Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

International Conference on Mobility in Computing- ICMiC13, Organized by Mar Baselios College of Engineering and Technology during December 17-18, 2013 at Trivandrum, Kerala, India, pg.66 – 73

RESEARCH ARTICLE

A Personalized Travel Recommender Model Based on Content-based Prediction and Collaborative Recommendation

Shini Renjith[1], Anjali C[2]

[1,2]Department of Computer Science and Engineering, Mar Baselios College of Engineering & Technology, Trivandrum, India

[1]shinirenjith@gmail.com; [2]anjalichandrika@gmail.com

Abstract **-Mobile Recommender is the essential companion for a modern traveller. The collaborative travel recommender describes the architecture of a personalized mobile travel predictor by performing content-based filtering and inductive learning techniques depending on travel foot prints. Further the intelligent recommender is powered by collaborative techniques to resolve the multi-armed contextual bandit problem of the exploration/exploitation dilemma traveller may face in his endeavour.**

Keywords **- Travel; Mobile; Predictor; Recommender; Inductive Learning**

I. Introduction

Travel and Tourism has become one of the significant contributors in the global economy with 30% share of the world's exports of services. As per UNWTO records there were 980 Million international tourists in 2011 and expected to grow to 1.8 Billion by 2030 [15]. From ages, people travel a lot due to different reasons – social, political, economic, leisure, etc. Tourism aims the travel for recreational, leisure or business purposes. It serves as a major source of income to the economy of most of the nations in the world. The motive of a tourist may be to explore new locations, new food, and new experiences or it may be for research purpose. The key challenge one may face in this endeavour is the availability of a guide at a new location for opting among a large number of alternatives to consider (a case of multi-armed contextual bandit problem) [11].

Past transactions are the clear indicators of the psychological needs of an individual. Studies have proved that past behaviour patterns can be utilized to unveil the personal interests and thereby helps in predicting the future behaviour. This is the guiding principle behind most of the prediction systems. Predictions can be further refined to generate meaningful recommendations to the intended recipients. Technology based travel recommenders play a key role in modern tourism industry.

The guaranteed companion for a person living in the modern era is a mobile device. Mobile Recommendation systems can play amomentous role in such scenarios. A system that can recommend various options based on the predicted user interest using his/her past behaviour is highly relevant in the context. Inductive learning techniques can help in gathering the user interest patterns from the transactional data [16]. Based on the inferred user interests, the future action of a traveller can be predicted. Recommendations can be provided for his/her current geographical position. The relevance of the recommendations can be improved by making it personalized based on user preferences and trust related factors [17]. The proposed travel recommender system has a two-step strategy consisting of prediction and recommendation processes.

Section II deals with the methodology used for predictions and recommendations. Section III explains the system overview and the architecture of the proposed system. The discussion is given in section IV and finally the paper concludes in section V.

II. Methodology

The prediction process helps to estimate the forthcoming user action based on his/her past behaviours as those are considered as the best indicators for any future actions. The recommendation process suggests various travel options to the user in line with the inferred behaviour. Inductive Learning techniques perform the vital role in the content-based filtering based prediction. This model adapts collaborative filtering based approaches to perform the actual recommendations. In addition to the results obtained via exploitation, an exploration mechanism is also proposed in combination so that the best recommendations are presented to the traveller.

A. *PREDICTION*

The first step in architecting a recommender system is the decision on a prediction mechanism to forecast the next user action. Content-based filtering can be applied in identifying user interests based on the past transactions of the individual [1]. This can help in formulating a user interest profile by analysing the attractions/sites that the user has already visited. This model can help predicting user behaviour that is highly consistent with his/her interests.

Content-based filtering also referred to as cognitive filtering, helps in predicting future travel intentions based on a comparison between the attributes of an attraction and the user profile. Each attraction is represented as a set of attributes like the type of attractiveness of the place - beach, theatre, park, shopping mall, etc. The user profile is formulated with the same attributes and populated by analysing the attributes of the places/attraction, visited by the user.

There are multiple considerations that need to be addressed when implementing a content-based filtering system. As the first step, decision has to be made on the attributes under consideration. This can be formulated automatically or manually. For easier implementation, a predefined set of attributes can be considered (manual model).When attributes need to be populated automatically an intelligent algorithm can be chosen which can extract attributes from the attractions under consideration.

Next, the attributes have to be represented/ stored in such a way that both the user profile and the attractions can be compared in a meaningful way. Precisely there should be a one-to-many relationship existing between the data set for users and attractions. Table 1 highlights the approach suitable for the prediction model of the recommender system.

Table I: Key Highlights of the Prediction Model

Sl. No	Consideration	Method
1.	Algorithm	Content Based Filtering
2.	Approach	Inductive Learning
3.	Input	Past User Transactions
4.	Output	Predicted User Interest

Finally, a learning algorithm has to be chosen that should be able to learn the user profile based on the attractions visited and should be able to make predictions based on the user profile.

B. *INDUCTIVE LEARNING*

Relevance feedback, genetic algorithms, neural networks, and Bayesian classifier are some of the learning techniques that can be considered for learning a user profile [2]. The key selection criteria should be the computational complexity and storage requirements of the algorithm. Bayesian classifier and relevance feedback are superior when their computational speed is considered.

1. Relevance Feedback

Relevance feedback is an iterative process where the feedback on the relevance of the results of a given query is used to refine the query further. Vector Space Model is used for representing any objects as vectors in a multi-dimensional space. It can be used for information filtering, information retrieval, indexing and relevancy rankings [5]. The Rocchio Algorithm [6] which is developed based on Vector Space Model is an implementation strategy for relevance feedback. This algorithm is based on the assumption that each user has a view on which object (here the attribute associated with an attraction) can be denoted as relevant or irrelevant. The mathematical representation of the algorithm can be denoted as below.

$$\overrightarrow{Qm} = \left(a.\overrightarrow{Q0}\right) + \left(b.\frac{1}{|Dr|}.\sum_{\overrightarrow{Dj}\in Dr}\overrightarrow{Dj}\right) - \left(c.\frac{1}{|Dnr|}.\sum_{\overrightarrow{Dk}\in Dnr}\overrightarrow{Dk}\right)$$

where

$\overrightarrow{Qm}$	Modified Query Vector
$\overrightarrow{Q0}$	Original Query Vector
$\overrightarrow{D}$j	Related Document Vector
$\overrightarrow{D}$k	Non-Related Document Vector
a	Original Query Weight having value typically around 1.0
b	Related Documents Weight having value typically around 0.8
c	Non-Related Documents Weight having value typically around 0.1
D_r	Set of Related Documents and
D_{nr}	Set of Non-Related Documents

2. Genetic Algorithms

Genetic algorithms also follow the iterative process to generate the population of possible predictions by continuous evolution through a fitness function that ranks the solutions. The best solutions are retained and the worst ones are removed as the iteration continues. This is useful in learning the interests of a user from a set of attractions the user has visited and rated. The three genetic operators - reproduction, crossover and mutation are applied to formulate the potential profile across iterations. The fitness of a potential profile is determined using the cosine similarity function[7], [8] given below.

$$\cos\theta = \frac{A.B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A\mathrm{i}.B\mathrm{i}}{\sqrt{\sum_{i=1}^{n}(A\mathrm{i})^2}\,.\sqrt{\sum_{i=1}^{n}(B\mathrm{i})^2}}$$

Where A is the Document vector and B is the Query vector

3. Neural Network

Neural networks [9] can be used to model the user interests/ preferences. The neural network is formed and modified as a result of the attractions a user has visited or avoided. A new node is introduced as the result of a new attribute which is present in an attraction that is visited by the user. A link is formed between two nodes when corresponding attributes appear together in an attraction. The weight of the link is decided by the frequency in which the attributes are found together. The value of a node is increased when the user visit an attraction with that particular attribute and is decreased when an attraction with this attribute is avoided.

Mapping of the existing attributes to neural networks and activation of the matching nodes leads to the ranking of new attractions. A predefined value acts as the firing threshold for the network. If the input to a node connected to a collection of other active nodes is greater than this firing value, the former also becomes active. The sum of all the output values determines the rank of an attraction. This node firing is an iterative process, however the count of iterations in the network are subjected to its computational complexity.

4. Bayesian Classifier

The Bayesian classifier is a statistical method based on inverse probability theory which can be used for the classification purpose. Naive Bayes classifier assumes that the presence or absence of a particular attribute is not related to the presence or absence of any other attribute, given the class variable [10]. According to Bayes Theorem,

posterior = (prior x likelihood)/ evidence

Mathematically

$$P(C|F1,\ldots,F\mathrm{n}) = \frac{P(C).P(F1,\ldots,F\mathrm{n}|C)}{P(F1,\ldots,F\mathrm{n})}$$

In the travel prediction context, an attraction Acan be classified as relevant for a traveller if the probability that A belongs to classC that contains (interested attributes for the user)or does not contain(uninterested attributes for the user) is larger than the probability that A does not belong to class C given the attributes of A.

$$P(C|A) > P(\bar{C}|A)$$

where, $P(C|A) = \frac{P(C).P(A|C)}{P(A)}$ (according to Bayes' rule)

Under the assumption that the attributes of an attraction are conditionally independent, the probability is proportional to

$$P(C) \prod_{k}^{m} P(t_k|C),$$

where, t_k is the k^{th} attribute when there is total of m attributes.

C. *RECOMMENDER*

Once the user interests are made predictable, the system should be able to provide recommendations. Additional information required to make meaningful recommendations are the current user location and potential target attractions based on user preferences. The location information can be made easily available for processing by leveraging the GPS (Global Positioning System). The list of attractions at the locations is available in the recommender server database along with its attributes. The first cut filtering is directly achievable by using the user preferences setup at the time of enrolment (or refined later). Fig.1represents the recommendations generated using the location information and user preferences.

The next level of filtering makes use of the individualistic predictions generated using inductive learning techniques. In this stage, there may be multiple attractions that may qualify for presenting to the user. It is required to rank/order the attractions before performing the final recommendations.

Collaborative filtering [3] also referred to as social filtering, can be applied at this stage. It is capable of filtering and ordering attractions by using the recommendations/transactions of other people – the society. This approach is based on the principle that an individual who agreed in the evaluation of a group in the past is likely to agree to the same group again in the future. A person who wants to make a leisure trip for example, might ask for recommendations from friends who suggested an interesting travel spot in the past.

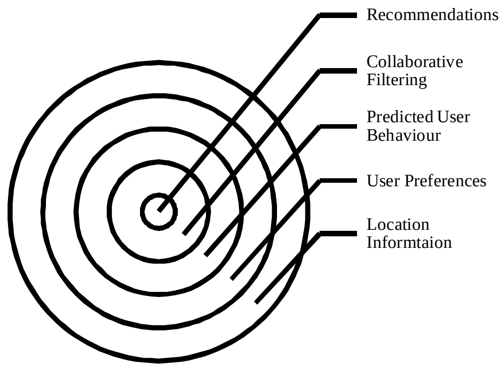

Fig. 1 Fine-tuning Recommendations

The approaches used for the recommendation portion of the system are shown in Table 2.

Table II: Key Highlights of the Recommendation Model

Sl.No	Sections	Contents
1.	Algorithm	Collaborative filtering
3.	Model	Memory based model
4.	Input	Predicted user behaviour
5.	Output	Recommendations

The key constraint of content-based filtering for prediction is that it restricts the recommendations to the type of attractions similar to those already visited. The interests of a user keep changing over time. The key expectation from a recommender system is that it could predict a user's future interests and recommend options that are entirely new to the user. This calls for two types of recommendation models [11],[12].

1) *Exploitation:* Attractions similar to those for which the user has already has an interest/ preference, ie. An existing attribute in the user interest profile having high rank – one with the highest frequencies.

2) *Exploration:* Attractions where the user profile does not provide evidence to predict the user's reaction, ie. An interest profile which is having zero matches against it for the user, but has high ranking in a collaborative system.

Recommendations can be generated by mainly using user-based (Neighbourhood based) or item-based algorithms.User-based algorithms tend to be more suggestible when there are more items than users, and item-based when the situation is reversed. Both these methods require minimal offline computation at the expense of somewhat greater computational demands when recommendations are generated [4].

1. Neighbourhood based Approach

Neighbourhood-based technique [13] is a user based collaborative filtering and is the most commonly used collaborative filtering model. In this approach a number of users are selected based on their similarity with the user under consideration. The prediction/ recommendations for the current user are performed by calculating a weighted average of the ratings provided by these selected users. The weight for a similar user is calculated based on the correlation of the user with the person for whom to make a prediction. Pearson correlation coefficient rfor a sample can be used as a measure of correlation here. The correlation between X and Y is then given by

$$r(X,Y) = \frac{\sum_{i=1}^{n}(X\mathrm{i} - \bar{X})(Y\mathrm{i} - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X\mathrm{i} - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y\mathrm{i} - \bar{Y})^2}}$$

Where Xi,Yi - ratings of person X and Y of the attraction i
$\overline{X}$, $\overline{Y}$- mean values of their ratings

So, the prediction for the rating of person X (i.e. the recommendation for X) of the attraction i based on the ratings of people who have rated item i can be computed as below

$$P(X\mathrm{i}) = \frac{\sum_{i=1}^{n} Y\mathrm{i} - r(X,Y)}{n}$$

The Spearman rank correlation coefficient, cosine similarity, mean-squared difference etc. also can be used as an alternative to Pearson correlation coefficient for measuring the similarity. In general, Pearson correlation has been found to provide the best recommendation results.

2. Item-to-Item Approach

Item-based collaborative filtering [14] is one of the most widely deployed collaborative filtering techniques in these days – an example is Amazon.com. This approach is an inversed variant of the neighbourhood-based approach. Here the similarities between ratings are taken into consideration for measuring the correlation between items. If two items tend to have the same users like and dislike them, then they are considered similar and users are expected to have similar preferences for similar items. The item similarity is deduced from user preference patterns rather than extracted from item data. In this model too the Pearson correlation coefficient can be utilized as a measure of correlation.

III. The System Overview

The proposed system consists of two phases, the enrolment phase and the recommendation phase.

1. Enrolment Phase

This phase covers the one time registration and preference setup in the user perspective. The user authentication information shown in [Fig.2] and the individual preferences like the preferred distances, accommodation likings, expenditure constraints, etc. at any attraction are captured in this stage. It also takes care of the administration operations required for site/attraction management like addition or deletion of sites and change in their attributes. There are two types of attributes for any attraction – static and dynamic[Fig.3]. Static information includes details like longitude, latitude, climate, average cost of expenditure etc. Dynamic attributes include the category of the attraction like beach, park, shopping mall etc. These dynamic attributes contains the same values as a user preferences which helps in a direct comparison.

User	**Authentication & Preferences**					**Interest Profile**				
	Id	**Password**	**Pref. Distance**	**Pref. Expenditure**	**…**	**Beach**	**Theatre**	**Park**	**Shopping Mall**	**…**
User 1										
User 2										
…										
User n										

Fig. 2 User Profile

Attraction	**Static Attributes**					**Dynamic Attributes**				
	Longitude	**Latitude**	**Climate**	**Avg. Expenditure**	**…**	**Beach**	**Theatre**	**Park**	**Shopping Mall**	**…**
Attraction 1										
Attraction 2										
…										
Attraction n										

Fig. 3 Attraction Profile

2. Recommendation Phase

This phase includes two types of activities. The first type of activities keeps track of the transactions performed by the user. This data is used to generate the user interest profile which is the base for the prediction process. Prediction process generates the potential attribute types that the user may interest to go. Any of the inductive learning techniques (covered earlier) may be chosen while attempting an implementation. For example, the past transactions of the user give a sense of the user's inclination towards a certain type of attraction that can be one of the most preferred choices even at the current scenario. Prediction process can be performed offline - may be at the start/end of each day to update the user's interest profile[Fig.4].

The second types of activities in the recommendation phase include the actual recommendation process. This is an online process. Here the current location of the user is the starting point for which the lists of available attractions are filtered out based on a match with user preferences and dynamically populated interest profile/ predicted user behaviour. The attractions in this list can be further ranked to form the final recommendations using collaborative filtering techniques. The explicit or implicit feedback from other registered users will lead to the ranking process - trust based recommendation systems rely on explicit feedback whereas implicit feedbacks are leverage in some systems based on neural networks.

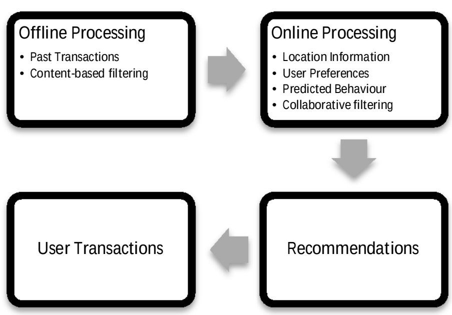

Fig. 4 Prediction → Recommendation→ User Action

The Global Positioning System (GPS) has an important role in the mobile system. The current positioning of a person is monitored with the help of the global positioning system. The location of a person is mapped with the latitude and longitude values obtained by the geocoding technique and later the recommendations are provided to the user through reverse geocoding. A spot is considered for recommendation process when the distance from the current position of the person to the attraction (the Euclidian distance) matches with the preferred distance set by the user.

There are 3 key database tables to be maintained as part of the proposed system –user table, attraction table and transaction table. The user table contains single record per user. It stores authentication, preferences and a tuple containing user interest profile. The user can manage the authentication and preference information by logging into the system. The interest profile will get automatically updated periodically, say once in a day by mining the user's transaction history which is available in transaction table. The attraction table contain the location specific information and a tuple containing the attributes of the attraction. The data in this tuple will be in match with the data in the user interest profile. The weight of each attribute in the tuple will get updated periodically, say once in a day by mining the transaction history at the particular attraction which is available in transaction table.

When a registered user goes to an attraction which is part of the system and spends specified time - say 4hours, a transaction record should get automatically added to the transaction table. (Logic should be implemented to avoid multiple entries if the user spends long duration at a place). This table gets updated frequently as user visits. The records from this table are used to perform content-based filtering to predict the user interest profile/ user behaviour pattern. This also can be leveraged to perform a collaborative filtering to get the rank of various attractions. This table undergo periodic refresh by trimming out very old transactions (say more than 2 year old) to make the predictions and recommendations more relevant.

The recommendations can be made available to the user in the form of push messages in the mobile device. If the user visits the place and spends a predefined time there it will be automatically entered in the table. User feedback may not be captured here. Being a social animal the user will automatically convey his feedback regarding the spots to his friends, through different media like blogs or direct interaction and will automatically affect the visit count.

One of the key questions in the model is that how the recommendations can be formulated when there is no transaction evidences exist for a user (like newly registered user). This will lead to empty set of recommendations from exploitation algorithm. In such a case the most popular attraction around the current user location should get recommended due to exploration algorithm.

The main difference in this work from the existing system is that the user need not manually update the interests. Also no feedback may be required to be fed into this system. It can/will be conveyed by some other blogs, talks, logs etc. which will automatically reduce/increase the visitors count. Minimal data need to be stored. Real time data is stored in transaction table only. This is because a seasoned user will not have a change in its preference overnight and it will usually change after a periodicity. So we need to update the table offline i.e. not real time so it is a very fast application.

IV. Discussion

The implementation of the current model is planned using genetic algorithm based prediction mechanisms and item based collaborative filtering for recommendations. This model focuses only on recommending attractions based on past behaviour of the user and societal feedback from other users. The same can be refined further to address the scenarios mentioned below and more.

A tourist using the personalized travel recommender can be provided with an additional helping hand by assisting him/her for identifying the best mode of transportation to reach the recommended destination. Also the recommendations can be refined further to accommodate external parameters that may impact the relevance of an attraction like rain, fog, snow, etc. The best mode of transportation can be recommended by extending the prediction and recommendation processes to cover various conveyance mechanisms associated with an attraction. Combining the same with user preferences and past behaviour will assists in achieving relevant recommendations. In this manner, the further study is planned to cover the following:

Examples for more parameters to fine tune recommendations are past transactions of the user, current seasonal/climatic conditions, demographic information, safety parameters, etc. Similarly examples for more coverage of recommendations are most relevant attractions, best and cheapest mode of conveyance, best accommodation facilities in vicinity, best deals for shopping etc.

V. Conclusion

A modern man spends considerable time in travel – either for business/job or for leisure. The person will come across various options to choose between as he moves. The best assistant a traveller possesses is the handheld devices, he/she carries. In this study, a travel recommender system based on predictions of user behaviour identified via inductive learning of the past transactions is proposed and designed. This system is capable of providing personalized recommendations for the users on the move. In the proposed architecture a location dictionary is maintained with large pool of data about possible target locations along with its attributes/features. Each visit of the user to any of these locations is recorded and serves as a repository that can be mined to identify the person's preferences. As soon as the person reaches a new location, recommendations can be provided to him/her based on the past behaviour pattern.

Acknowledgment

We acknowledge support from the Department of Computer Science and Engineering, Mar Baselios College of Engineering and Technology for all guidance, reviews, valuable suggestions and very useful discussions. Also we would like to thank Mr. Renjith Ranganadhan for his motivation and the encouragements afforded.

References

[1] Robin van Meteren and Maarten van Someren, "Using Content-Based Filtering for Recommendation," in *Proc.11th European Conference on Machine Learning (MLnet / ECML2000) Workshop*, 2000.

[2] Marco de Gemmis, Leo Iaquinta, Pasquale Lops, CataldoMusto, FedelucioNarducci and Giovanni Semeraro, "Preference Learning in Recommender Systems," in *Proc. 12th European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD-09) Workshop*, 2009, p. 41-55.

[3] J. Ben Schafer, Dan Frankowski, Jon Herlocker and ShiladSen, *Collaborative Filtering Recommender Systems,* The Adaptive Web, Lecture Notes on Computer Science 4321. Berlin, Germany: Springer, 2007, vol. 61.

[4] Michael D. Ekstrand, John T. Riedl and Joseph A. Konstan, *Collaborative Filtering Recommender Systems*, 1st ed., Now Publishers, 2011.

[5] YuanhuaLv and ChengXiangZhai, "Adaptive Relevance Feedback in Information Retrieval," in *Proc. 18th ACMConference on Information and Knowledge Management (CIKM '09)*, 2009, p. 255-264.

[6] Christopher D. Manning, PrabhakarRaghavan and HinrichSchütze, *An Introduction to Information Retrieval*, 1st ed., Cambridge University Press, 2012.

[7] John Pagonis and Adrian F. Clark, "Engene: A genetic algorithm classifier for content-based recommender systems that does not require continuous user feedback," in *Proc. 10th UK Workshop on Computational Intelligence (UKCI2010)*, 2010, p. 1-6.

[8] J. Usharani and K Iyakutti, "A Genetic Algorithm based on Cosine Similarity for Relevant Document Retrieval," *International Journal of Engineering Research & Technology (IJERT)*, vol. 2 (2), Feb. 2013.

[9] Manolis Wallace, IliasMaglogiannis, Kostas Karpouzis, George Kormentzas and StefanosKollias, "Intelligent One-Stop-Shop Travel Recommendations Using an Adaptive Neural Network and Clustering of History," *Journal of IT & Tourism*, vol. 6(3), pp. 181–193, Jan. 2003.

[10] Luis M. de Campos, Juan M. Fernández-Luna, Juan F. Huete and Miguel A. Rueda-Morales, "Combining content-based and collaborative recommendations: A hybrid approach based on Bayesian networks," *International Journal of Approximate Reasoning*, vol. 51, pp. 785–799, Apr. 2010.

[11] Tyler Lu, D´avidP´al and Martin P´al, "Contextual Multi-Armed Bandits," *Journal of Machine Learning Research*, vol. 9, pp. 485–492, 2010.

[12] DjallelBouneffouf, AmelBouzeghoub and Alda Lopes Gançarski, *Hybrid-e-greedy for Mobile Context-Aware Recommender System*, Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD) Lecture Notes in Computer Science. Berlin, Germany: Springer, 2012, vol. 7301.

[13] Weiqing Wang, Zhenyu Chen, Jia Liu, Qi Qi and Zhihong Zhao, "User-based Collaborative Filtering on Cross Domain by Tag Transfer Learning," in *Proc. 1st International Workshop on Cross Domain Knowledge Discovery in Web and Social Network Mining (CDKD '12)*, 2012, p. 10-17.

[14] BadrulSarwar, George Karypis, Joseph Konstan and John Riedl, "Item-Based Collaborative Filtering Recommendation Algorithms" in *Proc. 10th International World Wide Web Conference (WWW10)*, 2010, p. 285-295.

[15] (2013) World Tourism Organization UNWTO website. [Online].Available: http://www2.unwto.org/

[16] Jin'AnXu,ToshihikoItoh, Kenji Araki and Koji Tochinait, "A Point-Pass-Based Action Prediction Method," in *Proc. 4th IEEE International Symposium on Communications and Information Technology (ISCIT 2004)*, 2004, p. 103-108.

[17] F. Chong Tat Chua and Ee-Peng Lim, "Trust Network Inference for Online Rating Data Using Generative Models," in *Proc. 16th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD2010)*, 2010, p. 889-897.

