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

IJCSMC, Vol. 3, Issue. 7, July 2014, pg.222 – 229

RESEARCH ARTICLE

Recommender System

Varsha Negi

Student, M.Tech (Computer Science Engineering), MERI College of Engineering & Technology

(Email_id: varshanegi0403@gmail.com)

Abstract- One of the major data mining applications is Recommender System. It is the intelligent system that basically investigate the dataset present in existing system and based on which it will give some suggestions to the user regarding further process. These recommender systems are generally application specific and work on certain parameters. In this present work we define a hybrid recommender system for the movies ranking. A movie based recommender system suggests the user about the movie that he should rank after performing the intelligent analysis. In this present work, we are defining three dimensions to get the concept of hybridization. This kind of dataset having two main dimensions called users, movies and the relationship. The first level analysis will be based on user side where the content based weighted similarity analysis will be performed. Once the similar users will be identified, the next work is performed on movie side. The similar movies based on different aspects are identified using content based weighted analysis. At the third level, the similarity analysis between the relationships is identified using collaborative analysis. To perform the collaborative analysis, correlation coefficient will be used. Once these three level analysis will be completed, the next work is to conclude the relationship using weighted approach. The weightage will be applied all three methods and obtain the analysis. Another improvement here defined is the analysis under the temporal vector. It means, instead of analysis on whole dataset, the dataset in same time domain will be considered only. The work will be implemented in Matlab environment.

I. INTRODUCTION

One of the effective terms defined along with data mining is recommendation system. The recommender system is some kind of predictive system that actually analyzes the users or the customer behavior and based on this on this analysis; recommend some product or service to the user. Today most of the available product or service providers are connected through the web. These all users gets the feedback or the reviews from the users in an online system. These products or services are available online. Based on the user interest analysis to these products or services, a recommender system suggests some product or service to the user based on user feature analysis. It actually

compare the new user features with existing user features so that the decision regarding the product or service purchase. The recommender system is also defined as the approach that guide the user in a personalized way so that he can select the most useful item from the large set of available alternatives. There are number of online recommender systems adapted by different web service providers such as Amazon uses the recommender system to guide the users about the book or article purchase. In this research work we have defined a hybrid recommender system by combining the content based and collaborative recommendation as the single unit. .

II. PREVIOUS ARTS

Many approaches have been proposed in order to produce quality recommendations. Gediminas Adomavicius and Alexander Tuzhilin [1] in their paper entitled "Towards the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions", have presented an overview of the field of recommender systems along with description of the current generation of recommendation methods that are usually classified into the following three main categories: content-based, collaborative, and hybrid recommendation approaches. They have evaluated the limitations of various approach and suggested the possible extensions for making better recommendations.

Prodan Andrei-Cristian [7] discussed general aspects of a recommender system and collaborative filtering in detail in paper "Implementation of a Recommender System Using Collaborative Filtering" without dealing with the major issues of recommender system i.e. sparsity and scalability.

Maltz and Ehrlich [4] describe active CF as new technique in which people who find interesting documents actively send pointers to those pointers to their colleagues. A "pointer" contains a hypertext link to the source document as well as contextual information to help the recipient determine the interest and relevance of the document prior to accessing it.

Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl [8] proposed Item based collaborative filtering that relies on Item-Item similarity in paper entitled as "Item-Based Collaborative Filtering Algorithms". Although the approach provides quality predictions and performs fairly well than user-based approach still it is quite static approach while the recommendation system deals with the dynamic information.

SongJie Gong [9] proposes an extension to traditional CF approach that considers user similarity as well as item similarity considering both rating and attribute information. The approach is implemented on Movielens and IMDB datasets and results evidenced that it can alleviate scarcity problem.

Zheng Wen [5] presented several CF algorithms (Item-based k nearest neighbor , sparse SVD algorithm etc) for RS and test the performance of each algorithm and their mixtures on part of the Netflix data.

Michael J. Pazzani and Daniel Billsus [10] in their paper entitled "Content Based Recommendation Systems" have discussed systems that recommend an item to a user based upon a description of the item and a profile of the user's interests. However, such systems work only when sufficient amount of content information is available i.e. they suffer from problem of lack of content information.

Michael J. Pazzani discusses an approach to combine recommendations from multiple sources viz., CF, CB and demographic information. CF and CB methods can also be combined under a single unifying model.

Hydra: A Hybrid Recommender System [17] discusses the combination of CF and CB approaches in the context of web-based recommendations. This hybrid approach is special in that rating data as well as content information are joined in a unified model, which leads to less parameters and more reasonable prediction results.

Dipankaj G Medhi and Juri Dakua described the recommendation system MovieReco that consider the trustworthiness of user. RS usually depends on information provided by users to gather its knowledge but if many users provide wrong information it will not be possible for RS to provide accurate recommendations. In [35], an approach to deal with these bad users is proposed which includes the concept of testing the knowledge of users. MovieReco provides an interface to user to enter the information like actor, director etc and then decide whether the user is passing the knowledge test or not. After that standardized rating is calculated and stored in database which is further used for providing robust and effective recommendations using algorithm explained in [35].

III. PROPOSED METHOD

The presented work is movie based recommender system that will accept the user input as the basic information to get the suggestions in the form of recommended movies so that the movie ranking will be identified. In this we have defined a three dimensional hybrid recommendation system based on the similarity ratio. The first two dimensions are defined using content based analysis applied on user and movies separately. The third dimension is defined for relationship analysis using collaborative similarity analysis.

The presented work is actually in the form of an architecture that will accept the user request and identify the user similarity ratio based on different parameters. In this work have taken 3 main parameters.

1. The first level is applied on user similarity analysis. It this a weighted content based analysis is performed. This analysis is performed on different fields like age, gender, occupation etc.

2. The second level is applied on movie similarity analysis. In this a weighted analysis is applied on server side for similar movie identification. Here weighted content based analysis is performed.

3. The third level is based on statistical measure here, the relationship analysis is performed using co-relation coefficient analysis.

4. Use the temporal vector to get effective results.

SYSTEM ARCHITECTURE

Here the system architecture of this movielens dataset is defined. Here figure 4.2 is showing the information analysis present in the dataset.

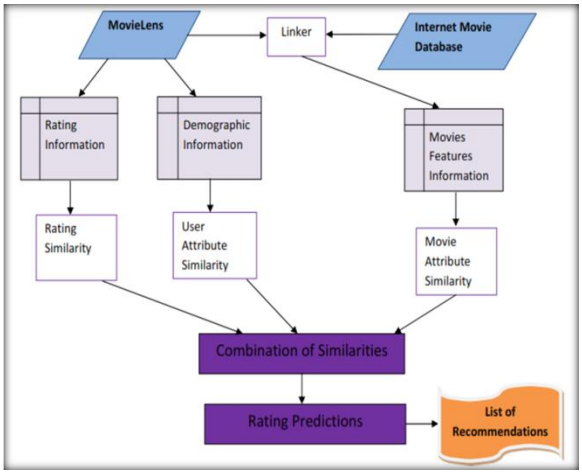

Figure : Overview of System Architecture

As shown in figure 4.2, The dataset is having the information about the user, movie and the rating assigned by the user to movies. These three are represented as three different dataset tables. The prediction over the dataset is performed under the recommendation analysis. The rating prediction is performed to identify the similarity between the users and to perform the prediction task.

SIMILARITY CALCULATION

The recommender system is about to perform the similarity based analysis. This analysis is required to performed at all three levels for all the available three tables. The first level is performed on user database to perform the user similarity analysis. In second stage, the analysis will be performed on movie database for movie similarity analysis. In third stage, the ranking similarity analysis is performed. To perform these three level analysis a hybrid model is presented. In which two stage mo

del is presented. At first stage, the content based analysis is performed and in second stage, the collaborative analysis is performed for similarity analysis.

Attribute Similarity

The effectiveness of the recommender system is based on the User attribute similarity. The attribute of this dataset includes the gender, age, occupation and location. The similarity analysis between two users is shown here

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

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 and

$Sim(u_{1f}, u_{2f})$ represents the similarity between values of *f* for u_1 *and* u_2.

On the movie dataset, the item similarity analysis is performed under different vectors such as release date, genre, country actor etc. Based on this analysis, the similarity between the movies will be identified.

Rating Similarity

Another similarity analysis is here required for the rating similarity analysis. The movie ranting is defined in rating dataset. This dataset contains mainly three attributes called userid, movieid and the rating. It signify the ranting assigned by the particular user for a particular movie. There are number of statistical approaches and formulas available to perform the estimation or the prediction. One of such similarity analysis formula is represented by Pearson Correlation formula. This formula is capable to identify the relation between two attribute values i.e. user and movie. The rating formula is given here under

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

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.

Item Rating similarity between item x and y is given by IB-PCC as

$$p_{xy} = \frac{\sum_{i \in U_x \cap U_y} (r_{ix} - \overline{r_x})(r_{jx} - \overline{r_y})}{\sqrt{\sum_{i \in U_x \cap U_y} (r_{ix} - \overline{r_x})^2 \sum_{i \in U_x \cap U_y} (r_{iy} - \overline{r_y})^2}} \quad (3)$$

Where $\overline{r_x}$ is the average rating given to item x and $\overline{r_y}$ is the average rating of item y.

IV. CONCLUSION

In this present work we have defined a recommender system to identify the ranking assigned to a movie. In this work, we have represented a hybrid model to perform the analysis. The hybrid model performed the analysis based on content based as well as the collaborative filtering. To perform this we have taken an authenticated dataset with three tables. One for user , second for site and third for ranking. Now while predicting the ranking of a new user.

At first the content based similarity match is performed to identify the similar users in the dataset. In the second stage, the rank provide by same kind of users is analyzed under the collaborative filtering and at the final stage the collective decision is taken regarding the site rank. Another dimension included in this work is the temporal factor. It means instead of analyzing the rank of all similar users a time based range is setup in this work. The presented

work is analyzed under the error rate. The obtained results show that the presented work is effective enough to provide correct results.

V. FUTURE WORK

The present work can be improved under different dimensions.

The main work can be done in future to reduce the error rate. In this work we have used 3-4 parameters while performing the content based match. More attributes can be considered to obtain more accurate results.

Another work can be done in same area with different datasets.

REFERENCES

[1] 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)

[2] Juan A. Recio-García," Prototyping Recommender Systems in jCOLIBRI", RecSys'08, October 23–25, 2008, Lausanne, Switzerland. ACM 978-1-60558-093-7/08/10 (pp243-250)

[3] Paul Resnick," The Influence Limiter: Provably ManipulationResistant Recommender Systems", RecSys'07, October 19–20, 2007, Minneapolis, Minnesota, USA. ACM 9781595937308/07/0010 (pp 25-32)

[4] Linas Baltrunas," Exploiting Contextual Information in Recommender Systems", RecSys'08, October 23–25, 2008, Lausanne, Switzerland. ACM 978-1-60558-093-7/08/10 (pp 295-298)

[5] Erich Christian Teppan," Implications of Psychological Phenomenons for Recommender Systems", RecSys'08, October 23–25, 2008, Lausanne, Switzerland. ACM 978-1-60558-093-7/08/10(pp323-326)

[6] Michael P. O'Mahony," A Recommender System for On-line Course Enrolment: An Initial Study", RecSys'07, October 19–20, 2007, Minneapolis, Minnesota, USA. ACM 978-1-59593-730-8/07/0010 (pp 133-136)

[7] Olga C. Santos," A Recommender System to Provide Adaptive and Inclusive Standard-based Support Along the eLearning Life Cycle", RecSys'08, October 23–25, 2008, Lausanne, Switzerland. ACM 978-1-60558-093-7/08/10 (pp 319-322)

[8] Jiyong Zhang," A Recursive Prediction Algorithm for Collaborative Filtering Recommender Systems", RecSys'07, October 19–20, 2007, Minneapolis, Minnesota, USA. ACM 978-1-59593-730-8/07/0010 (pp 57-64)

[9] Pearl Pu," A User-Centric Evaluation Framework for Recommender Systems", RecSys'11, October 23–27, 2011, Chicago, Illinois, USA. ACM 978-1-4503-0683-6/11/10 (pp 157-164)

[10] Toshihiro Kamishima," Personalized Pricing Recommender System-Multi-Stage Epsilon-Greedy Approach", HetRec '11, October 27, 2011, Chicago, IL, USA ACM 978-1-4503-1027-7/11/10

[11] Jingjing Zhang," Anchoring Effects of Recommender Systems", RecSys'11, October 23–27, 2011, Chicago, Illinois, USA. ACM 978-1-4503-0683-6/11/10 (pp 375-378)

[12] Bart P. Knijnenburg," A Pragmatic Procedure to Support the User-Centric Evaluation of Recommender Systems", RecSys'11, October 23–27, 2011, Chicago, Illinois, USA. ACM 978-1-4503-0683-6/11/10 (pp 321-324)

[13] Rong Hu," Acceptance Issues of Personality-based Recommender Systems", RecSys'09, October 23–25, 2009, New York, New York, USA. ACM 978-1-60558-435-5/09/10 (pp 221-224)

[14] Rong Hu," Design and User Issues in Personality-based Recommender Systems", RecSys'10, September 26–30, 2010, Barcelona, Spain. ACM 978-1-60558-906-0/10/09 (pp 357-360)

[15] Punam Bedi," Aspect-Oriented Mobility-Aware Recommender System", 2011 World Congress on Information and Communication Technologies 978-1-4673-0125-1@ 2011 IEEE (pp 191-196)

[16] Pooja Vashisth," Interest-Based Personalized Recommender System", 2011 World Congress on Information and Communication Technologies 978-1-4673-0125-1@ 2011 IEEE (pp 245-250)

[17] R. Arora, P. Pahwa, S. Bansal, "Alliance Rules of Data WarehouseCleansing", IEEE , International Conference on Signal Processing Systems, Singapore, May 2009, Page(s): 743 –747

[18] S. Chaudhuri, K. Ganjam, V. Ganti, "Data Cleaning in Microsoft SQL Server 2005", In Proceedings of the ACM SIGMOD Conference,Baltimore, MD, 2005

[19] S. Reddy, A. Lavanya, V. Khanna, L.S.S. Reddy, "Research Issues on Data Warehouse Maintenance", IEEE, ICACC '09. International Conference Advanced Computer Control, Singapore, Jan 2009, Page(s):623 – 627

[20] S.E. Madnick, Y.W. Lee, R.Y. Wang, and H. Zhu, "Overview and framework for data and information quality research", ACM, Journal of Data and Information Quality, Vol. 1, No. 1, Article 2, June 2009.

[21] Sang-goo Lee, Seoul Nat, Univ Seoul, "Challenges and Opportunities in Information Quality", E-Commerce Technology and the 4th IEEE International Conference on Enterprise Computing, E-Commerce, and E-Services, 2007, Tokyo, Jul 2007, Page(s): 481 – 481

[22] Tom Johnsten and Robert B. Sweeney, "A Methodology for Hiding Knowledge in XML Document Collections", Proceedings of the 27th Annual International Computer Software and Applications Conference (COMPSAC'03) 0730-3157/03 © 2003 IEEE

[23] Xingzhi Sun, "A Border-Based Approach for Hiding Sensitive Frequent Itemsets", Proceedings of the Fifth IEEE International Conference on Data Mining (ICDM'05) 1550-4786/05 © 2005

[24] Yu et al. Yu Huang, Xiao-yi Zhang, Zhen Yuan, Guo-quan Jiang, "A universal data cleaning framework based on user model", IEEE, ISECS International Computing, Communication, Control, and Management, Sanya, China, Aug 2009, Page(s): 200 – 202

[25] Timothy E. Ohanekwu," A Token-Based Data Cleaning Technique for Data Warehouse Systems".

[26] Dr. Payal Pahwa," Domain Dependent and Independent Data Cleansing Techniques", IJCST ISSN : 2229 - 4333

[27] Yingping Huang,"Infrastructure, Data Cleansing And Mining For Support Of Scientific Simulations".

[28] Payal Pahwa, "Uclean: A Requirement Based Objectoriented Etl Framework", International Journal of Computer Science & Engineering Survey (IJCSE).

