EMPIRICAL EVIDENCE OF CONTRIBUTION TO E-TOURISM BY APPLICATION OF PERSONALIZED TOURISM RECOMMENDATION SYSTEM

Saso KOCESKI
Faculty of Computer Science
GoceDelcev University
Stip, Macedonia
saso.koceski@ugd.edu.mk

Biljana PETREVSKA
Faculty of Tourism and Business Logistics
GoceDelcev University
Stip, Macedonia
biljana.petrevska@ugd.edu.mk

Abstract

This article examines the necessity of applying recommendations by introducing web portal in the line of enriching e-tourism. It argues the dramatically change of tourist motives and interests due to the rapid usage of Internet. In order to cope with ever-changing tourist preferences and to contribute to tourism and economic development, a tourism recommendation system is proposed. Hence, a software module is developed based on collaborative filtering method which accuracy testing performed satisfactory outcomes. Moreover, acting as a national tourism web portal it enables to be introduced as a tool for assisting tourists in identification of a tailor-made tourist place. So, the study indicates a solution for generating personalized list of interesting tourism items by matching them to the tourists’ preferences. In this line, “MyTravelPal” is strongly recommended as a way-out in increasing the awareness for certain tourist destination that is capable of fulfilling travellers’ favorites.

Keywords: Internet, e-tourism, web portal, recommendation system
JEL classification: L83, L84, L86

1. INTRODUCTION

Everyone identifies tourism industry as a source of generating numerous positive impacts. Generally, tourism contributes to economic growth and development, promoting global community and international understanding and peace, providing tourism and recreational facilities to local people, improving living standards, stimulating local commerce and industry, reinforcing the preservation of heritage and tradition (Goeldner et al, 2000). The ground for enhancing all that lies in the quantity of tourists and travellers.

Yet, attracting a bigger number of tourists is not a trouble-free process, particularly in times of ever-changing travel preferences. The rapid development of
Internet, particularly in the past two decades, has changed tourism consumer behaviour dramatically (Mills and Law, 2004). It had an enormous impact on tourism industry, specifically to the way how tourists search for information. Moreover, the Internet has influenced tourism in significant manner by providing a great variety of services and products on-line (Kabassi, 2010). So, the Web became the leading source of information particularly important in times of increased number of competitors in tourism market. It was detected as the only way-out to be steady-ready to take prompt action. With the increased importance of search in travellers’ access to information, tourist destinations and businesses were forced to detect more adequate approaches to adapt to the fast-pace change in the environment (Pan et al., 2011). This particularly addresses the on-line tourism supply since tourist destinations have a strong need to acquire data for potential and present tourists and travellers. By the mediation of digital environment, it is noticeable the obvious tourists’ transformation from “passive audiences” to “active players” (Prahalad and Ramaswamy, 2000). A noteworthy transformation was made from just passive searching and surfing to creating content, collaborating and connecting. Hence, the development of the Internet empowered the "new" tourists who became knowledgeable and ask exceptional value for their money and time (Buhalis and Law, 2008). In this line, the web-booking systems gain in interest as a direction for detecting differences in the ways that active/passive tourists use the Internet for seeking different kinds of information, booking trips, paying and so forth.

Despite the variety of options regarding tourist destination or attraction, tourists frequently are not capable to cope with such a huge volume of choice. Moreover, they need advice about where to go and what to see. In e-tourism, recommendations may refer to indicate cities to go to, places to visit, attractions to see, events to participate in, travel plans, road maps, options for hotels, air companies, etc. Such scope of work is very often a robust and needs a facilitating factor. Then, a recommendation system is introduced with a main aim to assist the tourists in finding a way-out in the e-tourism “chaos”. The main idea of the recommendations is to contribute by facilitating personal selection and prevent tourists and travellers from being overwhelmed by a stream of superfluous data that are unrelated to their interest, location and knowledge of a place. So, it is much easier for tourists to access the information they need thus resulting in shorter lead-time for bookings, making last-minute decisions and generally, tailoring their own packages from a suite of options.

Solution is seen in personalization of the information delivery to each traveler, together with the travel history. Yet, the advanced tourist information systems must offer more than just relatively static information about sights and places. The way out is detected in application of recommendation systems as a promising way to differentiate a site from the competitors. So, user-generated content will gain in significance thus enabling developing more accurate recommendation systems.

Generally, the contribution of this paper lies in the fact that it enriches the poorly-developed empirical academic work within this scientific area in the Former Yugoslav Republic of Macedonia (FYROM). Additionally, the empirical investigation may alarm the relevant tourism-actors in the country, that the time has changed and that the on-line experience has shifted from searching and consuming to creating, connecting and exchanging. Previously passive consumers and web
surfers are now generating content, collaborating and commentating. So, this research proposes development of national tourism recommendation system since only if being prepared in due time, one may struggle the unexpected challenges.

2. LITERATURE REVIEW ON E-TOURISM AND TOURISM RECOMMENDATION SYSTEMS

The successful introduction of the Internet to e-tourism is fully supported by the search engines which became a dominant source in tourists’ use to access particular tourism and travel products. Due to its significance, this issue raised an interest within academia and practitioners. Generally, they argue regarding the understanding how search engines work and how travellers use the Internet and booking systems as tools in e-tourism (Morrison et al., 2001; Pan et al., 2007; Buhalis and Law, 2008; Pan et al., 2011; Xiang and Pan, 2010). Moreover, the success of search engine marketing requires a good understanding of consumer behaviour in order to provide the information desired by different consumers. Furthermore, the necessity of developing digital technology that will support the personalized services to address individual needs is fully justified. Tourism actors should collect customer information before, during and after a visit in order to better understand consumer behaviour choices and determinants (Buhalis and O'Connor, 2005).

Some researches address different approaches dealing with variety of relationships that appeared in e-tourism. So, Weber and Roehl (1999) explored demographics between Internet users and tourists at the same time. However, little research has been done on the travel-related behaviours of Internet travellers. In this respect, Morrison et al. (2001) found that some book travel online, while others go to travel agents or call the toll-free numbers of travel providers after getting travel information on-line. With regards to the behavioural dimensions, it may be utilized to segment travel markets as a powerful tool in managing e-tourism (Hennessey et al., 2008). Regardless the approach, it must be underlined that tourism needed this kind of information some years ago, while today we are faced with tourists with different travel patterns which cause different activity while travelling.

One may argue the inevitable relationship between tourism and information. Moreover, it is a widely-recognized fact that information and decision-making have become the foundation for the world economy (Wang, 2008). Due to the importance of tourism, recommendation systems applied in tourism have been a field of study since the very beginnings of artificial intelligence. It is a matter of identifying a class of intelligent applications that offer recommendations to travellers, generally as a response to their queries. They mostly leverage in-built logical reasoning capability or algorithmic computational schemes to deliver their recommendation functionality. So, the recommenders attempt to mathematically model and reproduce the process of recommendations in the real world.

Due to the rapid expansion of e-tourism, the tourism recommendation systems attracted a lot of interest in academia. In this respect, Mirzadeh et al. (2004), McSherry (2005) and Jannach (2006) elaborate the need for developing intelligent recommendation systems which can provide a list of items that fulfill as many requirements as possible. Ricci et al. (2002) and Wallace et al. (2003) discuss a
recommender system dealing with a case-based reasoning in order to help the tourist in defining a travel plan. However, as the most promising recommendation systems in the tourism domain are the knowledge-based and conversational approaches (Ricci and Werthner, 2002; Thomson et al, 2004). Yet, some other variants of the content-based filtering and collaborative filtering are engaged for recommendation, like knowledge-filtering, constraint-based and case based approaches (Kazienko and Kolodziejski, 2006; Ricci and Del Missier, 2004; Zanker et al, 2008). In the same line, the recommendation systems based on a text mining techniques between a travel agent and a customer through a private Web chat may easily be applied (Loh et al, 2004).

Some recent academia work refers to more sophisticated outcomes than the above noted. Namely, the introduction of a personalized tourist information provider as a combination of an event-based system and a location-based service applied to a mobile environment is suggested by Hinze et al (2009). In the other hand, an investigation on sources and formats of on-line travel reviews and recommendations as a third-party opinions in assisting travellers in their decision making during the trip planning is presented in the work of Zhang et al (2009). Noticeable are the findings regarding development of a web site in order to enable Internet users to locate their own preferred travel destinations according to their landscape preferences (Goossen et al, 2009). Furthermore, the usage of the orienteering problem and its extensions to model the tourist trip planning problem is elaborated as efficient solution for number of practical planning problems (Vansteenwegen and Wouter, 2011). It is evidently that the research area is extending resulting in improving dependability of recommendations by certain semantic representation of social attributes of destinations (Daramola et al, 2010). Moreover, most of the recommendation systems focus on selecting the destination from a few exceptions (Niarakiand Kim, 2009; Charou et al, 2010).

3. DATA AND METHODOLOGY

The research methodology was prepared in terms of creation of efficient and accurate personalized recommendation system based on novel algorithms. Specifically, it applies one of the most prevailing and efficient techniques to building recommender systems - collaborative filtering. This technique implements the idea for automating the process of “word-of-mouth” by which people recommend items to one another. It uses the known preferences of a group of users who have shown similar behavior in the past to make recommendations of the unknown preferences for other users. The main aim of the research was to develop and propose national tourism web portal which will assist and support tourists visiting the FYROM by helping them to identify relevant tourist objects that matches to their personal interests. The data was collected between October 2011 and January 2012, by the mixed research group composed of researchers from the faculties of Computer Science and Tourism at the “GoceDelcev” University from the FYROM. The data set was consisted of 9840 ratings from 265 users for 445 tourist objects. Additionally, each user rated at least 25 objects, and each object has been rated at least once.

To accomplish the main objective of the research, a several step methodology was developed. The first step foresees tourist and tourist objects profiling. The sys-
tem uses tourist types taken from the scientific tourism literature to model the tourist personal profile. The tourist profile indicates the degree to which tourists identify themselves with the given types. Typically, individual tourist cannot be characterized by only one of these archetypes but has unique combination of these personalities, although to varying degrees. Thus, tourist types model the tourists’ generic interests in an abstract form. Vectors are suited to model such tourist profile, whereby each dimension corresponds to a certain tourist type while the value indicates how much the tourist identifies him- or herself with the corresponding type.

4. ANALYSIS AND DISCUSSION

Tourist profiling is a two-step process which involves creating the profile and then reviewing the profile to make any necessary adjustments. The initial tourist profile for each system user is created by the user himself during the process of registration, by determining the degree of membership to each of the tourist types. Considering the fact that the human preferences change over time due to various factors, the tourists might change their behavior too. To make the system capable to cope with these changes, the tourist profile adjustment has been enabled. It is based on the ratings the tourist gives for each visited object (Eq. 1).

\[ \overline{U_{t+1}^i} = \overline{U_t^i} + R_{ik} \cdot w \cdot \overline{O_k} \]  (1)

where \( \overline{U_{t+1}^i} \) represents the profile vector of the i-th user in the moment of time \( t \) and \( U_i \in U \), \( U \) - is the complete set of users registered to the system. \( \overline{O_k} \) represents the profile vector of the k-th object in the set of all objects \( O \) registered in the system \( O_k \in O \), \( w \) - is the weighting factor and \( R_{ik} \) is the rating of the k-th tourist object given by the i-th user. The weighting factor (Eq.1) is used simply to prevent significant change of the tourist profile from a single rating. Very similarly, the tourism object is modeled through a vector as well, which quantitatively describes how much the object is related to the given types. In order to prevent information overloading, the system must recommend a particular subset of tourist objects in accordance to tourists’ personal preferences. The intention is to make a possibility for avoiding unpleasant tourists’ experiences.

The next step is introducing the matchmaking algorithm. The idea is to match the tourist profiles with the set of tourist objects. The result is a list of objects for each given tourist type. If a tourist profile matches the characteristics of an object, this object should be recommended to the respective tourist. The more similarities they have in common, the more contributes the tourist object to the tourist’s satisfaction. Therefore it should be ranked in higher position within the list. In order to estimate the similarity degree between tourist profiles and tourist objects, the system contains a special module based on a vector matchmaking function. For this purpose, a vector space model is applied, which dimensions fully correspond to the selected types of tourists noted in the respected literature (Gibson and Yiannakis,
Moreover, each tourist type represents one dimension in that space. The proposed matchmaking function is presented by Eq. 2.

\[
F_i(U_i, O_j) = \frac{\sum_{k=1}^{N} U_{ij_k} \cdot O_{jk}}{\sqrt{\sum_{k=1}^{N} U_{ij_k}^2} \sqrt{\sum_{k=1}^{N} O_{jk}^2}}
\]  

(2)

where \( U_{ij_k} \) is the degree of membership of the \( i \)-th tourist to the tourist type \( T_{kj} \), \( O_{jk} \) is the degree of membership of the \( j \)-th tourist object to the tourist type \( T_{kj} \), and \( N \) is the number of tourist types. Furthermore, tourist objects may be ordered by the value of the matchmaking function for a given tourist, and only those objects that have positive value for this function will be considered for recommendation.

The methodology is additionally enriched with a behavioral dimension. Namely, it is a common fact that before and during the travelling, people rely on recommendations from reference letters, news reports, general surveys, travel guides, and so forth. In addition, they desire personal advice from other people with similar preferences or people they trust. In fact, over 80% of travellers participating in a TripAdvisor.com survey agree that “reading other travellers” on-line reviews increases confidence in decisions, makes it easier to imagine what a place would be like, helps reduce risk/uncertainty, makes it easier to reach decisions, and helps with planning pleasure trips more efficiently” (Gretzel et al., 2007).

In this line, the authors introduce the collaborative filtering technique in order to develop a recommender. Yet, this technique is faced with many challenges, among which the ability to deal with highly sparse data and to scale with the increasing numbers of users and items, are the most important. So, in case of rarity and insufficiency of available ratings data, the sparsity problem occur which causes poor recommendation quality. In order to cope with all noted obstacles, the cloud-model collaborative filtering approach has been adopted. It is constructing the user’s global preference based on perceptions, opinions and tastes, which are subjective, imprecise, and vague (Palanivel and Siavkumar, 2010), and it seems to be an appropriate paradigm to handle the uncertainty and fuzziness on user preference.

The main goal of this model is to construct the global preference for each user (Zhang et al., 2009) by calculating a triple of three digital characteristics (Eq. 3). The first digital characteristic is the expected value (Ex) which represents the typical value of user ratings, that is, the average of user ratings. The second digital characteristic is the entropy (En) and represents the uncertainty distribution of user preference being measured by the deviation degree from the average rating. The third digital characteristic is the hyper-entropy (He) and is a measure of the uncertainty of the entropy En, which is measured by the deviation degree from the normal distribution.
The $k$ similar (neighbor) users, for an active user are selected based on the cloud model similarities between the active user and the users that already rated the object. In this regard, a likeness similarity function $F_2$ using the cosine measure was computed as in Zhang et al., 2009. Upon these two defined models, the authors define the similarity between the characteristic vectors. Furthermore, a total recommendation function for a given tourist object is calculated as a weighted average of the functions $F_1$ and $F_2$ (Eq. 4).

$$F_{rec} = (F_1 * w_1 + F_2 * w_2) / (w_1 + w_2)$$

where $w_1=0.7 > w_2=0.3$. Based on that, the objects are ordered and further classified into five categories (Eq. 5).

$$Cat_{k,j} = \begin{cases} 
  k = 1, \forall Oj \in O_{i_{rec}} \land 0 \leq Frec \leq 0.2 \\
  k = 2, \forall Oj \in O_{i_{rec}} \land 0.2 < Frec \leq 0.4 \\
  k = 3, \forall Oj \in O_{i_{rec}} \land 0.4 < Frec \leq 0.6 \\
  k = 4, \forall Oj \in O_{i_{rec}} \land 0.6 < Frec \leq 0.8 \\
  k = 5, \forall Oj \in O_{i_{rec}} \land 0.8 < Frec \leq 1 
\end{cases}$$

5. SETTING TOURISM WEB PORTAL

The developed national tourism web portal is structured in the form of a social network. Namely, it has a significant improvement compared to the existing national tourism and travel websites in the FYROM. One may find useful information on the first created national web tourism portal (www.exploringmacedonia.com) as well as on several other private initiatives which in the same line, act as additional tourism portals (www.travel2macedonia.com, www.go2macedonia.com, www.simplymacedonia.com, www.macedonialovesyou.com, www.mysticalmacedonia.com, www.macedonia-timeless.com, etc.). However, none of them provide the tourists and travellers with a customized, unique, and enriching travel experience as “MyTravelPal”. The main attribute that makes the biggest differentiation with the existing tourism web portals is that the new one incorporates some standard plugins typical for social networks, like Facebook, Twitter etc. So, it advances the concept by including custom plugins, like the recommended objects plugin which is the core of the portal. It uses the Google Map of the FYROM in order to visualize both: static and dynamic tourist objects. In this
regard, the static tourist objects are those that do not have temporary horizon, like churches, museums, archeology localities, etc., while the dynamic tourist objects are those that have limited time duration, like events, expositions, etc. Additionally, all tourist objects are respectfully displayed on the map in accordance to their geographical location. In order to gain more profound findings in an in-depth manner, they are geographically grouped into municipalities. The designed national tourism web portal consists of all 84 municipalities in the FYROM and thus is being recommended to the user in the form of circles (Figure 1).

The size of the circle indicates the user’s affinity for a particular municipality. So, the larger the circle - it matches more precisely the users profile. It means that a large circle indicates a municipality with many tourist objects with high recommendation marks. By displaying the user’s affinity through the size dimension of the circle, users can easily observe which municipalities would be of most interest to them.
“MyTravelPal” additionally indicates highly relevant objects on the map. Namely, the tourist object are marked and displayed as icons in the location of the correspondent object (Figure 2). The size indicates how closely the object meets the user’s interests. Each image of the icon indicates the type of tourist objects such as a museum, church, or restaurant and has own information window. Generally, the window consist information about the name of the attraction supported by a picture. In case the attraction is accessible to rain, an umbrella icon is introduced. The information window also displays a general idea of the time consumption of the attraction as well as other users and friends who have visited the attraction. The additional plus is the option to view travellers’ narratives in either video, audio, or text format. Through this window, the user has a possibility to rate the tourist object and hence to give or not, positive recommendation to other users. This operation is presumed and highly recommended to be done after visiting the tourist object and according to the tourist’s personal experience and satisfaction. So, this operation is encountered to assists in updating the user’s profile, and simultaneously to make the process of recommendation more accurate.

6. ACCURACY TESTING

In order to measure recommendation accuracy more precisely the authors applied information-retrieval classification metrics. This is done by evaluating the capacity of the recommender system by suggesting a list of appropriate objects to the user. With such metrics it is possible to measure the probability that the recommender system takes a correct or incorrect decision about the user interest for an item. When using classification metrics, one may distinguish four different kinds of recommendations (Table 1).

<table>
<thead>
<tr>
<th>Recommended</th>
<th>Non recommended</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interesting</td>
<td>True-positive</td>
</tr>
<tr>
<td>Uninteresting</td>
<td>False-positive</td>
</tr>
</tbody>
</table>

If the system suggests an interesting tourist object to the user, we have a true positive (second row-second column in the Table 1). In a contrary, the object is uninteresting and we have a false positive (third row-second column in the Table 1). If the system does not suggest an interesting tourist object we have a false negative (second row-third column in the Table 1). In opposite, the tourist object is classified as uninteresting and the system does not recommend it to the user, so it is a true negative (third row-third column in the Table 1).

Further step in accuracy testing of the suggested recommendation system is application of one of the most popular classification accuracy metrics, meaning the recall and the precision. Generally, these metrics can be calculated by counting the number of test object that fall into each cell in the table. In that line, the recall measures the percentage of interesting objects suggested to the users, with respect to the total number of interesting objects. In the other hand, the precision measures the percentage of interesting objects suggested to the users, with respect to the total number of suggested objects.
In order to understand the overall quality of the recommended tourism system, the authors proceeded by combining the recall and precision by means of the F-measure (Eq. 6).

\[
F\text{-measure} = \frac{2 \cdot \text{recall} \cdot \text{precision}}{\text{recall} + \text{precision}}
\] (6)

Thus, to evaluate the suggested system a methodology which uses the k-fold and the leave-one-out together with classification metrics recall and precision was used. According to the k-fold, users in the dataset are partitioned into k parts, whereas the k-1 parts are used to construct the model, while the remaining part represents the testing set. The model created with the k-1 partition is tested on the remaining partition by an algorithm (Table 2).

**Table no. 2 Algorithm**

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td>One user in the testing set is selected (the active user)</td>
</tr>
<tr>
<td>Step 2</td>
<td>One rated tourist object (the test object) is removed from the profile of the active user</td>
</tr>
<tr>
<td>Step 3</td>
<td>An order list of recommended tourist objects is generated</td>
</tr>
<tr>
<td>Step 4</td>
<td>If the test item is in the top-3 categories of recommended objects, either the true positive or false positive counter is incremented, depending whether the user liked or disliked the test item</td>
</tr>
</tbody>
</table>

In the line of accuracy testing, two distinct user groups were considered. So, in one hand, the group A contained all users who have rated 30-60 tourists” objects being referred as the few raters” user group. In the other hand, the group B contained all users who have rated 61-100 tourists” objects being referred as the moderate rater’s user group. Consequently, the Step 1 of the proposed algorithm (Table 2) was repeated for all the users in both groups. The Steps 2-4 were repeated for all tourists” objects rated by the active user.

In order to understand if a user likes or dislikes a rated tourist object, a hypothesis was introduced that the object is interesting to the user if satisfies two conditions. Firstly, a rate is given by the user for particular tourist object, and secondly, a mean average of ratings for particular user is introduced. The first constraint reflects the absolute meaning of the rating scale, while the second the user bias. If a rating does not satisfy the defined conditions, it is assumed that the item is not interesting to the user.

So, the conducted accuracy testing of the recommendation system performances are based on system precision, recall and the f-measure. The outcomes were averaged for each of the groups and presented in the Table 3.

**Table no. 3 Average values for recommendation system precision, recall and f-measure**

<table>
<thead>
<tr>
<th>Group</th>
<th>Precision (%)</th>
<th>Recall (%)</th>
<th>F-measure (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A</td>
<td>77.24</td>
<td>81.34</td>
<td>79.24</td>
</tr>
<tr>
<td>Group B</td>
<td>82.65</td>
<td>87.48</td>
<td>85.00</td>
</tr>
</tbody>
</table>
According to the obtained results, the developed national tourism web portal with its collaborative recommender system seems to be robust as it achieves good results in both scenarios (users with few and moderate ratings). It also accomplishes a good trade-off between precision and recall, a basic requirement for all recommendation systems. Experimental results show that the proposed approach can provide satisfactory performance even in a sparse dataset.

7. CONCLUSION AND FUTURE WORK

The paper proposed developed national web based tourism portal reach on accurate recommendations and guidelines for tourists and travellers in the line of identifying ideal trip and holiday. Due to the fact that tourism is defined as one of the most economically-oriented industries in the world it is noticeable that it contributes in large manner in enhancement and strengthening of national economies.

So, introduction of such software module is fully justified since it increases the awareness of tourist destination that is capable of fulfilling travellers’ preferences, and respectfully in raising net tourism income. Although still being in its initial phase of development, “MyTravelPal” performed satisfactory outcomes. Therefore, it is highly endorsed to apply this module which is based on the method of collaborative filtering. In case of acting as a national tourism web portal it may be introduced as a tool for assisting tourists in identification of a tailor-made holiday place. Moreover, it is in the line of supporting the national economy through improvement of the tourism supply in more qualitative manner. Having in mind that this portal indicates the motives, preferences and reasons for travelling to the FYROM, it may be of high importance to the key-tourism actors in the process of identifying measures and implementing activities necessary for creating comprehensive tourism policy.

Yet, the discussed results and findings should be interpreted as selected samples to underline the usefulness of the proposed approach in contribution to the e-tourism. So, the future work includes additional insights on the improvement in the presented web based national tourism portal.

Acknowledgements

The authors thank IvicaKocev and NatasaKoceskafor their useful support in data gathering, analysis, as well as software development.

References


