ISSN 23499842(Online), Volume 1, Special Issue 1(NCFCSIT 15), March 2015
International Journal of Innovative Trends and Emerging Technologies

KNOWLEDGE BASED USER PROFILING IN TRAVEL PACKAGE RECOMMENDATION

[1]S.BHUVANESWARI, [2]MRS.S.JAYANTHI, M.E.,
[1]PG Scholr, [2]Assistant Professor [1,2]Department of CSE, Agni College of Techmology, Chennai,
bhuvaneswarisa10 @gmail.com, jayanthi.cse@act.edu.in

***Abstract*- Rapid growth of online travel information imposes increasing challenges for tourists who have to choose from large number of travel package to satisfy their personalized needs. The demand for online travel services increase dramatically to improve living standards, even an ordinary family can do extended travel very comfortably on a small budget. In this paper, we first examine the unique characteristics of travel data for existing travel packages, and then develop a tourist-area-season topic (TAST) model. This TAST model helps to extract topic conditioned on both travel packages and tourists that can be represent in different topic distributions. Then, based on this model, we propose a hybrid recommendation to generate the lists for travel package recommendation. Besides, we extend the TAST model to the tourist-relation-area-season topic (TRAST) model that captures relationships among the tourists in each travel group. Thus results show that the TAST model can effectively capture the unique characteristics and the hybrid recommendation is more effective than traditional recommendation techniques.**

Keywords: collaborative filtering, gibbs sampling algorithm, TAST, Hybrid recommendation, TRAST.

I.INTRODUCTION

Recommender systems which are defined as applications that e-commerce sites make use to suggest products and provide consumers with information based on their decision-making processes. They implicitly assume user needs and constraints, through suitable recommendation algorithms, and convert them into the intelligent recommender by using product selections. Knowledge is extracted either by content- or knowledge-based approaches. Furthermore, the interaction process that turns needs into products, and view of user which is purely depends on the recommendation technology and algorithms. For example, if the system used for guiding the behavior of other users in the recommendation, it explicitly shows reviews or rates of the selected products from a similar user.

Besides, there are technical and domain issues to implement an effective travel package recommendation. Precisely, we first analyze the characteristics of the travel packages. Along this line, travel time and destinations are split into different seasons and areas. Then, also a Tourist-Area-Season Topic (TAST) model was developed, which can extract the topics conditioned on both the tourists and the landscape (i.e. locations, travel area). As a result of TAST model can represent the content of the travel packages as well the interests of the tourists. Based on this TAST model, a hybrid recommendation approach is developed for travel package recommendation by including some additional factors such as the behaviors of tourists based on seasonal, the prices of given travel packages, and some common issues is that cold start problem. The tourist-relation-area-season topic (TRAST) model helps to understand the reasons why tourists form a travel group. This is very helpful for capturing the latent relationships among the tourists in each travel group.

II. RELATED WORK

In general, the related work that can be categorized into two classes: the first class will be tourism domain, where some system has been design to help tourists which can be sub divided into other two groups.

In first group, people are particularly focused on intelligent systems for the tourists before planning, information filtering. For example, Yin et al. proposed an automatic trip planning framework. Also, Hao et al. proposed a location-topic model from large collection of travel logs and to recommend the travel destination. Wu et al. designed a system to generate the personalized tourism summary.

In the second subgroup, people target on providing more personalized travel information. For example, Averjanova et al. designed map-based mobile system. Moreover, Carolis et al. used a map for outlining the location and the information of landscapes in area. Finally, a more revealing on-tour support system, MobyRek, was designed by Ricci and Nguyen.

The second class includes the research work related to topic models and their applications on recommender systems. Topics models are usually based on the documents (such as messages, emails, etc.) are mixtures of latent topics.

III. SYSTEM DESIGN

System Architecture Design depicts to communicate with client server interfaces. In that design user will compare the different travel packages for same location. These travel packages retrieve information from package content database and also look for similar travel packages that are already available in travel log database. The travel log will maintain the history of travel packages. Admin will upload the topics for each packages based on user rating. We can find his/her nearest neighbors by ranking their similarity values. Thus, the packages, favored by these neighbors but have not been traveled by the given tourist, can be selected as candidate packages which form a rough recommendation list, and they are ranked. Suppose 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. Admin will provide recommend the items that are preferred by the users who have similar tastes with her. After removing the packages which are no longer active, we have the final recommendation list. These recommended lists will send to user. In real applications, new travel package recommendation list can be separated from the general list.

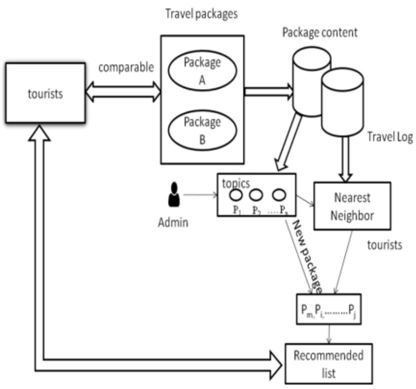

Figure 3.1: System Design

ISSN 23499842(Online), Volume 1, Special Issue 1(NCFCSIT 15), March 2015
International Journal of Innovative Trends and Emerging Technologies

3.1 Publishing the services

The travel packages will satisfy their information which is provided to users. So, develop a tourist-area-season topic (TAST) model, this model analyzed travel packages and tourists through different topic distributions. As results, the extraction of topics is conditioned based on the tourist's needs and either locations or travel seasons of the landscapes. And also shows the content of the travel packages and the interests of the tourists. Each and every choice of a travel package represents the strong interest of the tourist in the content provided in the package.
There are three major challenges to characteristics: First, how to estimate the tourist interests and the travel package; second, how to recommend the packages to each tourist; third, how to form a travel group based on tourists relationship. Finally, it is determined to develop more suitable approaches for travel package recommendation.

3.2 Machine Learning TAST Model

In general, a learning problem considers a set of n samples of data and tries to predict properties of unknown data. If each sample is more than a single number, and for instance a multi-dimensional entry, is it said to have several attributes, or features. Machine learning is about learning some properties of a data set and applying them to new data. This is why a common practice in machine learning to evaluate an technique is to split the data at hand in two sets, one that we call a training set on which we learn data properties, and one that we call a testing set, on which we test these properties.

3.3 Normalizing

First, we have obtained the topic distribution of each tourist and package by the TAST model; then we can compute the similarity between each tourist by their topic distribution similarities. We recommend the items that are preferred by the users who have similar tastes with her that as been given to the user. Thus, the packages which are favored by these neighbors but have not been traveled by the given tourist can be selected and forms a rough recommendation list; also they are ranked by the probabilities which are examined by the collaborative filtering.
Normally, services are normalized by quality of services (QoS). The QoS performed based on certain factors such as response time, availability, throughput and restability that should be satisfied the services.

IV. PROPOSED ALGORITHM

In order to access recommended list for each given tourist, there are a variety of algorithms have been advanced to estimate the parameters of these model. And also show how to estimate the packages and packages by topic based model which based on the Bayesian network. From this results helps to measure the similarity between packages and tourists.

4.1 Gibbs sampling algorithm

A variety of algorithms are developed to estimate the parameter the TAST model. Here, we achieved the Gibbs Sampling method, a form of Markov chain Monte Carlo (MCMC), which is easy to implement and efficient way of extracting large sets from travel logs. The generation of each landscape provides token for given travel log depends on topic distributions. Finally posterior estimates are calculated from given training set. The main aim this algorithm to construct a Markov chain that has the target posterior distribution as its stationary distribution.
Let $f_{(j)}(x_{(j)} \mid x_{(1)}, \ldots, x_{(m)})$ be the topic distribution of the element by number of landscape given all the other parameters minus the j^{th}, then **Gibbs Sampling** for an m-component variable of

$$X^t = (x^t_{(1)}, \ldots, x^t_{(m)})$$

tourist-season pair is given by the transition from to X^{t+1} generate as:

Given an arbitrary initial value: $X^0 = (x_1^0, x_2^0, \ldots, x_m^0)$.

$$1. X^{t+1}_{(1)} \sim f_{(1)}(X^{t+1}_{(1)} \mid x^t_{(2)}, \ldots, x^t_{(m)})$$

$$2. X^{t+1}_{(2)} \sim f_{(2)}(X^{t+1}_{(2)} \mid x^{t+1}_{(1)}, x^t_{(3)}, \ldots, x^t_{(m)})$$

$$\vdots$$

$$m. X^{t+1}_{(m)} \sim f_{(m)}(X^{t+1}_{(m)} \mid x^{t+1}_{(1)}, x^{t+1}_{(2)}, \ldots, x^{t+1}_{(m-1)})$$

4.2 Information gain method

Information gain (IG) measures the amount of information in bits about the prediction rates that should be available information (ie., the presence of a feature and the corresponding class distribution). Concretely, it measures the expected reduction in entropy (uncertainty associated with a random feature).
Information gain method will divide the entire location in our data set into multiple areas according to the travel area segmentations provided by the travel company. We use an information gain-based method to get the season splits. The information entropy of the season S^P is

$$\mathbf{Ent(S^{P}) = -\sum_{i=1}^{|S^P|} P_i \log(P_i)}$$

where $|S^p|$ is the number of different packages in S^p and P_i is the proportion of package Pi in this season.
The best split month induces a maximum information gain given by E(i) which is equal to

$$\mathbf{Ent(S^P) - WAE(I; S^P)}$$

4.3 Collaborative Filtering

Collaborative filtering (CF) is a technique used by travel package recommender systems. Collaborative filtering is the process of filtering for information or patterns using techniques involving collaboration among multiple travel agents, viewpoints etc. Applications of collaborative filtering typically involved for very large data sets.

Here we use to generating the personalized candidate package set for each tourist by using the collaborating filtering algorithm. The packages which are favored by these neighbors those does not been traveled by the given tourist and also form a rough recommendation list. In recommendation system, issues cold start problem arise so new packages that are similar to the ones already traveled by the given tourist. Thus, we propose to compute the similarity between the new package and the given number (e.g., 10) of candidate packages in the top of the recommendation list. The new packages that are similar to the packages which have been added into the recommendation list and their ranks in the list based on the average probabilities. It is expected that this method can not only deal with the cold-start problem but also avoid the overspecialization problem.

Suppose if two users have same rated items in common and also have similar tastes. Such user build a group, they get recommendation to those items that he/she hasn't rated before, but at same time they have positively rated by user in his/her neighborhood.

V. RESULTS AND DISCUSSIONS

From the results, we can see that the proposed hybrid recommendation that will works very well for predicting the tourists' travel preferences by deriving the unique characteristics of the travel package data. Also, we explained about domain depended such as travel.

Also, if we want to deploy this work for real-world services, we have to form more practical functions. Second, there are some restriction with the evaluation, which is based on the ability to recover removed test data and a simple user study. For real-world applications, more revealing online experiments are needed.

$$X^t = (x^t_{(1)}, \ldots, x^t_{(m)})$$

VI. CONCLUSION

In this paper explained about study on personalized travel package recommendation. Specifically, we first examined the unique characteristics of travel packages and developed the TAST model. The TAST model can find the interests of the tourists and extract information. Then, we are deriving the TAST model for developing a hybrid recommendation approach on personalized travel package recommendation. This hybrid recommendation strategy has the ability to combine several constraints existing in the real-world scenario. Furthermore, we enlarged the TAST model to the TRAST model, which can record the relationships among tourists in each travel group. Finally, an empirical study was deployed on real-world travel data. The hybrid recommendation can helps better performances of travel package recommendation, and the TRAST model which is more effective and also provides automatic prediction.

VII. FUTURE WORK

For future work in this paper so we could replace the collaborative filtering algorithm it’s a more complex algorithm to decrease the further cost. Work can be done on using our model for tourist packages as per the quality. To study deeply about the performance evaluation and what are the factors will affect can be done. Finally we can compare with other topic models for better performances.

REFERENCE

[1]. Qi Liu, Enhong Chen “A Cocktail Approach for Travel Package Recommendation” IEEE *IEEE Trans. Knowl. Data Eng*, Vol. 26, NO. 2, February 2014

[2]. Q. Liu, Y. Ge, Z. Li, H. Xiong, and E. Chen, “Personalized Travel Package Recommendation,” Proc. IEEE 11th Int’l Conf. Data Mining (ICDM ’11), pp. 407-416, 2011.

[3].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. Knowl. Data Eng.*, vol. 17, no. 6, pp. 734–749, Jun. 2005

[4]. H. J. Ahn, “A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem,” *Inf. Sci.*, vol. 178, no. 1, pp. 37–51, Jan. 2008.

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

[6]. A. Mccallum, X. Wang, and A. Corrada-Emmanuel, “Topic and Role Discovery in Social Networks with Experiments on Enron and Academic Email,” J. Artificial Intelligence Research, vol. 30,pp. 249-272, 2007.

[7]. Q. Liu, E. Chen, H. Xiong, C. Ding, and J. Chen, “Enhancing Collaborative Filtering by User Interests Expansion via Personalized Ranking,” IEEE Trans. Systems, Man, and Cybernetics, Part B: Cybernetics, vol. 42, no. 1, pp. 218-233, Feb. 2012.

[8]. R. Pan et al., “One-Class Collaborative Filtering,” Proc. IEEE Eighth Int’l Conf. Data Mining (ICDM ’08), pp. 502-511, 2008.

[9]. P. Lops, M. Gemmis, and G. Semeraro, “Content-Based Recommender Systems: State of the Art and Trends,” Recommender Systems Handbook, chapter 3, pp. 73-105, 2010.

[10]. 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.

[11]. J. Herlocker, J. Konstan, L. Terveen, and J. Riedl, “Evaluating Collaborative Filtering Recommender Systems,” ACM Trans. Information Systems, vol. 22, no. 1, pp. 5-53, 2004.

[12]. C. Wang and D. Blei, “Collaborative Topic Modeling for Recommending Scientific Articles,” Proc. ACM 17th ACM SIGKDD Int’l Conf. Knowledge Discovery and Data Mining, pp. 448-456, 2011.

[13]. X. Wu, J. Li, and S. Neo, “Personalized Multimedia Web Summarizer for Tourist,” Proc. 17th Int’l Conf. World Wide Web (WWW ’08), pp. 1025-1026, 2008.

[14] .J. Wu, H. Xiong, and J. Chen, “Adapting the Right Measures for KMeans Clustering,” Proc. 15th ACM SIGKDD Int’l Conf. Knowledge Discovery and Data Mining, pp. 877-886, 2009

[15]. 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.

[16]. 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.

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

[18] 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.

[19] J. Herlocker, J. Konstan, L. Terveen, and J. Riedl, "Evaluating Collaborative Filtering Recommender Systems," *ACM Trans. Information Systems,* vol. 22, no. 1, pp. 5-53, 2004.

[20] A. Jameson and B. Smyth, "Recommendation to Groups," *The Adaptive Web,* vol. 4321, pp. 596-627, 2007.

[21] Y. Koren, "Factorization Meets the Neighborhood: A Multifaceted Collaborative Filtering Model," *Proc. 14th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (SIGKDD '08),* pp. 426-434, 2008.

[22] P. Lops, M. Gemmis, and G. Semeraro, "Content-Based Recommender Systems: State of the Art and Trends," *Recommender Systems Handbook,* chapter 3, pp. 73-105, 2010.

[23] J. MacQueen, "Some Methods for Classification and Analysis of Multivariate Observations," *Proc. Berkeley Symp. Math. Statistics and Probability (BSMSP),* vol. 1, pp. 281-297, 1967.

[24] A. Mccallum, X. Wang, and A. Corrada-Emmanuel, "Topic and Role Discovery in Social Networks with Experiments on Enron and Academic Email," *J. Artificial Intelligence Research,* vol. 30, pp. 249-272, 2007.

[25] R. Pan et al., "One-Class Collaborative Filtering," *Proc. IEEE Eighth Int'l Conf. Data Mining (ICDM '08),* pp. 502-511, 2008.

[26] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, "GroupLens: An Open Architecture for Collaborative Filtering of Netnews," *Proc. ACM Conf. Computer Supported Cooperative Work (CSCW '94),* pp. 175-186, 1994.

[27] F. Ricci, D. Cavada, N. Mirzadeh, and N. Venturini, "Case-Based Travel Recommendations," *Destination Recommendation Systems: Behavioural Foundations and Applications,* chapter 6, pp. 67-93, 2006.

[28] F. Ricci et al., "DieToRecs: A Case-Based Travel Advisory System," *Destination Recommendation Systems: Behavioural Foundations and Applications,* chapter 14, pp. 227-239, 2006.

[29] F. Ricci and Q. Nguyen, "Mobyrek: A Conversational Recommender System for On-the-Move Travelers," *Destination*

Recommendation Systems: Behavioural Foundations and Applications, chapter 17, pp. 281-294, 2006.

[30] M. Rosen-Zvi et al., "The Author-Topic Model for Authors and Documents," *Proc. 20th Conf. Uncertainty in Artificial Intelligence (UAI '04),* pp. 487-494, 2004.

[31] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Application of Dimensionality Reduction in Recommender Systems-a Case Study," *Proc. ACM WebKDD Workshop*, pp. 82-90, 2000.

[32] S. Staab et al., "Intelligent Systems for Tourism," *IEEE Intelligent Systems,* vol. 17, no. 6, pp. 53-66, Nov. 2002.

[33] A. Venturini and F. Ricci, "Applying Trip@ Dvice Recommendation Technology to www.visiteurope.com," *Proc. 17th European Conf. Frontiers in Artificial Intelligence and Applications,* vol. 141, p. 607, 2006.

[34] C. Wang and D. Blei, "Collaborative Topic Modeling for Recommending Scientific Articles," *Proc. ACM 17th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining,* pp. 448-456, 2011.

[35] X. Wu, J. Li, and S. Neo, "Personalized Multimedia Web Summarizer for Tourist," *Proc. 17th Int'l Conf. World Wide Web (WWW '08),* pp. 1025-1026, 2008.