A Personalized Hybrid Tourism Recommender System

Conference Paper · October 2017

DOI: 10.1109/AICCSA.2017.12

Hela Masri
Université de Tunis

Saoussen Krichen
Institut Supérieur de Gestion de Tunis

A personalized hybrid tourism recommender system

*Abstract***—This paper focuses on building personalized recommender system in the tourism field. The application recommends to a tourist the best attractions in a particular place according to his preferences, his profile and his appreciation to previous visited places. This paper proposes a hybrid recommender system that combines the three most known recommender methods which are: the collaborative filtering (CF), the content-based filtering (CB) and the demographic filtering (DF). In order to implement these recommender methods, we have applied different machine learning algorithms which are the K-nearest neighbors (K-NN) for both CB and CF and the decision tree for the DF. The hybridization is a good choice to make the best of their advantages and to overcome the cold start problem. To enhance the recommendation accuracy, we use two hybridization techniques : switching and weighted. For the weighted approach, a novel linear programming model is applied to obtain the optimal weights' values. An extensive experimental study is conducted based on different evaluation metrics using extracted data from TripAdvisor. Our results show that the hybrid method is more accurate than the other recommender approaches used separately.**

*Index Terms***—Travel recommender system, machine learning, personalization, Hybrid recommender system, weighted hybridization, switching hybridization, Collaborative, Content-based, Demographic, Point of Interest.**

I. Introduction

The number of tourists in the world has increased in the last decades. The number of arrivals around the world in 2014 is 1,113 million and it's expected to be 1,800 million of arrivals in 2030 [1]. A tourist needs to plan his trip by selecting a destination and the different points of interest (POIs) to visit. He generally uses information from travel agencies, travel book guides and websites to organize his trip. Using internet, the tourist has an easy access to large amounts of travel information. The huge volume of available information about destinations, leisure activities and the previous reviews of other travelers turned the trip planning into a very challenging and time consuming task. The tourist gets eventually overwhelmed and he may have serious difficulties to discern the more interesting POIs from the rest. We should also note that the majority of websites are offering non personalized recommendations that are based on the number of visits of POIs (i.e. Visit Seattle website www.visitseattle.org) or the average rating given by previous users (i.e tripadvisor.com).

The Recommender Systems are developed to overcome these problems. They basically assist tourists in managing the big amount of information and to facilitate travel decisions [2]. In the tourism field, they are referred as Tourism Recommender Systems (TRSs). TRSs offer personalized information to users to improve the tourist experience. In other words, the system selects the more suitable and adequate POIs for a tourist by suggesting the most appropriate activities to their profile.

In a general context, RSs are classified in several types, based on the used data, the way they formulate recommendations and the algorithms they implement. The most known among them are [3]: CF, CB and DF. CF recommendation approach look at user interest (ratings and likes), it's based on the assumption that people who agreed in the past will agree in the future, and that they will like similar kinds of items as they liked in the past. CB recommendation approach focus on the item features to find good recommendations. The assumption behind CB approach is that people tend to like items that are similar to others that they have previously liked. DF recommendation approach uses information about demographic profile of the user such as the age, the gender and the country. Then, it recommends items preferred by users having similar features.

The RS where recently applied in tourism field. Most of the personalized TRSs are based on CB [2], [4], [5], DF [6], [7] and CF [8], [9]. Since each technique has some drawbacks, the combination of different approaches is also a widespread practice in the literature. The integration of these techniques can be done in different ways such as mixed [10], weighted [11] or switching [12].

In this paper, we propose to develop a new hybrid tourism recommender systems that combines three recommender filtering methods (CF, CB and DF) while using two hybridization techniques: switching and weighted. For the weighted technique, we propose an automatic approach to set the weights' values by applying a novel linear programming model. To validate our new method, a benchmark is built using crawled data from TripAdvisor, with a case study on the city of Paris. Based on different evaluation metrics, we proved that our method offers the best accuracy for the predicted rates. The combination of the two hybridization schemes improved considerably the results.

The rest of the paper is organized as follows. Section 2 discusses related works on TRSs. Section 3 presents the architecture of our proposed solution and details each component. The experimental results are presented in Section 4. Finally we conclude the paper in Section 5.

II. Tourism recommender systems

Recommender systems have been employed in the tourism sector providing many services, we can mention the suggestion of a destination or a tourist pack as done in Traveller [11], the suggestion of a trip plan [8], [13], [14] and the most utilized services in TRSs is the recommendation of attractions in a particular destination [9], [15]. Many platforms have been developed in the literature. They can be a desktop application, web application, and a mobile application.

Several recommender technique are used in the tourism field, we can mention the the CF, the CB, the DF and the hybrid method.

In CF approach, recommendation are made by using a group of users with similar preferences. The preferences are represented by likes and dislikes or by ratings. once the system identifies users having similar interest with the active users, it suggests to him items that preferred by those users. Generally determining the users with similar interest is done with K-NN algorithm like in GUIDME [9] which is a a mobile and web applications that recommend attraction to the tourist, according to his current location, preferences, and past visits. iTravel [16] is another mobile TRS that uses the CF approach to exchange rating between users and to predict the rate of the current user on non visited attractions.

The CB recommender compares the features of items that have not been rated to the active user with features of rated items to predict his preferences about the non rated items. for example CT-Planner 4 [17] gives a personalized recommendation to tourist by the CB approach, it determine user' rating about PoI, also the system e-Tourism [5] uses the CB approach to determine the tourist' preferences basing on semantic features of attraction.

The DF approach makes recommendation based on the user profile, it classifies users by their demographic information to predict at the final his rate or his like and dislike about an item. we can mention Wang et al., 2012 [7] who used demographic information and build a demographic recommender system by three machine learning algorithm: Naive Bayes, Bayesian network and Support vector machine (SVM). This system predict how a new tourist will rate an attraction. Results have shown that SVM method performed well on demographic information but results suggest that demographic information alone is not sufficient to do accurate prediction of ratings. There is other tourism recommender systems which use machine learning in uncertain case such as in Wang et al., 2011 [6] who develop a DF approach that recommends tourist attractions to users. The suggestion of activities and the computation of the probability of recommending the right activity are based on the traveler motivation and the traveler type, the traveler type also depends on user's age, occupation and personality.

The hybrid method try to avoid the limitation of each traditional recommender method by combining them, this method improves the prediction accuracy of rates. The system e-Tourism [14] creates plan according to the group of visitors' preferences. This system use an agent named General Recommender System Kernel to suggest activities in the city of Valencia to one or group of users. The personalization is obtained by mixed hybridization technique where they execute the CF approach, the CB approach separately then the results are combined. Turist@ [15] is a multi-agent-based system that provides personalized recommendations with a hybrid method. It combines CB and CF recommendation. The CB approach computes similarity between activities to obtain a final recommendation list and the CF uses the clustering method to generate users neighborhood. PSIS (Personalized Sightseeing Information System) [18] is also a hybrid TRS that suggests to tourist sightseeing places in the city of Porto, Portugal. This system group users into clusters by using CB, CF and DF, then it generates a set of rules with fuzzy algorithm, so that new users can be classified into many groups with different membership degrees.

The Hybrid weighted technique combines the results of different recommender methods and generates prediction by affecting weights to each used techniques by a linear formula. Traveller [11] uses this technique to combine the CF, the CB and the DF results.

Buddy@Move [12] is a hybrid travel recommender which combines the results of CB, CF and DF techniques. This system uses a switching techniques. It generates recommendation by DF for the new user cases and by the CB for the existing users cases and the CF is used to rank recommendation by using the users preferences. The CB, DF and CF recommendations apply the K-NN algorithm to predict ratings, the first uses the Pearson similarity to compute distances while the other use the Cosine similarity measures.

Our new hybrid recommender system is built on the data from TripAdvisor like in Wang et al., 2012 [7]. It combines the CF, the CB and the DF approaches by using two hybridization technique. Like in Buddy@Move [12], we use a switching technique that generates prediction in the case of new user by DF but in the case of new activity, our method generates prediction by CB and in the case of existing user and existing activity, we use a weighted technique that combines the CF, the CB and the DF results with a weighted sum formula.

III. THE PROPOSED APPROACH

Our study aims to develop a personalized recommender system that suggests to a tourist what are the most convenient PoIs in a particular destination, given his profile and his previous appreciations. The proposed TRS is hybrid as it combines the three recommender methods (CF, CB and DF) by using two hybridization schemes which are the weighted and switching approaches. The proposed approach benefits from advantages of each recommender method and overcomes their drawbacks.

Figure 1 shows the architecture of our proposed system, it contains five components: the data set generator, the three recommender systems and the hybrid engine. When a user log in the system, he must choose what type of activities to search (i.e. hotels, restaurants and destinations). Once he chooses the type, the hybrid engine generates a personalized recommendations according to the user's interests. The engine applies the three approaches (CF, CB and DF) and combines

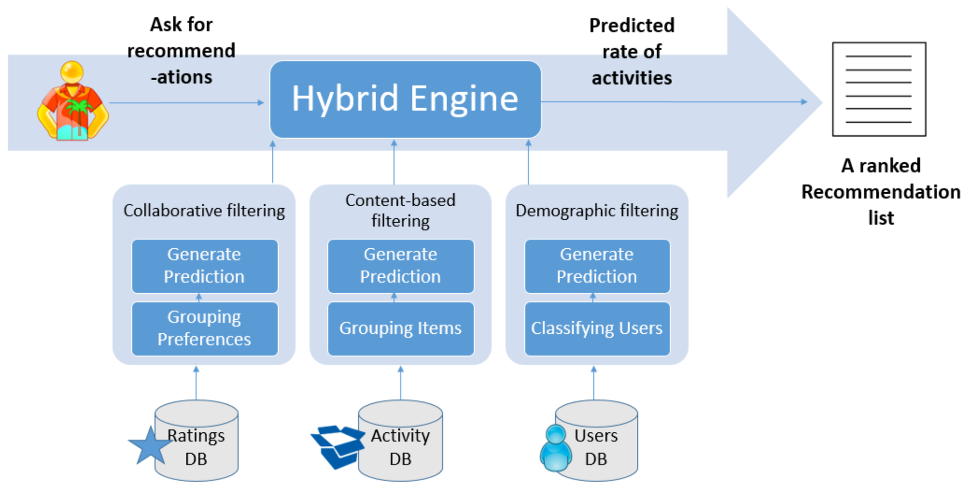

Fig. 1. The architecture of the proposed TRS

their results in order to provide a good prediction of the activities' rates. As an output, the user receives a final ranked list of recommended activities based on their predicted rates.

A. Data set

Our data were crawled from the e-tourism web site TripAdvisor (www.tripadvisor.com) through a web crawler named WebHarvy. This tools is able to navigate through web pages and converts its information presented in HTML form into a structured data. This web crawler is able to obtain product's information, user's information and user's review and evaluation of products. After collecting data by WebHarvy, data will then be exported automatically to Microsoft Excel with CSV format.

The extracted data needs to be pre-processed and cleaned so that it can be used by the recommender methods. We have used the Sql server integration services (SSIS) that transforms data from CSV format and load them into the database.

Our data is divided into three main tables :

- Users (id, login, age, gender, origin, region, travel style).
- Activities (id, activity name, activity category, activity price, Latitude, Longitude).
- Rantings (activity id, user id, rating).

B. Collaborative filtering

In our work, we have used the C# language and the NReco Recommender framework which is port of Apache Mahout CF engine, also we have chosen the user-based CF. This method aims to predict the active user' rates on the non rated products basing on the user' interest and the neighborhood interest. There are many similarity measure to compute the distance between two users in order to determine the neighborhood such as Uncentered Cosine similarity, Pearson correlation, Spearman correlation, Euclidean distance, Tanimoto coefficient and Log likelihood similarity [19]. After experimenting the performance of six similarity measure that will be presented in the next section, we chose the Tanimoto coefficient because it's the most accurate one in our data set. Its equation is given in (1) :

$$sim(i,j) = \frac{|f_i \cap f_j|}{|f_i| + |f_j| - |f_i \cap f_j|} \tag{1}$$

Where f_i is the set of item for which user i express preference, f_j is the set of item which user j express preference and $f_i \cap f_j$ is the intersection set of preferred item for user i and preferred item for user j.

The user -based collaborative filtering suffers from cold-start problem which appears when we have a new user that has no ratings enters into the system or when we have a new item that is no rated yet by users. In those cases the CF algorithm cant predict user's rates. This method contains two main steps:

- Computing interest similarity between the active user and all other users by using Tanimoto coefficient measure in Mahout.
- Searching neighbors users by the K-NN algorithm, here K is equal to 50.

C. Content-based filtering

In our work, we have used the C# language to implement the CB. In order to predict the rate of the user on one specific product, we used the nearest neighbor algorithm which computes the distance between the current item with all the rated item by the user. Each item is represented as a vector of features.

There is many distance measures between feature vectors that can be used to compute the similarity of two items, we have chosen the Euclidean distance formula (2) which was used and recommended by many existing tourism recommender system [15].

$$\delta = \sqrt{\sum_{k=0}^{n} (i_k - j_k)^2} \tag{2}$$

Where i_k and j_k are the feature k of item i and j. n is the number of features for item. This distance provides a value

between 0 and 1.
The CB overcomes the problem of the new item cold start problem but it suffers from the new user cold start problem where it can not provide results because there is no rated items by the active user to use in the comparison.
The main steps of this method are :

- Identify non rated activities by the active user.
- For each non rated activity, compute its similarity with all the rated activity by the active user by using the Euclidean distance similarity formula.
- Find for each non rated activity its nearest activity and predict its rate (the rate of the highest similar rated activity).

D. Demographic filtering

In our work, we have used the C# language to implement the DF and we have used the Accord.NET Framework [20] to build our decision tree.
The decision tree try to classify a user according to his information profile (age, gender, region, travel style)) to obtain his rate about one specific activity, so we assign the demographic information as the nodes and the ratings as the leafs.
There is many decision trees algorithm such as TDIDT, ID3, CART, C4.5, CHAID and MARS. We have chosen the ID3 decision tree proposed by Quinlan in 1986 [21] because it mostly used when we have discrete attributes and because it builds the fastest tree. ID3 suffers from the over-fitting or over-classification problem if a small sample is tested that's why we have pruned the constructed decision tree.
The DF approach overcomes the new user cold start problem that appears in the CF approach but it still suffers from the new item cold start problem where it cannot predict the rate of a new item that has no previous rates.
This method contains three main steps :

- Identify non rated activities by the active user.
- For each non rated activity create a decision tree that predict the rating.
- Apply each constructed decision tree on the active user profile to obtain his predicted rate.

E. Hybrid engine

In this paper, we propose a hybrid model that combines the three RS approaches: the CF, the CB and the DF. This method aims to overcome the drawbacks of each recommender approach used separately and especially the cold start problem. Moreover, this hybrid method tries to find the best combination of the cited approaches in order to increase the accuracy of prediction.
Each algorithm predicts the rate of one user u on one item i separately and the hybrid method combines all of them. This method takes into account the advantage of each one, for example, if we are on the case of a new activity where no one rated it yet then the CB will perform better than the other methods, and if we are on the case of a new user then the DF will perform better than the other methods.

To avoid the cold start problem and to take into account the advantages of each recommender approach, we have realized a particular hybrid method (Fig. 2) that uses a double hybridization techniques, a weighted hybridization combines the rating of users by the following formula (3):

$$\hat{r_w} = \alpha \cdot r_{\hat{DF}} + \beta \cdot r_{\hat{CB}} + \gamma \cdot r_{\hat{CF}} \tag{3}$$

Where, $r_{\hat{DF}}$ is the predicted rating using the DF approach, $r_{\hat{CB}}$ is the predicted rating using the CB approach and $r_{\hat{CF}}$ is the predicted rating using the CF approach. α, β and γ are fractions that represent the weight of each method.
A switching hybrid techniques switches between different recommender results in order to take advantage of each type at different situation and to take the best rating result. The switching techniques uses the weighted hybrid recommender result in the case of an existing user and an existing item situation, in the case of an existing user and a novel item situation, it uses the CB recommender result and in the case of a novel user and an existing item, it uses the DF recommender result. The Algorithm 1 summarizes the hybrid method.

Algorithm 1 Hybrid method

```
Input  : Dataset D, User u
for Each non rated item i do
    if Existing user And Existing item situation then
        r̂ = α · r_DF^ + β · r_CB^ + γ · r_CF^
    end
    else if New user And Existing item situation then
        r̂ ← r_DF^
    end
    else if Existing user And New item situation then
        r̂ ← r_CB^
    end
    Add r̂ to R
end
Output: Ratings R
```

The Algorithm 1 find the rates of all the non rated items by the active user in order to make recommendation. At every item it switches to the optimal solution as follows:

- Check whether no cold start situation is detected. If so, it use the average weighted sum of DF, CB and CF results.
- Check whether a new user cold start situation is detected. If so, it use the DF recommender result.
- Check whether a new item cold start situation is detected. If so, it use the CB recommender result.

In order to find the optimal and stable coefficient of the weighted technique, we did many experiments using the cross validation procedure. We proposed a new linear programming

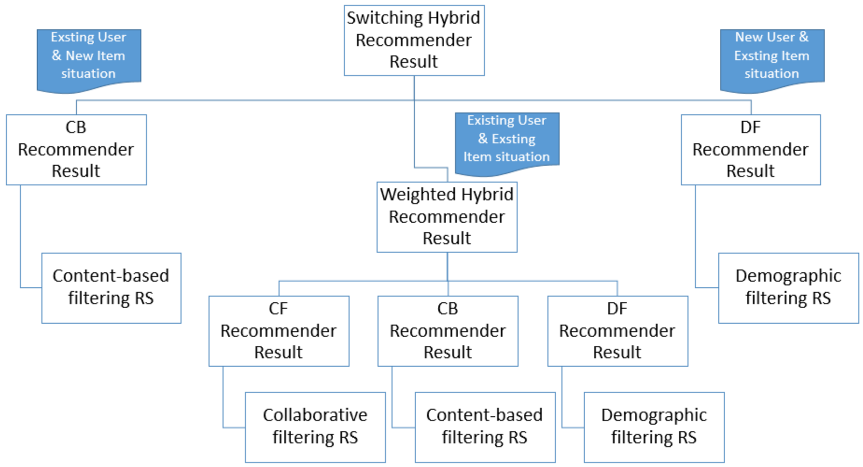

Fig. 2. Hybrid method

model (4) described as follows :

$$\begin{aligned} \text{minimize} \quad & \frac{\sum_{k=1}^{n} |\alpha \cdot \hat{r_{DF_i}} + \beta \cdot \hat{r_{CB_i}} \gamma \cdot \hat{r_{CF_i}} - y_i|}{n} \\ \text{subject to} \quad & \alpha + \beta + \gamma = 1; \\ & \alpha \succeq 0; \\ & \beta \succeq 0; \\ & \gamma \succeq 0. \end{aligned} \tag{4}$$

This linear programming problem minimizes the difference between the the weighted sum result $\alpha \cdot \hat{r_{DF}} + \beta \cdot \hat{r_{CB}} + \gamma \cdot \hat{r_{CF}}$ and the real value of prediction y_i, where n is the number of tested instances.

Using the cross validation, we got 10 subsets of training and testing sets. For each subset, we run our three recommender systems separately to compute the predicted values. Given the obtained results for each subset, the linear programming problem is then executed to generate the optimal values of the weights. The final coefficients to be included in our system, are the average of the optimal decision variables values of the 10 subsets. The final retained values are as follows $\alpha = 0,02, \beta = 0,83$ and $\gamma = 0.15$.

IV. Experimental results

In this section, we describe the implementation of our tourism recommender system and some scenarios of it.

A. Data set generation and pre-processing

We extracted data set from TripAdvisor website as the experimental data and we chose Paris as a destination because it's one of the most famous touristic destination in the world and because attractions in Paris have very large number of reviews in TripAdvisor.

Our data set contains $11,737$ reviews rated by 6576 users on 160 attractions.

In order to use the activities' attributes by the CB approach and to compute the Euclidean distance between the activities features, we have transformed the features to a vector with numeric value and the categorical attribute "activity category" is encoded into several binary attributes. Each activity can contains more than one category value from this selection (Points of Interest & Landmarks, Churches & Cathedrals, Historic Sites, Architectural Buildings, Monuments & Statues, Sacred & Religious Sites, Fountains, Educational sites). So we have transform each category value to one binary attribute where 1 represents the existence of the category and 0 represents the nonexistence of the category.

In TripAdvisor, there 20 possible styles for each activity, which are: Foodie, Beach Goer, Nature Lover, History Buff, Vegetarian, 60+ Traveler, Backpacker, Eco-tourist, Like a Local, Luxury Traveler, Trendsetter, Thrifty Traveler, Urban Explorer, Family Vacationer, Thrill Seeker, Art and Architecture Lover, Peace and Quiet Seeker, Shopping Fanatic and Nightlife Seeker. Each user can have more than one travel style. We merged these values into five groups in order to reduce the sparsity of our data set. The five groups are the following: style 1 (Foodie and Vegetarian), style 2 (Beach Goer, Nature Lover, Eco-tourist and Backpacker), style 3 (History Buff, Art and Architecture Lover, Peace and Quiet Seeker), style 4 (Urban Explorer, Like a Local, Family Vacationer) and style 5 (Thrill Seeker, Shopping Fanatic, Nightlife Seeker, Trendsetter and Luxury Traveler). This five travel style attribute are transformed also into binary attribute.

B. Experimental Implementation

In order to evaluate our proposed system, we divided ratings data (Table I) into 10 parts and a 10-fold cross validation approach is applied where at each iteration, one part is considered as testing set and the other are considered as training set. As an evaluation metric, we selected the Mean Absolute Error (MAE) [22], which is defined as follows :

$$MAE = \frac{\sum_{i=1}^{n} |x_i - y_i|}{n} \tag{5}$$

where x_i is the predicted rating value by the i^{th} tourist, y_i is the real rate and n represents the number of activities with

TABLE I
EXAMPLE OF DATABASE RATINGS

User-ID	Activity-ID	Rate
2	1	5
3	1	5
4	1	5
5	1	5
5	2	3
6	1	5
6	2	4
6	4	5

real rate in testing set. The lower MAE is, the more accurate is the recommendation result.
We have used also a normalized version of the MAE to express errors as percentages of full scale [23], this metric is defined by :

$$NMAE = \frac{\sum_{i=1}^{n} |x_i - y_i|}{n \cdot (r_{max} - r_{min})} \quad (6)$$

where r_{max} is the maximum value of rating and r_{min} is the minimum value of ratings.
We have also applied the Root mean squared error (RMSE) metric [24], which is widely used in the literature for evaluating recommender systems. It's defined by :

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (x_i - y_i)^2}{n}} \quad (7)$$

The RMSE emphasizes large errors by shedding a more severe penalization over them compared to the other metrics.

C. Results and Discussion

In this subsection, we present the experimental results of our method in generating ratings predictions.
In order to obtain the best similarity measure to be applied in the CF approach, we performed experiments over the 10 subsets. We compute for each subset its MAE measure and also we compute the MAE criterion for the averages. The Table II and the Figure 3 represent the obtained results.

TABLE II
MAE OF CF WITH FIVE DIFFERENT SIMILARITY MEASURE

Subsets	LL	Pearson	Spearman	Tanimoto	ED	Cosine
Subset 1	0,78	2,85	2,33	0,77	0,79	0,89
Subset 2	0,86	2,84	2,07	0,84	0,92	0,97
Subset 3	0,76	2,97	2,24	0,76	0,85	0,89
Subset 4	0,81	2,92	2,23	0,77	0,87	0,96
Subset 5	0,80	2,97	2,30	0,77	0,89	0,92
Subset 6	0,81	2,87	2,22	0,73	0,85	0,89
Subset 7	0,74	2,85	2,26	0,73	0,79	0,87
Subset 8	0,72	2,94	2,32	0,74	0,75	0,82
Subset 9	0,75	2,90	2,22	0,72	0,80	0,86
Subset 10	0,72	3,03	2,39	0,70	0,75	0,79
Average	**0,78**	**2,91**	**2,26**	**0,75**	**0,82**	**0,89**

We note through Table II that the Log likelihood (LL) CF and the Tanimoto Coefficient CF have the lowest value of MAE among the 10 testing subsets compared to the Cosine CF, the Euclidean distance, the Pearson CF and the Spearman CF.

Figure 3 shows the average MAE of each different CF approach. Overall, the Tanimoto Coefficient CF achieves mostly better results according to our data set comparing to the five other approaches with an average of 0.75 compared to 0.78 for the Log likelihood CF, 0.82 for the Euclidean distance 0,89 for the Cosine CF, 2,26 for the Spearman CF and 2,91 for the Pearson CF. Given these results, we opted for the Tanimoto Coefficient as a similarity measure of our CF approach.

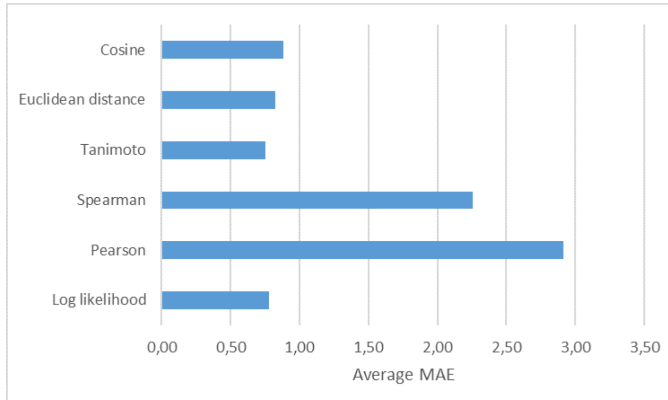

Fig. 3. The Average MAE in CF with five different similarity measure

In order to evaluate the proposed Hybrid method, we use three precision criteria which are MAE, NMAE and RMSE. We used the 10 fold cross validation and we computed the different criteria for each testing subset.
We compared our new hybrid method with the CF, the CB, the DF, the Switching Hybrid and the Weighted Hybrid methods. The weighted method consists on combining results of the three recommender approaches (CF, CB and DF) by a linear formula : $\alpha \cdot r\hat{_{DF}} + \beta \cdot r\hat{_{CB}} + \gamma \cdot r\hat{_{CF}}$ like in Traveller system [11]. The weights value are the same that we have obtained in the previous section. As used in Buddy@Move [12]. The switching hybrid method select among the current situation what recommender approach to use. It uses the DF for the new user and existing item situation, the CB for the existing user and new item situation and for the existing user and existing item situation it uses the CF approach.
The obtained results show that our hybrid recommender system presents a lowest values of MAE, NMAE and RMSE (figures 4, 5 and 6) than all the other used recommender methods.
Through these experiment results, we can notice that the double hybridization (switching and weighted) showed a better performance in term of accuracy then the other recommender approaches.

V. CONCLUSION

In this paper, we have presented an intelligent tourism recommender system that recommends to the tourist activities

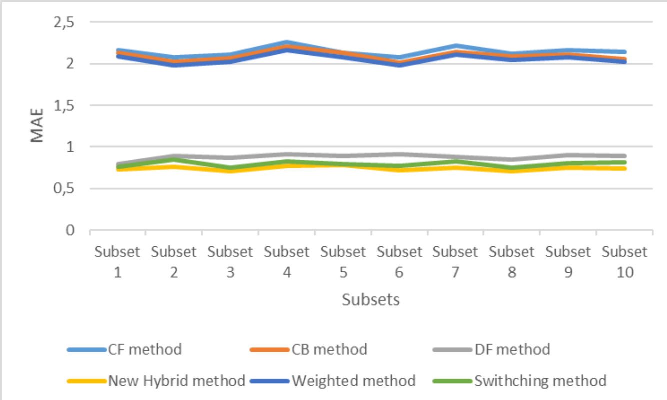

Fig. 4. The performances of the six RS in terms of MAE

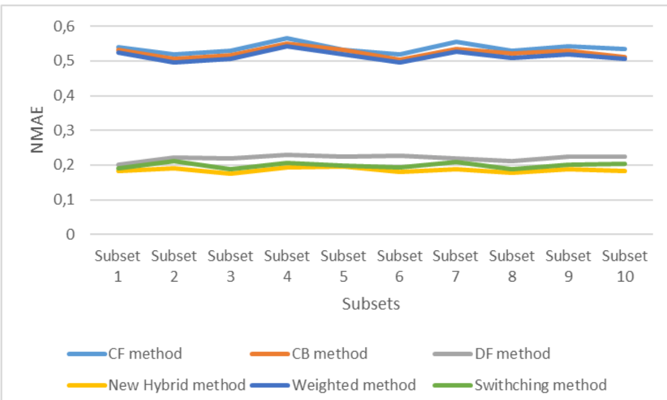

Fig. 5. The performances of the six RS in terms of NMAE

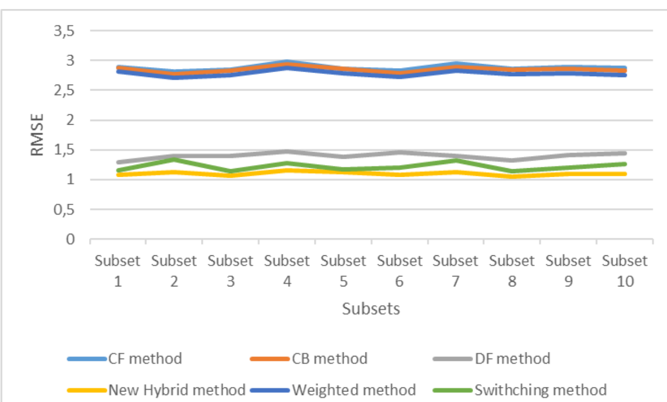

Fig. 6. The performances of the six RS in terms of RMSE

in a particular destination on a personalized way. We extracted data from TripAdvisor and we chose Paris as a destination to recommend its attractions for tourists. The recommender system is based on a hybrid approach which combines the CF, the CB and the DF methods, we use a double hybridization techniques : the switching and the weighted technique. Based on different evaluation metrics, the experimental results showed clearly that the hybrid method give more accurate prediction than the other methods used separately. As a future work, we are planning to add a new feature to our system, in order to generate the tour plan given the selected best rated attractions. This new service will be modeled as a vehicle routing problem.

References

[1] A. Holden, *Environment and tourism*. Routledge, 2016.

[2] F. Ricci, "Travel recommender systems," *IEEE Intelligent Systems*, vol. 17, no. 6, pp. 55–57, 2002.

[3] F. Ricci, L. Rokach, and B. Shapira, *Introduction to recommender systems handbook*. Springer, 2011.

[4] F. Ricci, Q. N. Nguyen, and O. Averianova, "Exploiting a map-based interface in conversational recommender systems for mobile travelers," *Tourism Informatics: Visual Travel Recommender Systems, Social Communities, and User Interface Design: IGI Global, Information Science Reference*, pp. 73–93, 2009.

[5] Á. García-Crespo, J. L. López-Cuadrado, R. Colomo-Palacios, I. González-Carrasco, and B. Ruiz-Mezcua, "Sem-fit: A semantic based expert system to provide recommendations in the tourism domain," *Expert systems with applications*, vol. 38, no. 10, pp. 13 310–13 319, 2011.

[6] W. Wang, G. Zeng, and D. Tang, "Bayesian intelligent semantic mashup for tourism," *Concurrency and Computation: Practice and Experience*, vol. 23, no. 8, pp. 850–862, 2011.

[7] Y. Wang, S. C.-F. Chan, and G. Ngai, "Applicability of demographic recommender system to tourist attractions: a case study on trip advisor," in *Proceedings of the The 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology-Volume 03*. IEEE Computer Society, 2012, pp. 97–101.

[8] F.-M. Hsu, Y.-T. Lin, and T.-K. Ho, "Design and implementation of an intelligent recommendation system for tourist attractions: The integration of ebm model, bayesian network and google maps," *Expert Systems with Applications*, vol. 39, no. 3, pp. 3257–3264, 2012.

[9] A. Umanets, A. Ferreira, and N. Leite, "Guideme–a tourist guide with a recommender system and social interaction," *Procedia Technology*, vol. 17, pp. 407–414, 2014.

[10] I. Garcia, L. Sebastia, and E. Onaindia, "On the design of individual and group recommender systems for tourism," *Expert systems with applications*, vol. 38, no. 6, pp. 7683–7692, 2011.

[11] S. Schiaffino and A. Amandi, "Building an expert travel agent as a software agent," *Expert Systems with Applications*, vol. 36, no. 2, pp. 1291–1299, 2009.

[12] S. Renjith and C. Anjali, "A personalized mobile travel recommender system using hybrid algorithm," in *Computational Systems and Communications (ICCSC), 2014 First International Conference on*. IEEE, 2014, pp. 12–17.

[13] P. Vansteenwegen, W. Souffriau, G. V. Berghe, and D. Van Oudheusden, "The city trip planner: an expert system for tourists," *Expert Systems with Applications*, vol. 38, no. 6, pp. 6540–6546, 2011.

[14] L. Sebastia, I. Garcia, E. Onaindia, and C. Guzman, "e-tourism: a tourist recommendation and planning application," *International Journal on Artificial Intelligence Tools*, vol. 18, no. 05, pp. 717–738, 2009.

[15] M. Batet, A. Moreno, D. Sánchez, D. Isern, and A. Valls, "Turist@: Agent-based personalised recommendation of tourist activities," *Expert Systems with Applications*, vol. 39, no. 8, pp. 7319–7329, 2012.

[16] W.-S. Yang and S.-Y. Hwang, "itravel: A recommender system in mobile peer-to-peer environment," *Journal of Systems and Software*, vol. 86, no. 1, pp. 12–20, 2013.

[17] Y. Kurata and T. Hara, "Ct-planner4: Toward a more user-friendly interactive day-tour planner," in *Information and communication technologies in tourism 2014*. Springer, 2013, pp. 73–86.

[18] J. P. Lucas, N. Luz, M. N. Moreno, R. Anacleto, A. A. Figueiredo, and C. Martins, "A hybrid recommendation approach for a tourism system," *Expert Systems with Applications*, vol. 40, no. 9, pp. 3532–3550, 2013.

[19] S. Bagchi, "Performance and quality assessment of similarity measures in collaborative filtering using mahout," *Procedia Computer Science*, vol. 50, pp. 229–234, 2015.

[20] C. R. Souza, "The accord .net framework," 2013.

[21] J. R. Quinlan, "Induction of decision trees," *Machine learning*, vol. 1, no. 1, pp. 81–106, 1986.

[22] J. S. Breese, D. Heckerman, and C. Kadie, "Empirical analysis of predictive algorithms for collaborative filtering," in *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 1998, pp. 43–52.

[23] K. Goldberg, T. Roeder, D. Gupta, and C. Perkins, "Eigentaste: A constant time collaborative filtering algorithm," *information retrieval*, vol. 4, no. 2, pp. 133–151, 2001.

[24] J. Bennett, C. Elkan, B. Liu, P. Smyth, and D. Tikk, "Kdd cup and workshop 2007," *ACM SIGKDD Explorations Newsletter*, vol. 9, no. 2, pp. 51–52, 2007.