International Journal of Computer Science and Engineering (IJCSE)
ISSN 2278-9960
Vol. 2, Issue 4, Sep 2013, 83-90

International Academy of Science, Engineering and Technology
Connecting Researchers; Nurturing Innovations

A HYBRID RECOMMENDER SYSTEM TO PREDICT THE LOCATION RANKING

AKSHITA[1] & SMITA[2]

[1]Research Scholar, Department of Computer Science and Engineering, PDMCEW, Bahadurgarh, Haryana, India

[2]Associate Professor, Department of Computer Science and Engineering, PDMCEW, Bahadurgarh, Haryana, India

ABSTRACT

Data Mining is about to perform the intelligent analysis on available statistics to derive the new results as well as the future aspects. This future aspect analysis is formed as a prediction system or the recommender system. The recommender system is about to identify the knowledge about the similar user or the event and derive the favorable aspect based on it.

In this present work, a hybrid recommender system is defined to identify the most favorable location for a user to spend his vacations. The presented hybrid model is based on the content based similarity as well as event based similarity. This collaborative system is implemented with authenticated dataset. The obtained results from the system gives acceptable results in terms of accuracy.

KEYWORDS: Collaborative, Hybrid, Location Based, Recommender System

INTRODUCTION

Data mining can be viewed as a result of the natural evolution of information technology. An evolutionary path has been witnessed in the database industry in the development of the following functionalities data collection and database creation, data management (including data storage and retrieval, and database transaction processing), and data analysis and understanding (involving data warehousing and data mining).

With numerous database systems offering query and transaction processing as common practice, data analysis and understanding has naturally become the next target. Efficient methods for on-line transaction processing (OLTP), where a query is viewed as a read-only transaction, have contributed substantially to the evolution and wide acceptance of relational technology as a major tool for efficient storage, retrieval, and management of large amounts of data. Several data mining tasks exist [5]. Here figure 1 is showing the basic data mining process flow.

The process begins with a valid data search or the collection. Once the database is generated or retrieved, the next work is to perform the selection of most required attribute and row set from this dataset. After this the cleaning and the pre processing stage is seen as the filtration stage. Finally the data evaluation and interpretation will be performed on this dataset to obtain the exact results from the system.

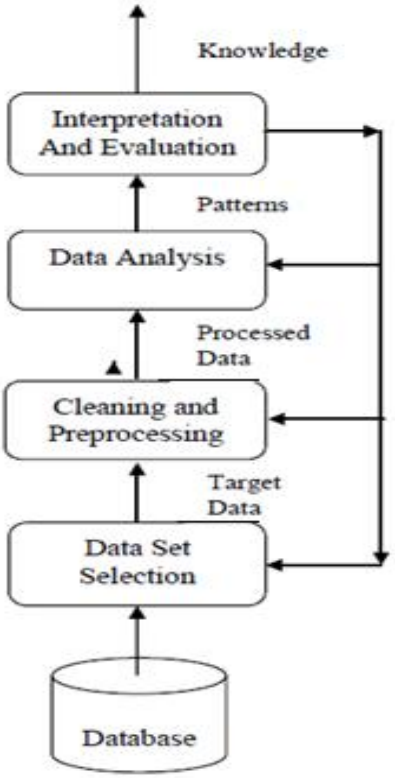

Figure 1: Data Mining Process

Recommendation techniques are information agents that attempt to predict which items out of large pool a user may b interested in and recommend the best one to the target user. Recommendation techniques have a number of possible classifications.

The whole classification is broadly categorized into the personalized and non personalized recommendation, and discusses all the personalized recommendation techniques shown in Figure 2.

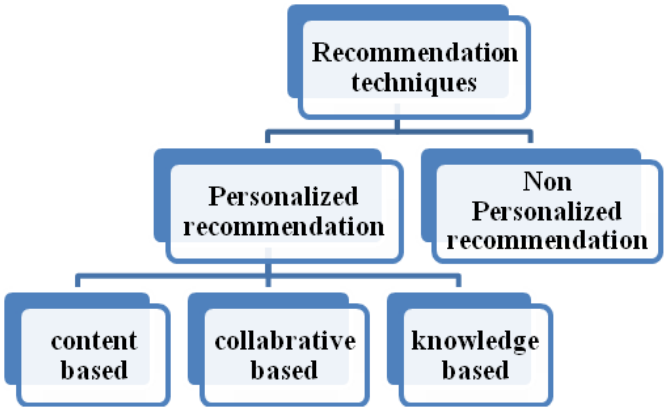

Figure 2: Classification of Recommendation Techniques

Personalized recommendation is an enabling mechanism to overcome information overload occurred when shopping in an internet marketplace, use personalized information for better recommendations to the user.

Non personalized recommendations are the simplest form of recommendations in which without any consideration of user's specifications some items are recommended. The most popular method is the recommendation based on ranking of items. However, since they don't take user's preferences into account, the quality of their results are low. For example in an electronic shop most sold items are recommended to all users [4].

Content based recommendation systems analyze item descriptions to identify items that are of particular interest to the user [8]. For instance, if a Netflix user has watched many cowboy movies, then recommend a movie classified in the database as having the "cowboy" genre. Collaborative based recommendation systems recommend items based on similarity measures between users and/or items. The items recommended to a user are those preferred by similar users.

Knowledge-based recommendation attempts to suggest objects based on inferences about a user's needs and preferences. Knowledge-based approaches are distinguished in that they have functional knowledge: they have knowledge about how a particular item meets a particular user need, and can therefore reason about the relationship between a need and a possible recommendation.

EXISTING WORK

In year 2004, Chien-Chih Yu presented a Framework on customer oriented intelligent decision support system. It was a web based frame work to personalize the B2C web e-services. The work includes the personal management, auction, negotiation, evaluation, planning, collaboration, transactions, payments, feedback and quality control. In this framework, almost all the business oriented functionality is implemented and described for investment and the tourism applications. The work is about to improve the efficiency and the effectiveness of the decision support system in the same area [1]. In 2010, an e-learning based decision support system was presented by Marta Zorrilla for E-Learning. The distance course learning and teaching process is processed in this application. The work includes the course description, course assignment, and other course based data mining operations are been handled by the system. The system includes the pattern based model along with probability analysis for the decision making. The work will help the instructor to answer all the student query and to find the outcome of standard process performed. The work also includes the reporting tool to present the work in an effective way [6].

In year 2007, Mohammed N. A. Abdelhakim presented a work on intelligent decision making for the evaluation and selection of educational multimedia. The work is a web based group decision system that will perform a statistical analysis on education provider with continuous evaluation to investigate the requirement for developing the educational application. The work also include the knowledge management along with design and implementation of performance evaluation to present the collect and process the data from instructors, producers and propose a solution for the educational consumers for the evaluation of the system[7].

In year 2008, Suresh Kalathur presented a work on the data mining operation on student driven content analysis while online teaching. The work is presented in the form of a web model integrated with data mining operations to handle the classroom discussions to predict the student faring. The model also provides the feedback on student discussion regarding the topic discussion in class, and a comparative analysis with other topics. The analysis includes the text mining operations relative to the answers submitted by the students [9].

Another mining based analysis on academic data is performed by J.M.Lauria in year 2012 for the analysis of college student retention. Author has presented an analytical research on academic risk using data mining approaches. The work is presented in the form of methodological framework to develop the query based model to analyze the course management respective to academic records and the classification process is performed to work on selected dataset [2].

In Year 2008, Kleanthi Lakiotaki performed a work," UTA-Rec: A Recommender System based on Multiple Criteria Analysis". UTARec, a Recommender System that incorporates Multiple Criteria Analysis methodologies is presented. The system's performance and capability of addressing certain shortfalls of existing Recommender Systems is demonstrated in the case of movie recommendations. UTARec's accuracy is measured in terms of Kendall's tau and ROC

curve analysis and is also compared to a Multiple Rating Collaborative Filtering (MRCF) approach [3]. In Year 2008, Juan A. Recio-García performed a work," Prototyping Recommender Systems in jCOLIBRI". Presented goal is to support system developers in rapid prototyping recommender systems using Case-Based Reasoning (CBR) techniques. In this paper Author describe how jcolibri can serve to that goal. Jcolibri is an object-oriented framework in Java for building CBR systems that greatly benefits from the reuse of previously developed CBR systems [10].

PROPOSED WORK

In this present work a hybrid architecture is defined to perform the recommendation about the location selection to spend the vacations. In this system, we have defined a dataset with three main tables. First table contains the details related to the user such as age, gender, occupation etc. The another table is defined to represent the locations. The locations are defined in terms of distance, type of location, cost factor etc. The third table is the rank table that defines the ranking allotted by different users to a particular user. The ranking is here assigned between 1 and 5. The presented system is divided in three main layers.

First, the content based matching is performed to identify the most similar users from the dataset. Then matching is performed under different attributes such as age, gender and occupation. The age and gender are taken with higher priority factors whereas the occupation is having the least priority contribution.

Once the similar users are identified, Second the event based match is performed. Here the event is described as the ranking assigned to a particular location. Each user assigned some ranking to each location. The analysis is here performed on these assigned ranking. For this some ratio analysis mechanism is used in which the ranking assigned to particular location by the similar users is identified and a ratio is generated. The another factor while performing this analysis is the temporal factor. The temporal factor is here defined as the time based analysis. It means instead of analyzing the whole dataset, a selective dataset is taken for the recommendation process. This dataset is the most recently recommended locations by the users. Based on the temporal factor based model, the ratio analysis is obtained.

Now in last, these two vectors are combined to obtain the final result. Equal weightage is assigned to both kind of analysis and finally the rank for a particular location by a particular user is identified. Here these aspects are described in detail

Content Based Analysis

User Attribute Similarity is calculated by using Demographic information of users. The main idea behind making predictions using demographic data is the assumption that people with similar characteristics enjoy similar Sites. It is believed that age, gender, occupation and hometown play an important role on Site preferences and a set of users who have a high level of demographic similarity with the target user is found. Then the similarity is used as initial value for user-based similarity calculation.

The similarity between two users can be calculated as:

$$sim_D(u_1, u_2) = \frac{\sum_{f \in F} w_f \, X \, sim(u_{1f} - u_{2f})}{\sum_{f \in F} w_f} \quad (1)$$

Where *f* represents a feature of the user from the set of all demographic features *F*,

w represents the relative weight of feature *f*,

u_{1f} and u_{2f} represent the values of *f* for u_1 and u_2

Rank Similarity

Rating similarity is calculated by using Pearson Correlation Coefficient. Pearson's correlation coefficient is a measure of the strength of the association between the two variables.

Let U be the set of users, P be the set of items and data as set of triplet (i, x, r), where i ϵ U is a user, x ϵ P is an item and r is a rating of item x by user i. Moreover, r_{ix} denotes the rating of item x by user i. P_i is subset of P that denotes the set of item rated by i. U_x denotes set of users that have rated item x. User Rating similarity is given by UB-PCC as

$$u_{ij} = \frac{\sum_{x \in P_i \cap P_j}(r_{ix} - \overline{r_i})(r_{jx} - \overline{r_j})}{\sqrt{\sum_{x \in P_i \cap P_j}(r_{ix} - \overline{r_i})^2 \sum_{x \in P_i \cap P_j}(r_{jx} - \overline{r_j})^2}} \quad (2)$$

Where $\overline{r_i}$ is the average rating user i give to all items and $\overline{r_j}$ is the average rating user j give to all items, r_{ix} is the rating given by user i to item x and r_{jx} is the rating of j to item x.

Hybrid Approach

On the basis of calculated similarities, predictions are generated. This can be done using class formation. As after determining the similarities, various classes can be formed on the basis of similarity values. The rating for an unseen item or a new user can be determined.

The basic flow of the presented work is shown in figure 3.

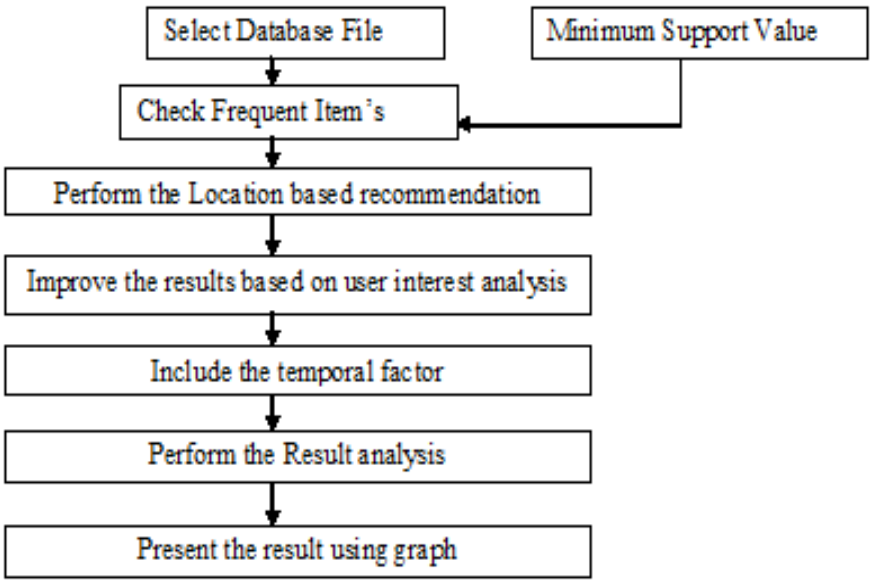

Figure 3: Flow Chart of Proposed Work

RESULTS

The presented work is implemented on an authenticated dataset in matlab environment. In this work, we have taken three tables called user table, location table and ranking table. At first the similarity based match is performed on user table to identify the similar users to process. In second layer, the rank similarity is performed on all three tables collectively and derives the ratio analysis. Finally these two are collected to obtain the rank for the particular location. To present the results, the dataset is divided in two sets called training set and testing set under 10 fold methods. According to this, 90% record dataset is taken as the training dataset and 10% is taken as the testing dataset. Now the analysis is performed on this testing dataset and the ranking is predicted. These predicted values are compared with existing ranked

values. The difference between these two values is taken as the error. The analysis of this work is done based on this error analysis. The Results obtained from this error analysis is given as under.

Table 1: Error Analysis

Method	Mean Absolute Error
10 elements dataset	0.131
20 elements dataset	0.139
30 elements dataset	0.127

As we can see in table 1, as the numbers of testing records are increased the accuracy of obtained results is also increased. The system is providing about 87% accuracy level.

Below graphs shows MAE calculation fir different size testing datasets.

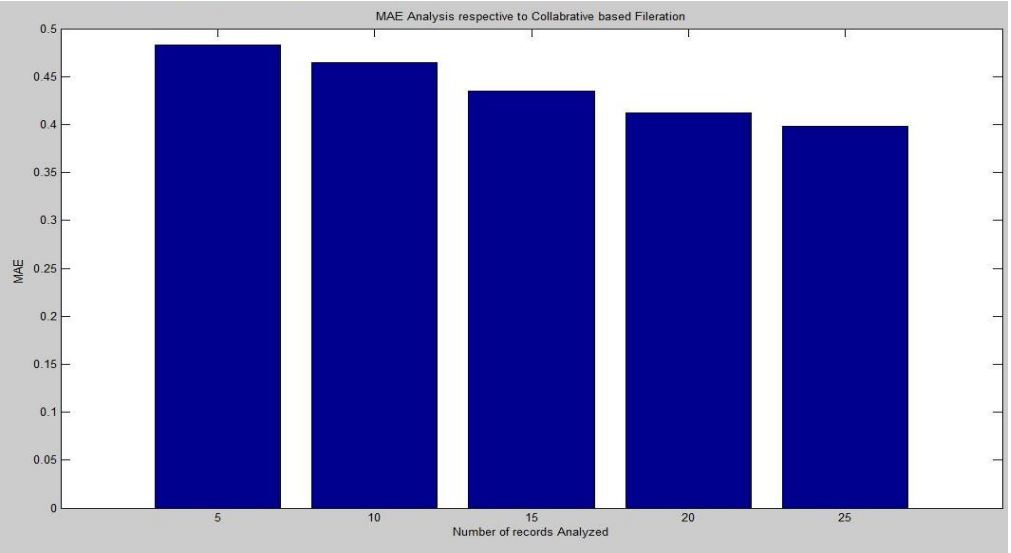

Figure 4: MAE Analysis Using Proposed Collaborative (Hybrid) Filtration

CONCLUSIONS

In this present work a hybrid recommendation system is presented to predict the user interest location based on similarity match. In this work, content based and rank based similarity measures are obtained and merged to obtain the collective results from the system. The obtained results show the effectiveness of the system in terms of higher degree of accuracy.

REFERENCES

1. Chien-Chih Yu," A Web-Based Consumer-Oriented Intelligent Decision Support System for Personalized E-Services".2004 Proceeding of the 6th ICEC 1-58113-930 (PP 229-237).
2. Eitel J.M. Lauría, Joshua D. Baron" Mining academic data to improve college student retention: An open source perspective", LAK'12,Proceeding of 978-1-4503-1111-3/12/04
3. Kleanthi Lakiotaki," UTA-Rec: A Recommender System based on Multiple Criteria Analysis", RecSys'08, October 23–25, 2008, Lausanne, Switzerland. ACM 978-1-60558-093-7/08/10 (pp 219-225)
4. Kyumars Sheykh Esmaili, Mahmood Neshati, Mohsen Jamali, Hassan Abolhassani and Jafar Habibi, "Comparing Performance of Recommendation Techniques in the Blogsphere" ,in ECAI 2006 workshop on recommender system.

5. Lukasz Kurgan and Petr Musilek (2006); A survey of Knowledge Discovery and Data Mining process models. The Knowledge Engineering Review. Volume 21 Issue 1, March 2006, pp 1–24, Cambridge University Press, New York, NY, USA doi: 10.1017/S0269888906000737.
6. Marta Zorrilla," A Decision Support System to improve e-Learning Environments", 978-1-60558-990-9/10/03
7. Mohammed N. A. Abdelhakim," A Web-Based Group Decision Support System for the Selection and Evaluation of Educational Multimedia", Emme ’07 Proceedings of the international workshop on Educational multimedia and multimedia education, 978-1-59593-783-4/07/0009.
8. Michael J. Pazzani and Daniel Billsus, “Content-Based Recommendation Systems”, the Adaptive Web, LNCS 4321, pp. 325 – 341, 2007.
9. Suresh Kalathur," Enriching Student Experience with Student Driven Content while Teaching an Online Data Mining Class", SIGITE’08 Proceedings of the 9th ACM SIGITE conference on Information technology education, 978-1-60558-329-7/08/10, pp 125-130.
10. Juan A. Recio-García, Belen Diaz Agudo "Prototyping Recommender Systems in jCOLIBRI", RecSys’08, Proceedings of the 2008 ACM conference on recommender system, October 23–25, 2008, Lausanne, Switzerland. ACM 978-1-60558-093-7/08/10 (pp243-250).