HYBRID RECOMMENDER SYSTEM FOR PERSONALIZED POI SELECTION

Božidara Cvetković[1, 2], Boštjan Kaluža[1, 2], Hristijan Gjoreski[1, 2], Mitja Luštrek[1, 2]
Jožef Stefan Institute – Department of Intelligent Systems[1]
Jamova 39, 1000 Ljubljana, Slovenia
International postgraduate School Jožef Stefan[2]
Jamova 39, 1000 Ljubljana, Slovenia
e-mail: {boza.cvetkovic, bostjan.kaluza, hristijan.gjoreski, mitja.lustrek}@ijs.si

ABSTRACT

An important phase of trip planning is the selection of relevant points of interest. Many recommender systems have been developed to assist in travel planning, but only few of them take into account user's preferences. This paper presents preliminary results of the hybrid recommender system which first, filters out the points of interest according to user preferences and second, predicts the attractiveness of unrated points of interest using a combination of expert rate, knowledge-based and collaborative filtering recommendation approach.

1 INTRODUCTION

The tourism industry in the European Union (EU) has increased its economic importance in the last 50 years [1]. It has become one of the most significant financial contributor to the budget (4 -11% GDP) of the EU and also individual countries. The recent statistics show that the crisis did not have an impact on holiday trips, but on number of days spent at one location [2]. This indicated that for the higher financial turnover, it is important to offer and provide information about the relevant attractions, the points of interest (POI) at certain location to the user.

The users usually spend a lot of time researching the travel guides and the web for places they would like to visit and things they would like to do and see before going on a trip. To facilitate the search and assist the user in creating suitable trip plan and choosing the relevant POIs, various trip planning web sites and services have been developed over the years [3-5]. Mostly, these sites contain a lot of helpful information about the POIs, but don't automatically recommend them according to the user's preferences. The recommendations are focused on choosing the appropriate flight or providing the list of the best rated POIs, hotels or restaurants according to the user's budget.

The effective way to overcome manual selection and reduce search complexity is by using recommender systems. The recommender systems are very popular in various domains such as book, news, music, movies etc. recommendation [6]. Main advantage of the recommender systems is the possibility of personalisation of search for each individual user. Personalisation involves matching the context in sense of the user specifics, preferences and history to infer on the selection procedure and provide relevant results.

There are various approaches to recommendations. In this paper we will overview the three major categories [6]: (i) the content-based approach; (ii) the collaborative filtering approach; and (iii) knowledge base approach. The content-based approach assumes that the user's behaviour is repeated under similar circumstances. The learner builds a model of user interests according to the users past behaviour. In domain of tourism this means, the predicted rate of POI will be influenced by the rates given by the user to similar POIs. The collaborative filtering is based on behaviour of groups of people under similar circumstances. In tourism, the rate of POI will be influenced by other like-minded people. There are two major categories of collaborative filtering: (i) memory-based collaborative filtering and (ii) model-based collaborative filtering. The memory-based collaborative filtering uses the nearest neighbour algorithm to determine similarity among the users. The neighbours' preferences influence the predicted rate of the new user. The model-based collaborative filtering uses a machine-learning algorithm to learn the model of ratings for known users. This model is afterwards used to predict the rates for the new user. The knowledge-based approach is based on some functional knowledge, such as expert rules, on how a certain item fits the needs of the type of user under certain circumstances.

Using recommender systems in tourism has become a popular research domain. There are many recommender systems focused on selecting the most suitable POIs. For example, the Traveller by Schiaffino et al. [7] uses collaborative filtering, demographic information and the content-based approach to make recommendation. The demographic information is used within collaborative filtering to determine similarity between two users. The rate is influenced by the previous rating of the similar users. The Huang et al. [8] uses tourism ontology and content-based approach based on Bayesian networks, thus using the past behaviour of the current user and other users. The Sepa system by Garcia-Crespo et al. [9] requires the user to explicitly define the preferences, interests and the type of places he/she likes to visit. The system connects to the

user's social network and utilizes the social information. It also uses real-time location via GPS as a feature in the recommender system. The user profile is built upon the information explicitly provided by the user and semantic information obtained from the social network. The user and the services are expressed in ontology like structures which allow the application of feature based similarity algorithms to be used. The recommender system algorithm is a combination of knowledge-based approach and content-based approach.

This paper presents the preliminary results of the hybrid recommender system used in the e-Turist project for selection of the most suitable POIs according to the user's specifics and preferences. The recommender system first, filters the POIs according to users' constraints and second, uses combination of expert rate, knowledge-based approach and memory-based collaborative filtering approach to predict the rate of the POI.

The rest of the paper is structured as follows. Section 2 presents the modules of the recommender system, Section 3 presents the experimental results and Section 4 concludes the paper.

2 HYBRID RECOMMENDER SYSTEM

The purpose of the e-Turist application is to provide a trip plan composed of the most suitable POIs according to the user's preferences and specifics. To obtain the demographic information of the user the user is asked to register before using the application; the preferences for the current trip plan are inserted before each plan creation as shown in Figure 1.

Figure 1: e-Turist application asking the user to specify the preferences of the tip.

The e-Turist recommender system is composed of three modules shown on Figure 2: (i) the constraints filtering module; (ii) the knowledge-based module based on knowledge-based approach; and (iii) collaborative filtering module based on memory-based collaborative filtering approach.

$$\text{rate}(\text{User}, \text{POI}_x)$$
$$=$$
$$f(\text{rate}_{KB}(\text{POI}_x), \text{rate}_{CF}(\text{POI}_x), \text{rate}_{EXPERT}(\text{POI}_x))$$
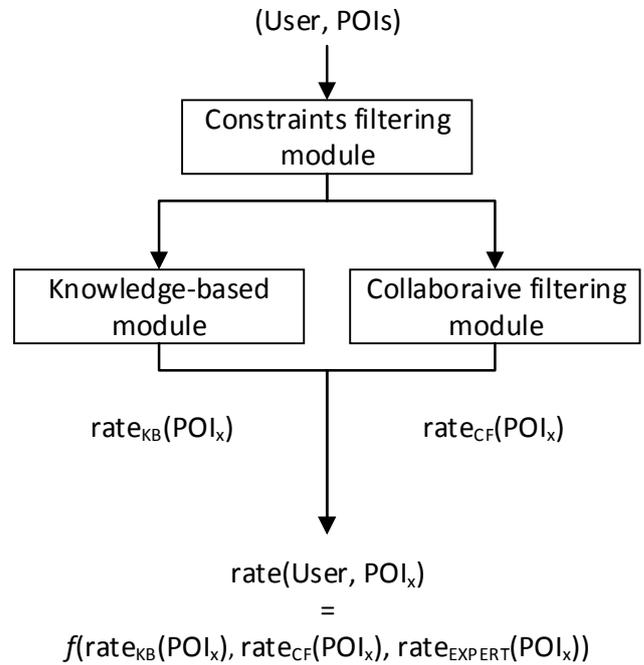
Figure 2: Workflow of the hybrid recommender system.

The final rate of hybrid recommender system is weighted sum of rates provided by the knowledge-based module, collaborative filtering module and expert rate provided by the experts for each POI. The expert rate is a constant.

2.1 The constraints filtering module

The constraints filtering module utilises "hard" constraints to keep only those POIs that satisfy the user's limitations. The hard constraints are: (i) the location; (ii) the purpose of the trip (active tourism, cultural heritage, gastronomy and entertainment); (iii) the working hours; and (iv) the mobility limitation of the user. The constraints filtering module returns POIs on a specified location that are open during the specified start of the trip and duration. The module filters out those POIs that are not categorised into purposes preferred by the user. In case the user has mobility limitations, the module filters out POIs that are not easily accessible.

2.2 The knowledge-based module

The knowledge-based module is based on knowledge-based approach. The approach is composed of expert rules that evaluate how a certain POI fits the needs of the type of user under certain circumstances.

The experts defined four sets of stereotypes that are important for the evaluation: (i) the age group; (ii) the education, (iii) the country of residence; and (iv) the budget. There are five age groups: (i) age up to 26; (ii) 27 to 36; (iii)

37 to 45; (iv) 46 to 55; and (v) 56 and higher. The education groups are three: (i) primary; (ii) secondary; and; (iii) tertiary. The budget groups are three: (i) low; (ii) medium and (iii) high. Each POI is categorised into one or more age groups, one or more education groups, one or more country and can have only one budget value. To evaluate the suitability the Euclidian distance is calculated between the user and POI characteristics. The final rate $rate_{KB}$ of the knowledge-based module is calculated using equation 1.

$$rate_{KB} = \frac{rate_{age}+rate_{edu}+rate_{country}+rate_{budget}}{4} \qquad (1)$$

2.3 Collaborative filtering module

The collaborative filtering is based on memory-based collaborative filtering approach. We used k-nearest neighbour algorithm [11] to find k similar users. Each instance represents one user. The feature vector is composed of rates per POI given by the individual user. In case the user did not rate the POI the value is defined as a missing value. The final rate for individual POI is an average value of rates per POI for the k-nearest neighbours.

3 EXPERIMENTAL SETUP AND RESULTS

To perform the experiment we collected data of 24 users with different age and background. The users were given a list of 90 POIs from Slovenian Istria, which are used by the e-Turist application. They were asked to rate given POIs they are familiar with from 1 to 5 stars. The data was used to evaluate each module separately and to define the final weights of the equation.

The goal of each module is to accurately predict the rate of POI as would be given by the current user. Result of each module was evaluated by mean absolute error (MAE) as defined by equation 2.

$$MAE = \frac{\sum_{i=0}^{n}(|rate_{true}-rate_{predicted}|)}{n} \qquad (2)$$

First, we evaluated the knowledge-based module. We used the demographic data and budget preference to predict the rate for each POI using the expert rules. An example of the expert rule for rating the suitability of POI according to budget is as follows:

IF user_budget notDefined:
    $rate_{bufget} = 0.5$
ELSE IF user_budget >= poi_budget:
    $rate_{budget} = 1$
ELSE:
    $rate_{budget} = 1 - \frac{abs(poi_{budget} - user_{budget})}{3}$

If the budget is not defined than the rate is 50% suitable otherwise the rate is normalised calculation of Euclidian distance between the budget specified by the user and the budget specified for the POI. The MAE of the knowledge-based module was 0.98 rate. The error can be translated into prediction accuracy of 75%.

Second, we evaluated the collaborative filtering module. The collaborative filtering is based on k-nearest neighbour algorithm. Before evaluation we had to define the number of neighbours that will be used by the algorithm. We tested the algorithm for k=1 to k=10. The results are shown in a graph in Figure 3. We can observe that higher the number of neighbours the lower the error. However, since we had data of only 24 users we had to limit the number of neighbours to the number from 1 to 5. The best results were obtained when k was set to 4.
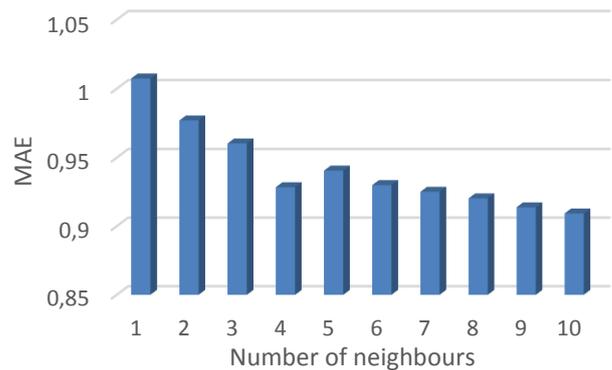
Figure 3: Number of neighbours and MAE of the collaborative filtering module.

The result of the collaborative filtering module using 4-nearest neighbours algorithm is MAE was 0.87 rate. The error can be translated into prediction accuracy of 78%.

The final result of the recommender system is calculated as a weighted sum of predictions of both modules and the expert value. The equation for the final result is presented as equation 3.

$$rate = w_1 * rate_{KB} + w_2 * rate_{CF} + w_3 * rate_{expert} \qquad (3)$$

The weights for each rate prediction were set based on the MAE for both modules and MAE of rate defined by the experts. The results are presented in Table 1.

Rate	MAE
Rate knowledge-based module	0.98
Rate collaborative filtering module	0.87
Expert rate	0.99

Table 1: Mean absolute error of knowledge-based, collaborative filtering module and expert rate.

We can observe that both modules perform with lower error as if the rate would be set equal to expert rate.

Therefore, the weights of both modules w_1 and w_2 should be set higher than the weight of the expert rate w_3. For the preliminary results, we decided the values of weights w_1 and w_2 should be equal and the weight w_3 lower. The results of the experiment defined the final algorithm and equations as follows:

IF $rate_{Expert}$ notDefined:
$$rate = 0.5 * rate_{KB} + 0.5 * rate_{CF}$$
ELSE:
$$rate = 0.4 * rate_{KB} + 0.4 * rate_{CF} + 0.2 * rate_{expert}$$

The result of the above equation was compared to a baseline approach. The baseline approach rates all POIs with rate 3, which is a medium rate. The results of all experiments are presented in MAE value and can be observed in Figure 4. The MAE value of the baseline approach is 1.05 rate and the MAE of the final rate is 0.86 rate, which is lower than MAE of both modules and the expert rate.
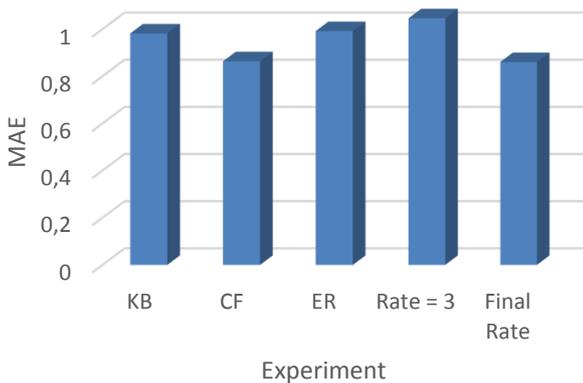
Figure 4: MAE value per each experiment the knowledge-based module (KB), the collaborative filtering module (CF), the expert rate (ER), the baseline approach, if rate is always equal to 3 (Rate=3) and the final rate calculated by the algorithm and equation above.

The result of the recommender system, the list of the most suitable POIs for the current user and the predicted rates, are afterwards processed with module for route planning, which is not a focus of this paper.

4 CONCLUSION

The paper presented hybrid recommender system consisting of knowledge-based module, collaborative filtering module and expert rate. The recommender system was used to predict the rates of tourist attractions or points of interest for individual user. The paper also presented the preliminary results of the recommendations. The experiments were performed on real users and data comprised from points of interest located in Slovenian Istria.

The results show that the collaborative filtering performs with lowest mean absolute error value compared to the other approaches. When the results of the three approaches (knowledge-based module, collaborative filtering module and expert rate) were combined, the mean absolute error of the predicted rate became a bit lower. The MAE of the final rate was 0.86 rate. Which can be translated into prediction accuracy of 79%.

References:

[1] The EU tourism industry, http://ec.europa.eu/
[2] Tourism statistics, http://epp.eurostat.ec.europa.eu
[3] Voyago, http://www.voyajo.com/
[4] TripAdvisor, http://www.tripadvisor.com/
[5] Fodors, http://www.fodors.com/
[6] F. Ricci, L. Rokach, B. Shapira. Introduction to Recommender Systems Handbook, Recommender Systems Handbook, Springer, 2011.
[7] S. Schiaffino, A. Amandi. Building an expert travel agent as a software agent, Expert Systems with Applications, 36(2), pp. 1291-1299, 2009.
[8] Y. Huang, L. Bian. A Bayesian network and analytic hierarchy process based personalized recommendations for tourist attractions over the Internet, Expert Syst. Appl. 36(1) , pp. 933-943, 2009.
[9] A. Garcia-Crespo, J. Chamizo, I. Rivera, M. Mencke, R. Colomo-Palacios, J. M. Gomez-Berbis. SPETA: Social pervasive e-Tourism advisor, Telemat. Inf., (26) 3, pp. 306-315, 2009.
[10] e-Tursit, http://www.e-turist.si/
[11] D. T. Larose. k-Nearest Neighbor Algorithm, Discovering Knowledge in Data, pp. 90-106, John Wiley & Sons, Inc, 2005.