Personalized Dynamic Recommendation System for Tourism Using Hybrid Approach in Web Mining

[1]Priyanka B. Tiwari, [2]Krunal J. Panchal
Student-ME, Assistant Professor
Computer Engineering Department,
L.J I.E.T., Ahmedabad, India

Abstract -Today's rapid exponential growth in the use of internet has given boom to all areas having access to it. This has helped creating great opportunities and options for the businesses as well as customers. But this rapid growth has resulted complex situations for the customers to find the product and services they really want due to variation in prices and features. Also, providing accurate, valuable and personalized information for the customers, has become crucial for the businesses. Here, we have proposed a hybrid approach based on multi-dimensional user behaviour. It combines various techniques for recommendation and generates a flexible recommendation for both registered/unregistered users. For unregistered user it gives recommendation based on their past session whereas for registered users, it gives the result with the hybrid approach that considers the past history of the user itself as well as recent highly searched items by other users. This approach helps the users to get what is trending currently and they are unaware of.Evaluation of effectiveness is done using standard data set and by comparison with the existing system. For the research work, we have taken tourism domain as it is the most profitable and trending area now a days. The result shows that the personalized recommendation in a dynamic manner not only has direct impact on customer's interest and gross-sale, but also increase the loyalty to/for the customer.

Keywords – Web Data Mining, Data Acquisition, Data Pre-processing, Data Cleansing, Data Mart Development, Collaborative Filtering, Item-Based Filtering

I. INTRODUCTION

People are exponentially moving towards the E-services in a rapid manner due to the indulgence of internet in all areas that have access to it. These areas include various industries like E-commerce, Tourism, Medical, Education, Music, Film industry and many more. This trend of E-Services has given boom to marketplaces, leading vast and boundless commercial information for the scenario. This has helped creating great opportunities for businesses as well as customers. The customers have ease in finding theservices and products of their choice whereason the other hand the businesses have the opportunity to offer their services directly tothe customers in dynamic manner. However, the customer have to process massive information from the commercial gateway before selection of the actual thing that they are looking for.

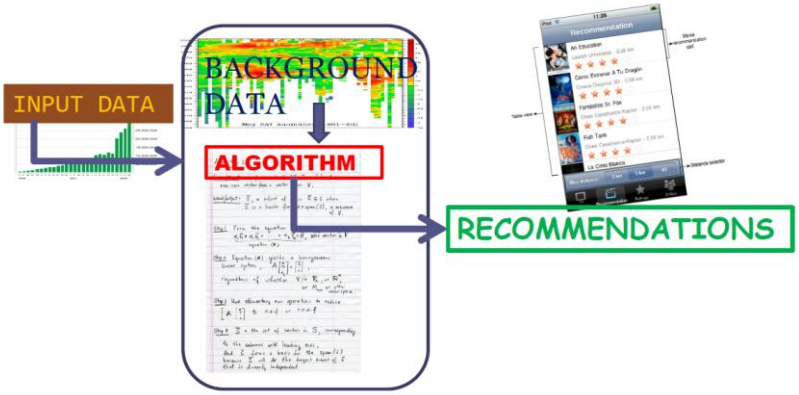

Figure 1. Recommendation System [12]

Recommender System, as kind of web-based support system, actively suggests a set of limited and ranked items from all available set of itemswithout taking direct input from the user. For this, in real situation it uses some form of data mining.

II. BASIC RECOMMENDER SYSTEM

Before getting into what the proposed model is all about, we are giving a brief idea about the basic recommendation system and its purpose. The purpose of any recommendation is to provide the users, the list of selected items which either they have viewed or of those items which are highly ranked by others. The recommender system is the system that uses some method or algorithm for making recommendations dynamically from a class of data sets. As all the data is present over the internet, the Web Mining is used. The recommender system includes the following basic processes:

1. Web Data Mining

Web data mining is the application of data mining techniques to discover the patterns in web content, structure and usage in various applications.

A. Data Acquisition

Data acquisition refers to the collection of data for mining purpose. This data is collected from three main sources: (i) web server, (ii) proxy server and (iii) web client

B. Data Pre-processing

Data processing is cleansing, formatting and grouping of web log files into the meaningful session for the sole aim of utilizing it for web mining purpose.

C. Data Cleansing

Data Cleansing is the stage in which irrelevant/noisy entries are eliminated from log files, removal of entries with "Error" or "Failure" status.

D. Data Mart Development

Data Mart is a logical subset of data warehouse. During this phase, the pre-processed data is stored in the warehouse.

After the data is stored in the warehouse, different recommendationtechniques can be implied.

2. Recommendation Techniques:

A. Content Based Filtering

The Content Based Filtering basically considers the attributes of the products. The algorithm recommends only best-matched products with similar attributes and the attributes get collected based on users' rating or the details.

B. Collaborative Filtering

The Collaborative Filtering approach basically considers the user-product interaction. This is based on collection and analysis of large amount of information available from user's behaviour, activity or preferences and also the prediction of what user would select based onsimilarity of other users.

C. Knowledge Based Filtering

The Knowledge Based technique is based on the explicit knowledge about interest of the user, item classification and recommendation standards. This is an alternative approach for Collaborative and Content based Filtering approach.

D. Demographic

The demographic technique is purely based on the users' information like gender, age, abilities, nationality,knowledge of languages, employment status, home ownership, location etc. The system does the recommendation based on the similaritiesin the demographic information.

E. Hybrid Approach

This approach implements the combination of two or moretechniques of recommendation.

Issues in Recommendation System

1. The web data keeps changing, and due to this, maintaining and updating the data warehouse is crucial.
2. The user's choice is different for each single user. So providing the recommendation in a personalized way it crucial.
3. Providing faster and more accurate recommendation to the client with desired qualities is complex.
4. Large number of in-frequent item set increase the space complexity and require too many data scans.
5. Uses complex algorithms in practical.

Due to these drawbacks, there is a necessity of making modifications in the recommendation techniques which are discussed in these papers. The next section contains the literature review done for the research work.

III. RELATED LITERATURE REVIEW

Numerous researches are available for web recommendation system. The term Web usage Mining was defined for the first time by Cooley and et.al which aims at predicting user's preferences and behaviour[10]. In [1], Prajyoti Lopes, Bidisha Roy, developed a dynamic recommendation system for all registered/unregistered visitors of the website. It was based on Action Based Rational Recommendation Technique which used formation of Lexical Patterns for products in each session and then its frequency and timestamp was counted. Another paper, presented recommendation to the user in efficient manner and schedule all activities of the users, using three different algorithms has namely HKB(Historical Knowledge Base), Cookies and Time Scheduling [2]. The data in the HKB was processed and ontology was made from where the matrix was used to find patterns for mining and ranking. In [3], Xuesong Zhao, KaifanJi, implemented Collaborative Filtering with implication of association rules. In User based collaborative filtering, matrix was constructed for evaluations on different products, then nearest neighbour set was found using KNN and then recommendation result was generated for target users. For final recommendation results, Association Rules was applied to discover relation between large data sets by finding minimum support and minimum confidence thresholds.

In [4], Xiaosheng Yu, Shan Sun, used recommendation engine that did recommendation based on the customer's browsing record and Case Study of Amazon.com online bookstore was shown which used Item-to-Item Collaborative Filtering, which didn't match between customers but between the items purchased. Liu Hai- Ling , Li Jun-Husi, Peng Jun, in [5] with objective to provide online recommendation system for optimized travel routes with least costs by integrated application of GIS, data mining, presentation recommendation technology and Web and GIS Technology based on both user and attraction data. Aymen El Kalifi ,Firas Ben Kharrat, Rim Faiz, presented a method for recommendation using live user interaction for providing best list of hotels[6]. For this, recommender algorithm used multidimensional user behaviour. In the proposed system, Web Services API traced all the user action, which got stored in database layer with additional information like date, time,user, kind of action, items.And then in recommendation layer, actions were grouped by type and then prediction was calculated. Each prediction had specific weight. Finally one RS was returned to Web Services API, for results.Use of java API (open service) has been done to extract hotel data from website TripAdvisor. In paper [7], Jen Hsiang Chen, Kuo-ming Chao, Nazaraf Shah have focused inpredicting active users' interest by getting a set of highly rated tourism places collecting information about preferences and taste of other tourists. For this, Item-Based Collaborative Filtering has been used. Also, an approach to minimize the travelling cost by scheduling travelling path from a set of selected places using genetic algorithm has been done. Another paper [8], focus on tourism e-commerce. For this, Collaborative Filtering and Association Rules of mining were implied for representing web mining based tourism e-commerce recommender system. Using CF, recommendation of top N list of products and services was done by finding nearest neighbour set and filtering implied item-based clustering or user-based clustering. Then, association rules were applied with an aim to increase the cross sales of tourism e-commerce. Honvia Zhang, Yuan Yang, in [9], focused on developing a personalized recommendation system model based on customer feedback, where results could be dynamically adjusted based on customer feedback information. The whole process consisted of two parts: Offline Mining and Online Recommendation. In Offline Mining, product data as well as customers' feedback was collected and processed. In online mining, the recommendation engine matched the mode according to customers' present conversation and produced personalised recommendation page set and got enclosed in the latest page of customer inquiry. The feedback included both Implicit and explicit feedbacks. The overall browsing behaviour of the user was stored in tabular form and degree of customers' interest was calculated based on time spent and number of hits.

A number of other methods and techniques have been implied by different researches and scholar.

IV. RESEARCH GAP

The problem exists that therecommendation techniques that have been applied are having either user based, item based or action based in particular creating limitation for the users. And the model providing recommendation using both techniques together are based on predetermined scheduling which makes the system rigid.

V. PROPOSED WORK

In proposed technique, hybrid approach has been applied for making the system more dynamic and in more personalized for the users. The focus is on generating the system that is more efficient than the existing one. Initially the system identifies whether the user is registered or not. Based on the user records, users' history is analysed. Then a combination of item –based and user- based filtering is applied for generating the list of items. Then collaborative filtering is applied on the generated list to make the final recommendation.

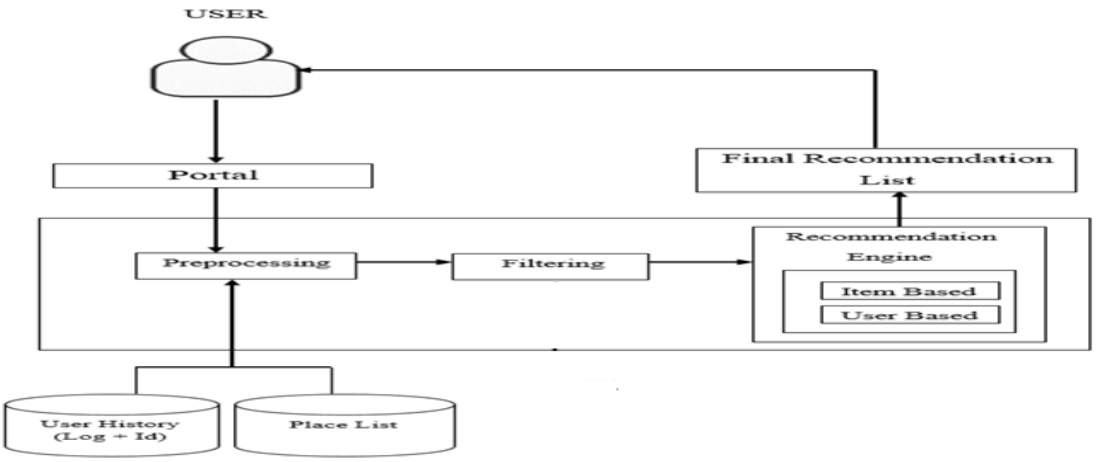

Figure 2. Architecture of Proposed Model

The system works for both registered and unregistered user. It makes use of the historical knowledge for the registered user whereas for unregistered it finds the ip-address of the user and then collects last several session. Based on the session records, it generates the list of items. The user interacts with the web portal and from these the system identifies whether user is registered/unregistered. Then based on the users' category, it is decided whether to consider the last few sessions or the overall access history. It uses different approaches for different types of user.

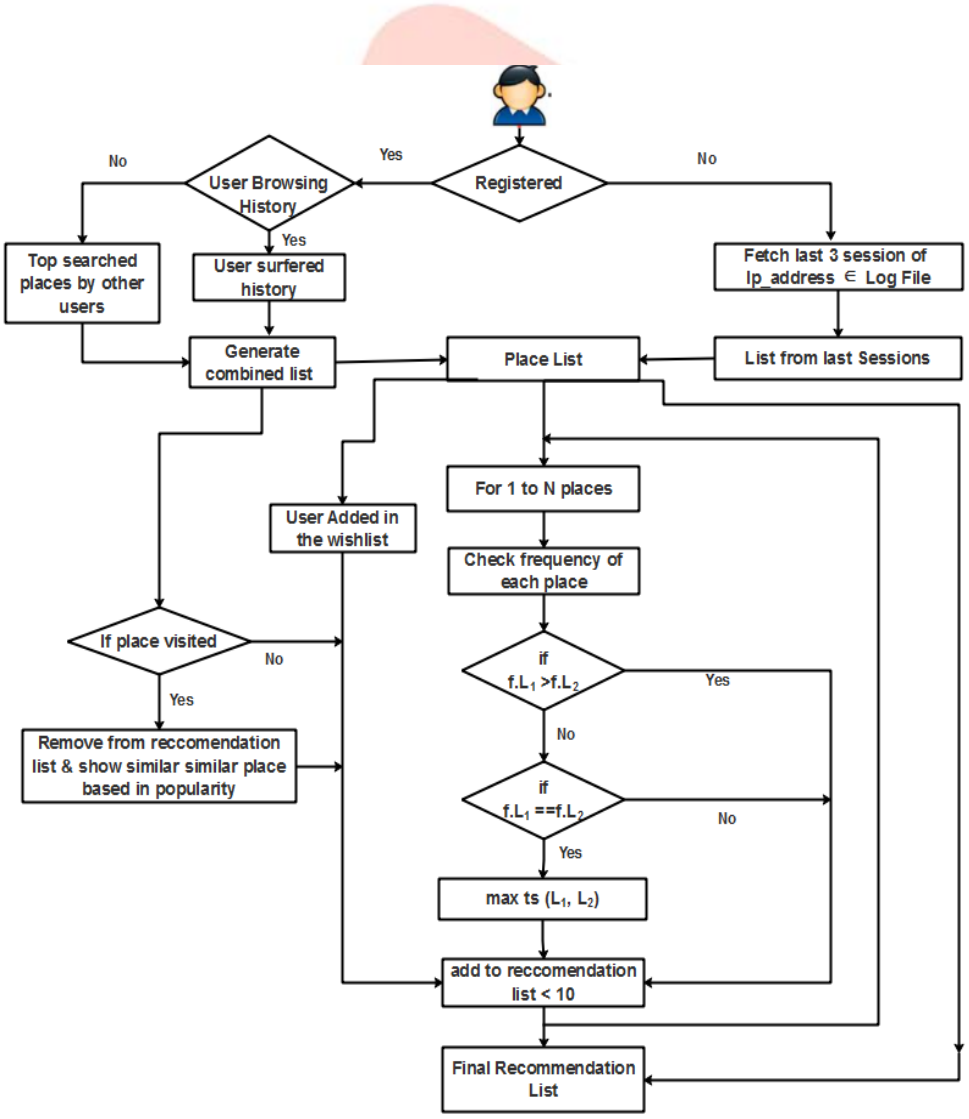

Figure 3. Diagram of Proposed Model

A. *Hybrid Approach For List Generation*

It checks whether the user's history is available or if he is the newly registered user. If the history is available it takes the places which are highly surfed by the user. Also it considers the recently searched places by other users having higher frequency. Here, the hybrid approach is applied for generating the combined list of places that are highly preferred and most frequently visited. And when the system finds that the user is unregistered , it considers the last few sessions of the user that is considered based on the IP address of the System.

B. *Recommendation Results*

In this, the final recommendation list is generated. Recommendation is made based on the category of the user. After the list of N places generated from the hybrid approach for registered user and last sessions of the unregistered user, the users' interest measure is calculated before considering the places in the recommendation list. For this, the frequency and the time stamp are used. Figure 3 explains the work flow of the proposed model.

C. *Basic Algorithm For The Proposed Model*

In order to enhance the dynamicity and making the system more personalized, hybrid approach has been proposed and following are the steps.

Step 1: User interacts with the system.
Step 2: System finds whether the user is registered or not.
Step 3: Apply hybrid approach for recommendation based on the user's history.
Step 4: Generates a combined list of highly preferred items.
Step 5: Checks the items with more frequency and adds it to the recommendation list.
Step 6: For the items with equal frequency, it checks the time stamp and adds the item with higher time stamp to recommendation list.
Step 7: Merges all the items filtered byhybrid approach.
Step 8: Generates final recommendation list.

VI. EVALUATION PARMETERS AND RESULT ANALYSIS

Evaluation of effectiveness is done using standard data set and by comparison with the existing system. For this, we have taken tourism domain as it is the most profitable and trending area now a days. Accuracy is the parameter that is used to evaluate the system. A matrix constructed for measuring the accuracy is in Table 1 as below.

Table 1 Recommendation Matrix

	Recommended items by the System	**Items Not Recommended by the System**
Expected Items	True Positive	False Negative
Not an Expected Items	False Positive	True Negative

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Based on these recommendation matrix and formula, the evaluation was done using 22786 records of standard dataset (Expedia) as compared to the existing system that used 1121records. We got the result with 65-70% accuracy for unregistered user and 75-85% accuracy for the registered users which is better than given by the existing system.

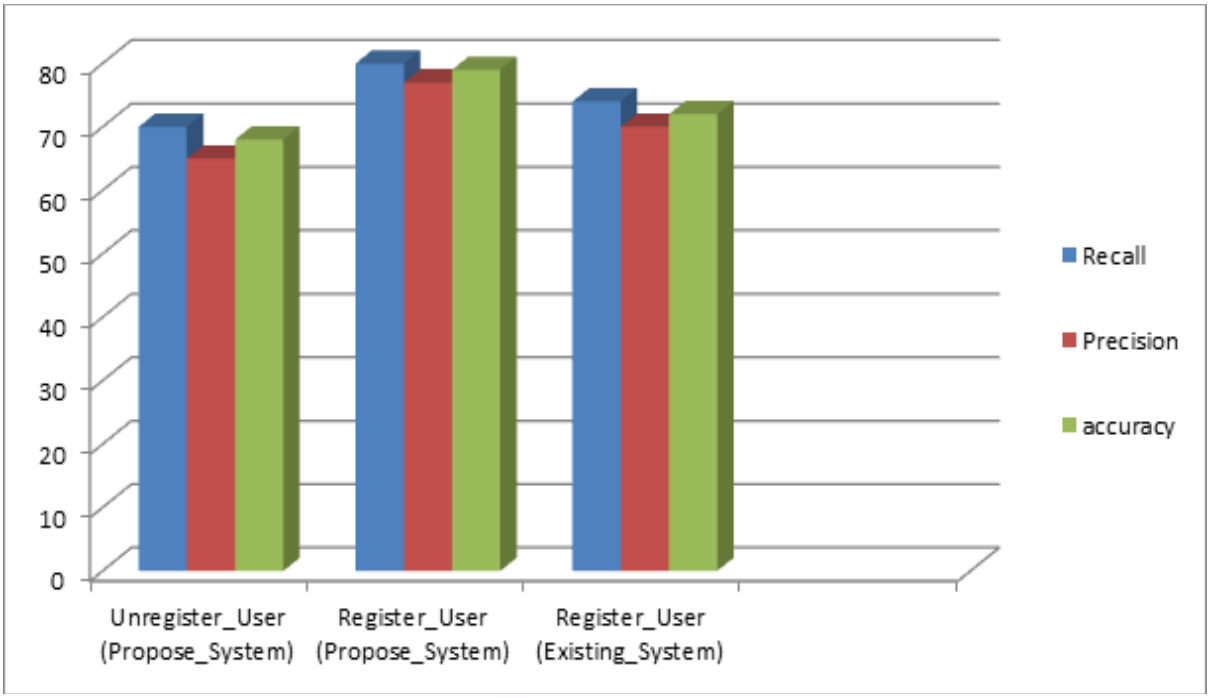

Figure 4. Result Comparison from Existing Approach

VII. CONCLUSION

Recommendation system is the base for providing needful and accurate data to the users, in the world of internet, nowadays. In this proposed model, use of hybrid recommendation technique has been done for providing data in more personalized manner and also maintaining the precision with changing user behaviour to maintain dynamicity. Even we have tried to add up the usage behavior of other users for enhancing the recommendation, so that the user may be benefited with the current trend that he/she is not aware of. Proposed system has several advantages like more personalized, dynamic, precise and efficient. It can be used in several domains like e-commerce, tourism, medical, entertainment and media, etc. Here, tourism domain has been taken for implementation of the proposed model, in which the user will be getting recommendation based on mutual browsing behavior in more flexible manner

VIII. REFERENCES

1) LopesPrajyoti, and Bidisha Roy. "Dynamic Recommendation System Using Web Usage Mining for E-commerce Users." Procedia Computer Science 45 (2015): 60-69. doi:10.1016/j.procs.2015.03.086.

2) AgarwalJuhi, Nishkarsh Sharma, Pratik Kumar, VisheshParshav, AnubhavSrivastava, and R.h. Goudar. "Intelligent Search in E-Tourism Services Using Recommendation System: Perfect Guide for Tourist." 2013 7th International Conference on Intelligent Systems and Control (ISCO) (2013): 410-415. doi:10.1109/ISCO.2013.6481190.

3) Zhao, Xuesong, and KaifanJi. "Tourism E-commerce Recommender System Based on Web Data Mining." 2013 8th International Conference on Computer Science & Education (2013): 1485-1488. doi:10.1109/ICCSE.2013.6554161.

4) Yu, Xiaosheng, and Shan Sun. "Research on Personalized Recommendation System Based on Web Mining." 2010 International Conference on E-Business and E-Government (2010):346-349. doi:10.1109/ICEE.2010.95.

5) Hai-Ling Liu, Jun-Huai Li, and Jun Peng. "A Novel Recommendation System for the Personalized Smart Tourism Route: Design and Implementation." 2015 IEEE 14th International Conference on Cognitive Informatics & Cognitive Computing (ICCI*CC) (2015): 291-96. doi:10.1109/ICCI-CC.2015.7259400.

6) Elkhelifi, Aymen, Firas Ben Kharrat, and Rim Faiz. "Recommendation Systems Based on Online User's Action."2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (2015):485-490. doi:10.1109/CIT/IUCC/DASC/PICOM.2015.69.

7) Chen, Jen-Hsiang, Kuo-Ming Chao, and Nazaraf Shah. "Hybrid Recommendation System for Tourism."2013 IEEE 10th International Conference on E-Business Engineering (2013): 156-161. doi:10.1109/ICEBE.2013.24.

8) Zhao, Xuesong, and KaifanJi. "Tourism E-commerce Recommender System Based on Web Data Mining." 2013 8th International Conference on Computer Science & Education (2013): 1485-488. doi:10.1109/ICCSE.2013.6554161.

9) Zhang, Hongxia, and Yuan Yang. "An E-Commerce Personalized Recommendation System Based on Customer Feedback." 2011 International Conference on Management and Service Science (2011): 1-3. doi:10.1109/ICMSS.2011.5998970.

10) Cooley, Robert, BamshadMobasher, and Jaideep Srivastava. "Data Preparation for Mining World Wide Web Browsing Patterns." Knowledge and Information Systems 1.1 (1999): 5-32. Web.

11) Wang, Ya-Min, Xue-Ling Han, and Xiao-Wei Liu. "E-commerce Recommendation System Based on CBR and Web Log Mining." 2011 IEEE 18th International Conference on Industrial Engineering and Engineering Management (2011): 311-315. doi:10.1109/ICIEEM.2011.6035165.

12) http://www.cse.lehigh.edu/~munoz/CSE335/classes/ComanRecommender.pdf