www.ijiarec.com

ISSN:2348-2079

Volume-5 Issue-2

International Journal of Intellectual Advancements and Research in Engineering Computations

A new e-guide for recommending travel package

[1]Mrs C.Navamani M.C.A., M.Phil., M.E., Assistant Professor,
[2]Mr C.Meenakshisundram Final MCA,
Department of MCA, Nandha Engineering College (Autonomous), Erode-52.
E-Mail ID: Navamanimca@gmail.com, m.s.satish008@gmail.com

Abstract* - As the worlds of commerce, entertainment, travel, and Internet technology become more inextricably linked, new types of business data become available for creative use and formal analysis. Indeed, this paper provides a study of exploiting online travel information for personalized travel package recommendation. A critical challenge along this line is to address the unique characteristics of travel data, which distinguish travel packages from traditional items for recommendation. To this end, we first analyze the characteristics of the travel packages and develop a Tourist-Area-Season Topic (TAST) model, which can extract the topics conditioned on both the tourists and the intrinsic features (i.e. locations, travel seasons) of the landscapes. Based on this TAST model, we propose a cocktail approach on personalized travel package recommendation. Finally, we evaluate the TAST model and the cocktail approach on real-world travel package data. The experimental results show that the TAST model can effectively capture the unique characteristics of the travel data and the cocktail approach is thus much more effective than traditional recommendation methods for travel package recommendation.

1. INTRODUCTION

As an emerging trend, more and more travel companies provide online services. However, the rapid growth of online travel information imposes an increasing challenge for tourists who have to choose from a large number of available travel packages for satisfying their personalized needs. Moreover, to increase the profit, the travel companies have to understand the preferences from different tourists and serve more attractive packages. Therefore, the demand for intelligent travel services is expected to increase dramatically.

2. CONCEPTS AND DATA DESCRIPTION

In this section, we first introduce the basic concepts, and then describe the recommendation scenario of this study. Finally, we provide the detailed information about the unique characteristics of travel package data.

Definition 1. A travel package is a general service package provided by a travel company for the individual or a group of tourists based on their travel preferences. A package usually consists of the landscapes and some related information, such as the price, the travel period, and the transportation means.

Specifically, the travel topics are the themes designed for this package, and the landscapes are the travel places of interest and attractions, which usually locate in nearby areas. Following Definition 1, an example document for a package named “Niagara Falls Discovery” from the STA. It includes the travel topics (tour style), travel days, price, travel area (the northeastern US), and landscapes (e.g., Niagara Falls), and so on. Note that different packages may include the same landscapes and each landscape can be used for multiple packages. Mean-while, for some reasons, the tourists for each individual package are often divided into different travel groups (i.e., traveling together). In addition, each package has a travel schedule and most of the packages will be traveled only in a given time (season) of the year, i.e., they have strong seasonal patterns. For example, the “Maple Leaf Adventures” is usually meaningful in Fall.

As illustrated in our preliminary work, there are some unique characteristics of the travel data. First, it is very sparse, and each tourist has only a few travel records. The extreme sparseness of the data leads to difficulties for using traditional recommendation techniques, such as collaborative filtering. For example, it is hard to find the credible nearest neighbors for the tourists because there are very few cotraveling packages.

Second, the travel data has strong time dependence. The travel packages often have a life cycle along with the change to the business demand, i.e., they only last for a certain period. In contrast, most of the landscapes will still be active after the original package has been discarded. These landscapes can be used to form new packages together with some other landscapes. Thus, we can observe that the landscapes are more sustainable and important than the package itself.

Third, landscape has some intrinsic features like the geographic location and the right travel seasons. Only the landscapes with similar spatial-temporal features are suitable for the same packages, i.e., the landscapes in one package have spatial-temporal autocorrelations and follow the first law of geography-everything is related to every-thing else, but the nearby things are more related than distant things [10]. Therefore, when making recommenda-tions, we should take the landscapes’ spatial-temporal correlations into consideration so as to describe the tourists and the packages precisely.

Fourth, the tourists will consider both time and financial costs before they accept a package. This is quite different from the traditional recommendations where the cost of an item is usually not a concern. Thus, it is very important to profile the tourists based on their interests as well as the time and the money they can afford. Since the package with a higher price often tends to have more time and vice versa, in this paper we only take the price factor into consideration.

Fifth, people often travel with their friends, family, or colleagues. Even when two tourists in the same travel group are totally strangers, there must be some reasons for the travel company to put them together. For instance, they may be of the same age or have the same travel schedule. Hence, it is also very important to understand the relation-ships among the tourists in the same travel group. This understanding can help to form the travel group.Last but not least, few tourist ratings are available for travel packages. However, we can see that every choice of a travel package indicates the strong interest of the tourist in the content provided in the package.

In summary, these characteristics bring in three major challenges. First, how to compare the interests of tourists and the content of the travel package; second, how to make package recommendations for each tourist; third, how to capture the tourist relationships to form a travel group. As a result, it is necessary to develop more suitable approaches for travel package recommendation.

3. THE TAST MODEL

In this section, we show how to represent the packages and tourists by a topic model, like the methods in [5] and based on Bayesian networks, so that the similarity between packages and tourists can be measured. Table 1 lists some mathematical notations in this paper.

3.1 Topic Model Representation

When designing a travel package, we assume that the people in travel companies often consider the following issues. First, it is necessary to determine the set of target tourists, the travel seasons, and the travel places. Second, one or multiple travel topics (e.g., “The Sunshine Trip”) will be chosen based on the category of target tourists and the scheduled travel seasons. Each package and landscape can be viewed as a mixture of a number of travel topics. Then, the landscapes will be determined according to the travel topics and the geographic locations. Finally, some additional information (e.g., price, transportation, and accom-modations) should be included. According to these processes, we formalize package generation as a W hat-Who-When-Where (4W) problem. Here, we omit the additional information and each W stands for the travel topics, the target tourists, the seasons, and the correspond-ing landscape located areas, respectively. These four factors are strongly correlated.

3.2 Model Inference

While the inference on models in the LDA family cannot be solved with closed-form solutions, a variety of algorithms have been developed to estimate the parameters of these models. In this paper, we exploit the Gibbs sampling method, a form of Markov chain Monte Carlo, which is easy to implement and provides a relatively efficient way for extracting a set of topics from a large set of travel logs. During the Gibbs sampling, the generation of each land-scape token for a given travel log depends on the topic distribution of the corresponding tourist-season pair and the landscape distribution of the area-topic pair.

4. COCKTAIL RECOMMENDATION APPROACH

In this section, we propose a cocktail approach on persona-lized travel package recommendation based on the TAST model, which follows a hybrid recommendation strategy [6] and has the ability to combine many possible constraints that exist in the real-world scenarios. Specifically, we first use the output topic distributions of TAST to find the seasonal nearest neighbors for each tourist, and collaborative filtering will be used for ranking the candidate packages.

4.1 Seasonal Collaborative Filtering for Tourists

In this section, we describe the method for generating the personalized candidate package set for each tourist by the collaborating filtering method. After we have obtained the topic distribution of each tourist and package by the TAST model, we can compute the similarity between each tourist by their topic distribution similarities. Intuitively, based on the idea of collaborative filtering, for a given user, we recommend the items that are preferred by the users who have similar tastes with her.

4.2 New Package Problem

In recommender systems, there is a cold-start problem, i.e., it is difficult to recommend new items. As we have explored in Section 2, travel packages often have a life cycle and new packages are usually created. Meanwhile, most of the landscapes will keep in use, which means nearly all the new packages are totally or partially composed by the existing landscapes. Let us take the year of 2010 as an example. There are 65 new packages in the data and only 2 of them are composed completely by new landscapes. Thus, for most of the new packages P^{new}, their topic distributions can be estimated by the topics of their landscapes.

4.3 Collaborative Pricing

In this section, we present the method to consider the price constraint for developing a more personalized package recommender system. The price of travel packages may vary from $20 to more than 3,000, so the price factor influences the decision of tourists. Along this line, we propose a collaborative pricing method in which we first divide the prices into different segments. Then, we propose to use the Markov forecasting model to predict the next possible price range for a given tourist.

In the first phase, we divide the prices of the packages based on the variance of prices in the travel logs, and the method is similar to the one used. We first sort the prices of the travel logs, and then partition the sorted list P L into several sublists in a binary-recursive way. In each iteration, we first compute the variance of all prices in the list.

4.4 Related Cocktail Recommendations

The previous cocktail recommendation approach (Cocktail) is mainly based on the TAST model and the collaborative filtering method. Indeed, another possible cocktail ap-proach is the content-based cocktail, and in the following, we call this method TASTContent. The main difference between TASTContent and Cocktail is that in TASTContent the content similarity between packages and tourists are used for ranking packages instead of using collaborative filtering. Since TAST Content can only capture the existing travel interests of the tourists, thus it may also suffer from the overspecialization problem. As there are many related topic models for the TAST model, it is also possible to design the similar cocktail recommendation approaches based on these models. Actually, it is quite straightforward to replace the TAST model by TT, TAT, and TST models in the cocktail approach to get the new recommendations. For example, in the experimental section, the notation TTER stands for the cocktail approach that is based on the TT model.

5. THE TRAST MODEL

In this section, we extend the current TAST model and propose a novel tourist-relation-area-season topic model to formulate the tourist relationships in a travel group. In the TAST model, we do not consider the information of the travel group. However, as noted in Section 2, each package has usually been used by many groups of tourists, and the tourists belong to different travel groups. Thus, if two tourists have taken the same package but in different travel groups, we can only say these two tourists have the same travel interest, but we cannot conclude that they share the same travel profile. However, if these two tourists are in the same group, they may share some common travel traits, such as similar cultural interests and holiday patterns. In the future, they may also want to travel together. Also, they may be family and always travel together during the holiday season. In this paper, we use the notation relation-ship to measure these commonalities and connections in tourists' travel profiles. Please also note that there are multiple tourist relationships simultaneously.

6. EXPERIMENTAL RESULTS

6.1 The Experimental Setup

The data set was divided into a training set and a test set. Specifically, the last expense record of each tourist in the year of 2010 was chosen to be a part of the test set, and the remaining records were used for training. The detailed information is described in Table 3.[4] Note that there are 65 new packages traveled by 269 tourists in the test set. However, only two of these packages are composed completely by new landscapes, and there are 11 new landscapes.

6.2 Season Splitting and Price Segmentation

In this section, we present the results of season splitting and price segmentation. For better we only show the travel logs with prices lower than $1,500. In the figure, different price segments are represented with different grayscale settings, and seasons are split by the dashed lines among months. In total, we have four seasons (i.e., spring, summer, fall, and winter), and five price segments (i.e., very low, low, medium, high, and very high). Since almost all the tourists in the data are from South China, this season splitting has well captured the climatic features there. Another interest-ing observation is that the peak times for travel in China include February (around the Spring Festival), July and August (the summer for students) and the beginning of October (National Day holiday). Fig. 7b describes the relationship between the percentage of the travel packages and the number of scheduled travel seasons. In Fig. 7b, we can see that most of the packages are only traveled in one season during a year, and less than 6 percent packages are scheduled in the entire year. At last, note that we do not give the illustration of relationship between each travel package and the number of its located areas. The reason is that almost all the packages in the data located in only one of the seven travel areas. These statistical results reflect the fact that landscapes in most packages have spatial-temporal autocorrelations, and the travel area and travel season segmentation methods are reasonable and effective.

6.3 Understanding of Topics

To understand the latent topics extracted by TAST, we focus on studying the relationships between topics and their landscapes'/packages' intrinsic characteristics. We have demonstrated that TAST can capture the spatial-temporal correlations among landscapes, and these landscapes, which are close to each other or with similar travel seasons, can be discovered. Meanwhile, the TAST model retains the good quality of the traditional topic models for capturing the relationships between landscapes locating in different areas and has no special travel season preference. Similarly, the topic distributions on each package can be also computed, and Table 4 illustrates many packages with highest probabilities from eight topics identified by the TAST model (Z ¼ 50) with the price factor being considered (as illustrated in Section 4.4). Based on the price-spatial-temporal correlations of packages (for many interpretations, there may contain some noise), all the topics can now be classified into eight types, which are noted from 1-1-1 (packages have price, spatial and temporal correla-tions) to 0-0-0 (packages have none of these correlations). Another interesting observation is that, the top travel packages in many topics are actually quite similar with each other, even though they are with different package IDs. For example, all the packages in topic 43 are about the Kunming-Dali-Lijiang tour. This finding once again demon-strates that, in addition to capture the intrinsic character-istics of the travel data, the TAST model still holds the capability of traditional models, such as the property of clustering documents (packages) [5].

6.4 Recommendation Performances

Since there are no explicit ratings for validation, we use the ranking accuracy instead. We adopt the widely used degree of agreement (DOA) and Top-K as the evaluation metrics. Also, a simple user study was conducted and volunteers were invited to rate the recommendations. For comparison, we recorded the best performance of each algorithm by tuning their parameters, and we also set some general rules for fair comparison. For instance, for collaborative filtering-based methods, we usually consider the contribution of the nearest neighbors with similarity values larger than 0. DOA measures the percentage of item pairs ranked in the correct order with respect to all pairs. Let NW_{Ui} ¼ P ðF_{Ui} E_{Ui} Þ denote the set of packages that do not occur in the training set (F_{Ui}) nor the test set (E_{Ui}) for U_i, and P R_{Pj} denote the predicted rank of package P_j in the recommendation list, and define check order$_{Ui}$ ðP_j; P_kÞ as 1 if P R_{Pj} P R_{Pk} otherwise 0.

For instance, an ideal(a random) ranking corresponds to a 100 percent (an average 50 percent) DOA, and we use DOA to stand for the average of each individual DOA. Under this metric, the ranking performance of each method is shown in Table 5, where we can see that Cocktail outperforms the benchmark methods. By integrating the price factor into the TAST model, Cocktail- performs nearly as well as Cocktail, and both of them perform better than TTER. Also, the methods that consider landscape informa-tion (i.e., LUCF, LBSVD, LItemRank, TTER, TASTContent, Cocktail) usually outperform those do not use such information (i.e., UCF, BSVD). As mentioned previously, it is harder to find the credible nearest neighbor tourists (and latent interests) only based on the cotraveling packages. Furthermore, TASTContent performs better than SContent, and TTER performs better than LUCF and LBSVD, and these demonstrate the effectiveness of model-ing latent topics. Meanwhile, unlike watching movies, most of the tourists seldom travel the packages that are similar to the ones that they have already traveled (e.g., have too many identical landscapes), thus content-based methods (i.e., SContent and TASTContent) perform worse than collaborative filterings (e.g., LUCF and Cocktail).

Computational performances. Also, we compare the com-putational performances of the algorithms. We run all the algorithms on the same platform. The execution time (i.e., the time used for building the model and making final recommendations for all the test tourists). We can see that many algorithms (e.g., LItemRank, TASTContent, Cocktail- and Cocktail) have the similar runtime. Among all the algorithms, BSVD and LBSVD are the most efficient, and Cocktail- has the worst computa-tional performance. Specifically, for the topic model-based methods, TTER does not have to consider the seasonal topic similarities of the tourists, thus it is the most efficient.

6.5 The Evaluation of the TRAST Model

Since we have little information about tourists, it is hard to interpret the identified relationships. However, we can test the effectiveness of the TRAST model from an alternative perspective; that is, the mined relationships will be used as features to help automatically form travel groups. We conduct two types of experiments. The first experiment is to use K-means clustering for grouping given tourists, and the second one is to find the tourists who would like to travel with given tourist.

6.6 Recommendation for Travel Groups

The evaluations in previous sections are mainly focused on the individual (personalized) recommendations. Since there are tourists who frequently travel together, it is interesting to know whether the latent variables (e.g., the topics of each individual tourist and the relationships of a travel group) as well as the cocktail approaches are useful for making recommendations to a group of tourists. To this end, we performed an experimental study on group recommendations.

7. CONCLUDING REMARKS

In this paper, we present study on personalized travel package recommendation. Specifically, we first analyzed the unique characteristics of travel packages and developed the TAST model, a Bayesian network for travel package and tourist representation. The TAST model can discover the interests of the tourists and extract the spatial-temporal correlations among landscapes. Then, we exploited the TAST model for developing a cocktail approach on personalized travel package recommendation. This cocktail approach follows a hybrid recommendation strategy and has the ability to combine several constraints existing in the real-world scenario. Furthermore, we extended the TAST model to the TRAST model, which can capture the relationships among tourists in each travel group. Finally, an empirical study was conducted on real-world travel data. Experimental results demonstrate that the TAST model can capture the unique characteristics of the travel packages, the cocktail approach can lead to better perfor-mances of travel package recommendation, and the TRAST model can be used as an effective assessment for travel group automatic formation. We hope these encouraging results could lead to many future work.

REFERENCES

[1] G.D. Abowd et al., “Cyber-Guide: A Mobile Context-Aware Tour Guide,” Wireless Networks, vol. 3, no. 5, pp. 421-433, 1997.

[2] G. Adomavicius and A. Tuzhilin, “Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions,” IEEE Trans. Knowledge and Data Eng., vol. 17, no. 6, pp. 734-749, June 2005.

[3] D. Agarwal and B. Chen, “fLDA: Matrix Factorization through Latent Dirichlet Allocation,” Proc. Third ACM Int’l Conf. Web Search and Data Mining (WSDM ’10), pp. 91-100, 2010.

[4] O. Averjanova, F. Ricci, and Q.N. Nguyen, “Map-Based Interac-tion with a Conversational Mobile Recommender System,” Proc. Second Int’l Conf. Mobile Ubiquitous Computing, Systems, Services and Technologies..

[5] D.M. Blei, Y.N. Andrew, and I.J. Michael, “Latent Dirichlet Allocation,” J. Machine Learning Research, vol. 3, pp. 993-1022, 2003.

[6] R. Burke, “Hybrid Web Recommender Systems,” The Adaptive Web, vol. 4321, pp. 377-408, 2007.

[7] B.D. Carolis, N. Novielli, V.L. Plantamura, and E. Gentile, “Generating Comparative Descriptions of Places of Interest in the Tourism Domain,” Proc. Third ACM Conf. Recommender Systems (RecSys ’09), pp. 277-280, 2009.

[8] F. Cena et al., “Integrating Heterogeneous Adaptation Techniques to Build a Flexible and Usable Mobile Tourist Guide,” AI Comm., vol. 19, no. 4, pp. 369-384, 2006.

[9] W. Chen, J.C. Chu, J. Luan, H. Bai, Y. Wang, and E.Y. Chang, “Collaborative Filtering for Orkut Communities: Discovery of User Latent Behavior,” Proc. ACM 18th Int’l Conf. World Wide Web (WWW ’09), pp. 681-690, 2009.