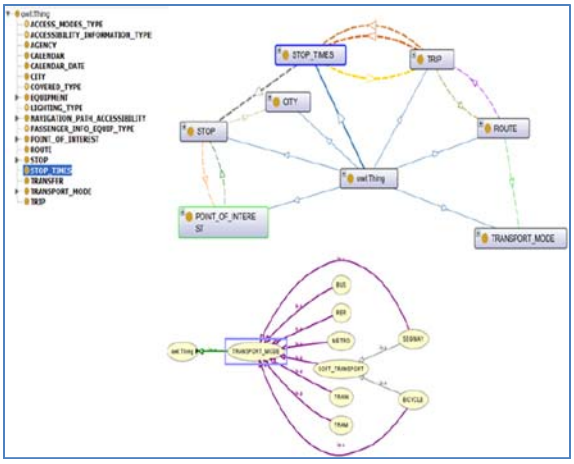

Figure 4: Classes, dataproperties, objectproperties, description.

4.2 Evaluation Results and Discussion

We have implemented our method in Java language and executed in Intel core i3 processor (1.9GHz) with 2GB memory running on windows 7 operating system.

- Clustering Quality

In order to justify the quality of clustering results, we use the Cohen's kappa value, which is widely used especially to measure the agreement between the proposed algorithm and the other competitors. The values between -1 and 1 assert that the accord is low, or very low, moderate agreement or strong agreement. The formula applied with Kappa value is:

$$K = \frac{P_0 - P_e}{1 - P_e} \tag{9}$$

With P_0: is the relative observed agreement among ratters and P_e is the hypothetical probability of chance agreement, using the observed data to calculate the probabilities of each observer randomly saying each category.

Table 2: Cohen's Kappa.

Results	New fuzzy K-modes Clustering	Fuzzy K-means clustering	K-modes clustering
New fuzzy K-modes Clustering	1	**0,67**	0,71
Fuzzy K-means clustering	**0,67**	1	0,4
K-modes clustering	0,71	0,4	1

[3] https://opendata.paris.fr/explore/

[4] https://developers.google.com/transit/gtfs/reference/

As shown in this table, the Kappa value is calculated for each pair of clustering techniques used. The new fuzzy K-modes clustering and the fuzzy K-means have good agreement strength of 0,67 (0,67∈ [0,6; 0,80], while new Fuzzy K-modes and K-modes clustering have an agreement strength of 0,71.

We admit that our proposed algorithm seems to show good results.

- Precision, Recall and F-measure

The Precision is defined by dividing the number of users correctly belonging to the positive cluster by the total number of users belonging to the positive cluster while Recall is calculated as the ratio between the number of users correctly to the positive cluster and the total number of elements that actually belong to the positive cluster. Indeed, the results is a set of clusters, we evaluate these measures by calculating the average of the precision and the recall in each clusters. Finally, we compute the F-measure for each query.

$$P = \frac{TP}{TP+FP}, R = \frac{TP}{TP+FN}, F = 2 * \frac{P*R}{P+R}$$

Table 3: Evaluation results.

Query	Estimated solution	Recommended Solution	Precision	Recall	F-measure
Q1	S8, S7, S27, S38	1) S12, S27, S42 2) S8, S15, S23, S30, S38 3) S14	0,33	0,75	0,46
Q2	S28, S13, S33	1) S13, S33, S28 2) S11, S26, S35, S37 3) S6, S21	0,33	1	0,50
Q3	S8, S28, S1, S30	1) S8, S15 2) S5 3) S13, S18 4) S20, S23, S30, S38	0,22	0,5	0,31

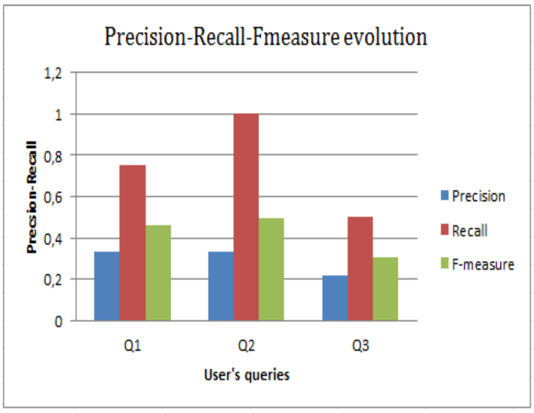

Figure 5: Precision, Recall and F-measure evolutions.

The found results presented in Fig. 5 shows that our algorithm returned most of the relevant results according to the high recall's values.

5 CONCLUSIONS

In this paper, we propose a new recommendation system based on fuzzy clustering algorithm jointly the ontologies in the public transportation field. The proposal supports both qualitative and quantitative data and aims to gather users who share common features into the same cluster. By building such clusters, called communities, we raise the problem of explicit information by learning from the similar profiles according to their interactions' histories with the system. The recommender solution fit his needs and responds to his formulated request. In our future work, we will detail how we reformulate the query sending by the user by adding the new implicit information inferred from the proposed clustering and we manage to use BabelNet to overcome the limits of the WordNet.

REFERENCES

Abolghasem Sadeghi, N & Kyehyun, K 2009, 'Ontology based personalized route planning system using a multi-criteria decision making approach', vol 36, no. 2250–2259.

Ahn, HJ 2008, 'A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem', vol Volume 178, Issue 1, 2 January 2008, Pages 37–51.

Bouhana , , Zidi, A, Chabchoub, & Abed, M 2015, 'An ontology-based CBR approach for personalized itinerary search systems for sustainable urban freight transport'.

Bouhana, , Fekih, A, Abed, M & Chabchoub, H 2013, 'An integrated case-based reasoning approach for personalized itinerary search in multimodal transportation systems'.

Castellano, , Mesto, , Minunno, M & Alessandra Torsello, 2007, 'Web User Profiling Using Fuzzy Clustering', 2007.

Choia, D-W & Chungb, C-W 2017, 'A K-partitioning algorithm for clustering large-scale spatio-textual data', Information Systems, pp. Information Systems 64 (2017) 1–11.

Gandy, L, Dept. of Comput. Sci., SIU,CIU, Rahimi, S & Gupta, B 2005, 'A modified competitive agglomeration for relational data algorithm', Fuzzy Information Processing Society, 2005. NAFIPS 2005. Annual Meeting of the North American, IEEE.

Hagen, 1999, 'mart Personalization," Forrester Report'.

Huang , Z & K. Ng, M 1999, 'A Fuzzy k-Modes Algorithm for Clustering Categorical Data', Journal IEEE Transactions on Fuzzy Systems, vol Volume 7 Issue 4, August 1999 Page 446-452.

Jalali, M, Mustapha, N, Sulaiman, MN & Mamat, 2010, 'WebPUM: A Web-based recommendation system to predict user future movements', vol Volume 37, Issue 9, September 2010, Pages 6201–6212.

Jeff M. , P 2013, 'Jaccard Similarity and Shingling', University of Utah.

Kodinariya, TM & Makwana, PR 2013, 'Review on determining number of Cluster in K-Means Clustering', vol Volume 1, Issue 6, November 2013, no. ISSN: 2321-7782.

Lakiotaki, & Matsatsinis, NF 2011, 'Multi-Criteria User Modeling in Recommender Systems', vol Vol26, no. 64 - 76.

Lazzerini, B & Marcelloni, F 2007, 'A hierarchical fuzzy clustering-based system to create user profiles'.

Liu , H, Hu, , Mian, A, Tian, H & Zhu, X 2014, 'A new user similarity model to improve the accuracy of collaborative filtering', vol 56 (2014) 156–166.

Liu, , Mehandjiev, & Xu, D-L 2011, 'Multi-Criteria Service Recommendation Based on User Criteria Preferences'.

Madhulatha, TS 2012, 'An overview on clustering methods,', CoRR abs/1205.1117.

Marçal de Oliveira, K, Bacha, F, Mnasser, H & Abed, M 2013, 'Transportation ontology definition and application for the content personalization of user interfaces', Expert Systems with Applications , www.elsevier.com/locate/eswa, pp. 3145-3159.

Martin-Bautista, MJ, Kraft, DH, Vila, MA, Chen, J & Cruz, J 2002, 'User profiles and fuzzy logic for web retrieval issues'.

Moreno , A, Valls , A, Isern , D, Marin, L & Borras, J 2013, 'SigTur/E-Destination:Ontology-based personalized recommendation of Tourism and Leisure Activities', vol 26(2013)633–651.

Moussa, , Soui2, M & Abed, 2013, 'User Profile and Multi-Criteria Decision Making: Personalization of Traveller's Information in Public Transportation', 17th International Conference in Knowledge Based and Intelligent Information and Engineering Systems -, Procedia Computer Science 22 (2013) 411 – 420.

Nilashi , M, bin Ibrahim, & Ithnin, N 2014, 'Multi-criteria collaborative filtering with high accuracy using higher order singular value decomposition and Neuro-Fuzzy system'.

Paireekreng, W & Wai Wong, K, 'Intelligent Mobile User Profile Classification forContent Personalisation'.

Resnick, , Iacovou, N & Suchak, M 1994, 'GroupLens: An Open Architecture for Collaborative Filtering of Netnews', no. pp. 175–186.

Ricci , , Rokach, L & Shapira, 2015, 'Introduction to Recommender Systems Handbook'.

Sarwar, , Karypis, G, Konstan, & Riedl, 2000, 'Item-based Collaborative Filtering Recommendation Algorithms', viewed 26 May 2016, <http://www10.org/cdrom/papers/519/node12.html>.

Shardanand, U & Maes, 1997, 'Social information filtering: algorithms for automating word of mouth,'.

Teran, & Meier, 2010, 'A Fuzzy Recommender System for eElections'.

Teran, & Meier, 2010, 'A Fuzzy Recommender System for eElections', *Electronic Government and the Information Systems Perspective, First International Conference, EGOVIS 2010, Bilbao, Spain, August 31 - September 2, 2010. Proceedings*.

Yager, RR 1988, 'On ordered weighted averaging aggregation operators in multi-criteria decision making', *IEEE Transactions on Systems, Man and Cybernetics*, vol volume 18 Issue 1, January/February 1988 , pp. Pages *183* - 190.