An Online Vacation Package Recommendation System

*Md.Asma Kousar   ****A.Madhavi**

*M.TECH student ,Dept of CSE, Vaagdevi College of Engineering

**Assistant Professor, Dept of CSE, Vaagdevi Engineering College

Abstract –

Tourism is one of the main reasons for the development and maintenance of any provinces. Its reliability does help in the growth of economic value of any country. And in these recent years, recommendation system is becoming so popular among online applications such as ecommerce and so on. This paper tries to overview how recommendation system is useful to tourism. A large travel and tourist dataset which includes travel packages, will be exploited and after analyzing it, records will be segregated into groups as per the tourist, area, season and most importantly in the basis of ratings. The personalized list of tourist package recommendation will be generated and recommended on the basis of collaborative pricing. Furthermore, the latent relationship among the entire tourist will be analyzed so that package can be recommended for similar tourists in a group. This approach can be used both by the travel agencies and the travel groups at low maintenance and cost.

Keywords – tourism; economic; dataset; tourist; list; recommendation.

I. INTRODUCTION

Recommendation system has successfully enhanced the quality of services for both consumer and service provider. So the travel agencies after considering the preferences of their clients (tourists), list of travel packages are offered to the tourist. On the other hand, instead of digging through all constraints separately, tourists can choose its favorable

travel package from the recommended personalized package list.

Recommendation system is categorize into

- Content based recommendation system: it is meant for only an individual user. The user is recommended with items similar to those that the user has preferred in the past.

- Collaborative filtering recommendation system: unlike content – based, it is meant for a number of users. The user is recommended with the list of items that other users with similar characteristics and preferences liked in the past.

- Hybrid recommendation system

Many technical and domain challenges are faced during the design and implementation of an effective recommendation system for personalized travel package recommendation.

First, travel data are always much fewer and sparser than other items, like movies for recommendation, because the costs for a

travel are much more expensive than for watching a movie.

Second, usually travel package are location based so they are said to be spatial or temporal for example the package contains locations which are geographically near. Also, different travel packages are usually developed for different travel seasons. Therefore, the landscapes in a travel package usually have spatial temporal autocorrelations.

Third, traditional recommender systems usually rely on user explicit ratings. However, for travel data, the user ratings are usually not conveniently available.

Finally, the old recommendation system usually has a long period of stable value, while the values of travel packages can easily be reduced over time. So the travel companies need to actively create new tour packages to replace the old ones based on the interests of the tourists.

To overcome these challenges, a different approach is implemented to personalized travel package recommendation. Starting with analyzing the key characteristics of an existing travel data set, travel time and travel destinations are divided into different

seasons and areas and a tourist-area-season theme (TAST) model is developed, which can represent travel packages and tourist's interests by different topic distributions. In this model, the extraction of themes is conditioned on both the tourists and intrinsic features such as locations and travel seasons of the landscapes. So, this model can well represent the content of the travel packages and the interests of the tourists.

This model is further enhanced by considering some additional factors including the seasonal behaviors of tourists, the prices of travel packages, and the cold start problem of new packages leading to the development of cocktail recommendation approach. Also tourist-relation-area-season theme (TRAST) model is introduced, which helps in understanding the reasons why tourists form a travel group. This goes beyond personalized package recommendations and is helpful for capturing the latent relationships among the tourists in each travel group. Finally, I will showcase the model on a travel data set to show that this model can effectively capture the unique characteristics of travel and performs much better than traditional techniques. A travel package is a general service package provided by a travel company for an individual or a group of tourists based on their travel preferences. A package consists of a combination of themes (area-season) and landscapes, which are the places of interest and attractions located in nearby areas the landscapes and some related information, such as the price, the travel period, and the transportation means. For this system, the users are the tourists and items are the existing packages, and we exploit a travel data set for building a recommender system.

II. BACKGROUND AND APPROACH

Content-based filtering analyzes the association between user's preferences and the descriptions of items. To recommend new items to a user, the content-based filtering approach matches the new items descriptions to those items that have been preferred by the user. On the other hand, the collaborative filtering (CF) approach does not need content information to make recommendations. Collaborative Filtering has been developed and improved over the past decade to the point where a wide variety of algorithms exist for generating

recommendations. Each algorithmic approach claimed itself to be superior for some purpose and they have solid reasons

behind it. Clearly identifying the best algorithm for a given purpose has proven challenging, in part because researchers disagree on which attributes should be measured, and on which metrics should be used for each attribute. Researchers who survey the literature will find over a dozen quantitative metrics and additional qualitative evaluation techniques.

III. TAST MODEL

Travel companies often consider the following issues

- Determine the set of target tourist, target season and the target places
- Travel theme will be determined according to the target tourist and travel season
- At last some additional information like price, transportation and accommodation should be included

As per the requirement, package generation is designed as a What- Who-When-Where (4W) problem. Here, each W stands for the travel topics, the target tourists, the seasons,

and the corresponding landscape located areas, respectively. These four factors are strongly correlated. The generation of a package can always be reprocessed. The landscapes for the package are drawn from the landscape set one by one. Steps to choose a landscape: (i) choose a topic (theme) from the distribution over topics (themes) specific to the given tourist and season (ii) landscape is generated from the chosen topic and travel area. Mathematically, the generative process corresponds to the hierarchical Bayesian model for TAST. TAST model is used to represent the packages and tourists by a topic model, so that the similarity between packages and tourists can be measured. In TAST model, each travel log can be distinguished by a vector of three attributes, i.e. package ID of one travel log, tourist/user ID, timestamp; the timestamp can be further projected to a season. Package ID of one travel log is again distinguished by three factors, i.e. vector of landscapes, located

area(s) and price factor. Then specific tourist – season pair and area – topic pair is derived from each topic distribution and landscape distribution.

A. Model Inference

For extracting a set of topics from a large set of travel logs, the Gibbs sampling method, a form of Markov chain Monte Carlo is preferred. And the generation of each landscape 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. For an example, let's take the topic (theme) "Central Park". In each iteration, the topic assignment of one "Central" token depends on both the topics of the landscapes traveled by the tourist in the given season and the topics of the other landscapes located nearby. Then all the tourists and packages are represented by the Z entry topic distribution vectors (Z, the number of topics, is usually in the range of [20, 100]). For example, a tourist, who traveled "Tour in Disneyland, Hongkong" and "Christmas day in Hongkong", may have high probabilities on the entries that stand for the topics such

as "amusement parks" and "Hongkong". By computing the similarity of the topic distribution vectors, similarity between the corresponding tourists and packages are defined.

B. Area/Seasons Segmentation

The entire location space in our data set is divided into seven big areas according to the travel area segmentations provided by the travel company, which are South China (SC), Center China (CC), North China (NC), East Asia (EA), Southeast Asia (SA), Oceania (OC), and North America (NA), respectively. Most packages are seasonal, and an information gain-based method is used to get the season splits. The entire year is partitioned into several seasons. In each iteration, the weighted average entropy (WAE) is used to find the best split.

IV. COCKTAIL APPROACH

This approach is used to derive personalized travel package recommendation based on the TAST model, which follows a hybrid recommendation strategy. Their step includes:

- To find the seasonal nearest neighbors for each tourist, the output topic distributions of TAST is used,
- Collaborative filtering is used to sort the candidate packages.
- New packages are added into the candidate list by computing

  similarity with the candidate packages generated previously
- Collaborative pricing is utilized to predict the possible price distribution of each tourist and reorder the packages. After removing the packages which are no longer active, we will have the final recommendation list.

A. Seasonal Collaborative Filtering for Tourists

Here items of the given user are recommended based on the preferred items of the other users who have similar taste with her. This concept is based on the idea of collaborative filtering of recommending system. The seasonal topic distribution for each tourist is derived from the TAST model. Methods like matrix factorization,

graphical distances or correlation coefficient can be used to compute these similarities. For computing the similarity between two tourists in a specific season, correlation coefficient is preferred due to its simple and effective problem solving scenario. Then the nearest neighbors of the given tourist are measured by ranking their similarity values. Thus, the packages which are favored by

these neighbors can be selected as candidate packages which provide a rough list for recommendation. In case, if a new package is introduced it is not easy for the system to recommend the new package to any tourist as it doesn't have any ratings. This complication is known as cold-start problem, i.e. it is difficult to recommend new items. So in order to recommend these new packages, the similarity between the new package and the given number (e.g., 12) of candidate packages in the top of the recommendation list is computed. The new packages which are similar to the candidate packages are added into the recommendation list and their ranks in the list based on the average probabilities of the similar candidate packages. This method

helps in solving the issue of cold – start problem and overspecialization problem which occurs if the recommendation system deals just with the current interest of the given tourist and not the rest.

B. Collaborative Pricing

For developing a more précised personalized package recommendation system, prices of travel packages which is the main influence of many tourists, are taken into

consideration and this concept is referred to as collaborative pricing method. Then, the Markov forecasting model is proposed to predict the next possible price range for a given tourist. Firstly, the prices of the packages are sorted and divided it into several sub lists in a binary – recursive way based on the variance of prices in the travel logs. Then the best split price having the minimal weighted average variance (WAV) is defined. Secondly, each price segment is marked as a price state and computed the transition probabilities between them. For this, in a scenario where a tourist used a package with price state A before traveling a package with price state B, then the weight of the edge from A to B will plus 1. These

weights from all the tourists are added and the packages are normalized into transition probabilities, which are later transformed a state transition matrix. From the current price state of a given tourist, the next possible price state is predicted by the one-step Markov forecasting model based on random walk. Finally, the obtained predicted probability distribution of the given tourist on each state used as weights to multiply the probabilities of the candidate packages in the rough recommendation list so as to reorder these packages.

V. TRAST MODEL

The tourist-relation-area-season topic (TRAST) model is another extension of TAST model to find the tourist relationships in a travel group with a new set of variables, with each entry indicating one relationship. Tourist travelling in same travel group shares some common travel traits, such as similar cultural interests and holiday patterns. But if tourist from different travel groups preferred same package then it can be predicted that these tourists are having same travel interest. In TRAST model, the

purchases of the tourists in each travel group are summed up as one single expense record and, thus, it has more complex generative process. If two selected tourists in a travel group who are young and dating with each other decide to travel in winter and the destination is North America. To generate a travel landscape (l), firstly a relationship is extracted (e.g. lover), and then find a topic for lovers to travel in the winter (e.g., skiing). Finally, based on this skiing topic and the selected travel area (e.g., Northeast America), landscape (e.g., Stowe, Vermont)

is drawn. To perform the inference, instead of Gibbs sampling formulae, to make it more efficient and easier, it is exercised in two distinct parts. Firstly, split TRAST model into two sub models. The first sub model TRAST1 is just like the TAST model, except for the two tourists are latent factors and some of the notations are with different meanings here. By this model, we use a sample to obtain topic assignments and tourist pair assignments for each landscape token. Then, in the second sub model TRAST2, we treat topics and tourist pairs as

known, and the goal is to obtain relationship assignments. After Gibbs sampling, each tourist's travel relationship preference can be estimated. Two tourists in a travel group are taken as a tourist pair for mining their relationships. By this TRAST model, all the tourists' travel preferences are represented by relationship distributions. For a set of tourists, who want to travel the same package, their relationship distributions are used as features to cluster them, so as to put them into different travel groups. For this, K-means clustering algorithm is used. Indeed, in real applications, when generating a travel group, some more external constraints, like tourists' travel date requirements, the travel company's travel group schedule are also considered.

VI. CONCLUSIONS

Large datasets are exploited to get deep details about tourists, packages and then derived new packages. Different topics or themes are analyzed and new ones generated. First of all, the TAST model is developed which captures unique

characteristics of travel data and thus discovers the interests of the tourists and extract the spatial temporal correlations among landscapes and outputs the topic and season recommendation. The TAST model is further expanded and developed cocktail approach for personalized recommendation for travel package. Along with new constraints, TAST model is extended to TRAST model which acquire the relations between tourists in each group. It is helpful for capturing the latent relationships among the tourists in each travel group and to understand why tourists form groups. TRAST model is used for effective analysis of automatic formation.

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] Lui, Chen,Xiong,Li and Wu, *"A Cocktail Approach for Travel Package Recommendation"* IEEE Transactions On Knowledge And Data Engineering, Vol. 26, No. 2, February 2014

[4] 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.

[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] 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. N.A.C. Cressie, Statistics for Spatial Data. Wiley and Sons, 1991.

[8] J. Delgado and R. Davidson, *"Knowledge Bases and User Profiling in Travel and Hospitality Recommender Systems,"* Proc. ENTER 2002 Conf. (ENTER '02), pp. 1-16, 2002.

[9] U.M. Fayyad and K.B. Irani, *"Multi-Interval Discretization of Continuous-Valued Attributes for Classification Learning,"* Proc. Int'l Joint Conf. Artificial Intelligence (IJCAI), pp. 1022-1027, 1993.

[10] F. Fouss et al., *"Random-Walk Computation of Similaritiesbetween Nodes of a Graph with Application to Collaborative Recommendation,"* IEEE Trans. Knowledge and Data Eng., vol. 19, no. 3, pp. 355-369, Mar. 2007.

[11] Y. Ge et al., *"Cost-Aware Travel Tour Recommendation,"* Proc. 17th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (SIGKDD '11), pp. 983-991, 2011.

[12] Y. Ge et al., *"An Energy-Efficient Mobile Recommender System,"* Proc. 16th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (SIGKDD '10), pp. 899-908, 2010. [13] M. Gori and A. Pucci, *"ItemRank: A Random-Walk Based Scoring Algorithm for Recommender Engines,"* Proc. 20th Int'l Joint Conf. Artificial Intelligence (IJCAI '07), pp. 2766-2771, 2007.

[14] U. Gretzel, *"Intelligent Systems in Tourism: A Social Science Perspective,"* Annals of Tourism Research, vol. 38, no. 3, pp. 757-779, 2011.

[15] T.L. Griffiths and M. Steyvers, *"Finding Scientific Topics,"* Proc. Nat'l Academy of Sciences USA, vol. 101, pp. 5228-5235, 2004.

[16] Q. Hao et al., *"Equip Tourists with Knowledge Mined from Travelogues,"* Proc. 19th Int'l Conf. World Wide Web (WWW '10), pp. 401-410, 2010.

AUTHOR 1:-

  * Md.Asma Kousar completed her B tech Vaagdevi Engineering College in 2014 and pursuing M-Tech in Vaagdevi College of Engineering

AUTHOR 2:-

   **A.Madhavi is working as Assistant Professor in Dept of CSE, Vaagdevi Engineering College