

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2015

User Genre Movie Recommendation System Using NB Tree

A.Saranya¹, A.Hussain²

Assistant Professor, Department of Information Technology, Adhiparasakthi Engineering College, Melmaruvathur,
India¹

Assistant Professor, Department of Information Technology, Adhiparasakthi Engineering College, Melmaruvathur,
India¹

ABSTRACT: The popularity among social network helps to share their views on anything by rating, generating blogs and analysis of user behavior (review). The new type of recommendation system in movie will provide best opportunity for users to solve cold start problems. This system is based on probabilistic matrix factorization. Four major factors are used for construction of this system datasets, prediction of rating, cosine similarity and NB tree. Moreover the ratings are made by experts which will be determined by number of accurate rating that they are made on movies. The accurate results are classifying by using cosine similarity; this system combines both user interest and social circle. Cosine similarity is mainly used to judge the similarity measure between the two users; interest is compared with high rate of accuracy this helps to find the interest of user more accurately. The NB tree is used to generate the user link formation. The user curiosity was classified into various level of ranges based on the rating of movie given by review experts. Experimental results show the proposed approach outperforms the existing Recommendation System approaches.

KEYWORDS: Recommendation System, Matrix Factorization, Movie Rating, Cosine Similarity, NB Tree, Prediction of Rating.

I. INTRODUCTION

Recommender systems typically produce a list of recommendations in one of two ways-through collaborative or content-based filtering. Collaborative filtering approaches building a model from a user's past behavior (items previously purchased or selected and/or numerical ratings given to those items) as well as similar decisions made by other users; then use that model to predict items (or ratings for items) that the user may have an interest in Content-based filtering approaches utilize a series of discrete characteristics of an item in order to recommend additional items with similar properties. These approaches are often mingled in Hybrid Recommender Systems. Every type of system has its possess strengths and weaknesses Recommender systems are a useful alternative to search algorithms since they help users discover items they might not have establish by themselves. Interestingly sufficient, recommender systems are frequently implemented using search engines indexing non-traditional data.

The recommender system compares the collected data to similar and dissimilar data collected from others and calculates a list of recommended items for the user. These systems often need a huge amount of existing data on a user in order to make accurate recommendations. Scalability: In many of the environments in which these systems make recommendations, there are millions of users and products. Thus, a large amount of computation power is often necessary to calculate recommendations. The number of items sold on main e-commerce sites is extremely huge. The most vigorous users will only have rated a small subset of the overall database. Thus, even the most popular items have very few ratings. Recommender systems usually have to deal with privacy concerns because users have to reveal susceptible information. Constructing user profiles using collaborative filtering can be problematic from a privacy point of view.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2015

The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for additional use. The actual data mining task is the automatic or semi-automatic analysis of large quantities of data to extract previously unknown interesting patterns such as groups of data records (cluster analysis), abnormal records (anomaly detection) and dependencies (association rule mining). This usually involves database techniques such as spatial indices. These patterns can then be seen as a kind of summary of the input data, and may be used in more analysis.

II. RELATED WORK

Most of people consider the friend’s opinions when they purchase some products or plan to watch the movies. As public network services have been recently popularized, many users attain and exchange their opinions on social networks. This information is consistent because they have secure relationships and trust each other. We propose a recommendation system based on advanced user modeling using community relationship of users. For the user modeling, both straight and indirect relations are considered and the relation weight between users is calculated by using six degrees of Kevin Bacon. Social filtering method can achieve better performance than a traditional user-based collaborative filtering method.

This research is an expansion of traditional collaborative filtering system and considers social relationship between users under the premise that if users have a closer relationship with other users, the preferred items are similar. Users are normally influenced by opinions of reliable friends rather than reviews of corresponding websites. Six degrees of separation is a theory that everyone is on average six steps away from any other person. An advanced user modeling using social relation weight for a movie recommendation system. In this method, we allocate relation weight to social relationship of users using six degrees of Kevin Bacon by considering both direct and indirect relations between users.

In this system we generate questionnaire using fuzzy rule (sugeno)[7] which consist of set of questions that has to be answered by the user at the time of registration. Cosine similarity is used to discover the interest of users more accurately by comparing it with the choice of recommendation and our interest. The rating prediction is used to recommend the movie which is having high rating. This helps the user to watch the worthy movie. For user association formation Hybrid approach is used (content + collaborative) to construct NB tree. Rating is not only given by the user itself, it is also given by the review experts.

III. SYSTEM ARCHITECTURE

The system of movie recommendation system is shown in Fig.3.1. It maintains the database which consists of list of movie details and user details. The profile about the user is consisted and it has the details like age, gender, interest. This also involves the fuzzy rule that create questionnaires to know the clear view about the interest of user.

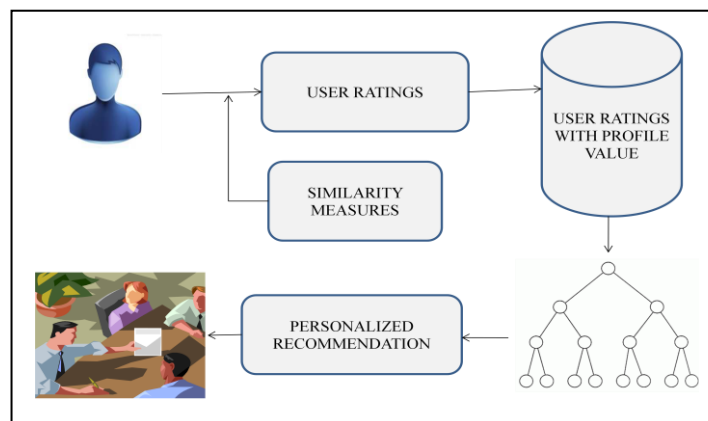


Fig.3.1 System Architecture

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2015

Cosine similarity is used for the measure of similarity between the interests of user with the movies which comes under similar category. The obtained result from this similarity measure is very much precise. The prediction rating is used to rate the favorite movie of user, the experts will also rate the movie. The high rated movie among all will be recommended based on the interest of user. Finally the user link formation is made by constructing a tree structure using naive Bayesian algorithm. This will helps to predict the rate of movies for recommendation and the set of movies are recommended to the user.

IV. MODULE DESCRIPTION

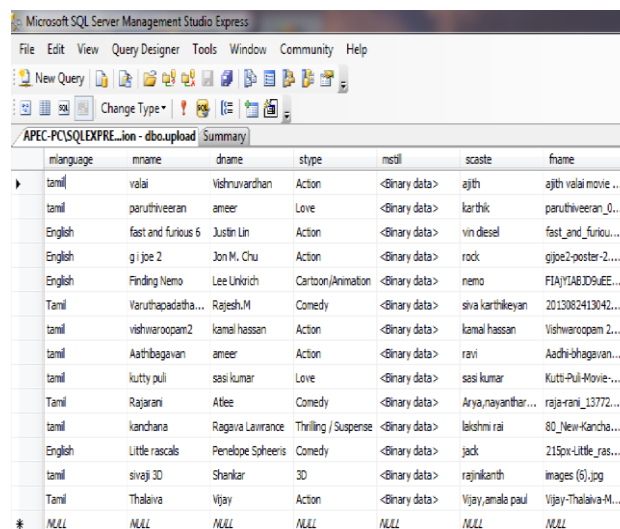
Modules are units of code written in access basic language. The developer can write and use module to automate and customize the database in very difficult ways. The modules are used to access and print information from the database. The report can situate records into many levels and calculate totals and average by checking values from many records at once. Also the modules are attractive and distinctive because the repository component available in SQL Server 3.5 is now called SQL Server management Studio Express Data Services. References to the element now use the term Data services.

The list of modules used in this project is:

- Data Collection
- Cosine Similarity
- Prediction Rating
- NB Tree Approach

A. Data Collection

Data collection has its own local directory service and user’s review with attribute of social networks. This has the collected information from thousands to lakhs of people and stores in a directory which is in the table form shown in Fig 4.1. Data collected from users based on their own interest and reviews from experts [1]. Artificial immune system with user and item based collaborative filtering to create an efficient and accurate recommendation system, researchers uses Movie Lens (ML) dataset as our testing target to evaluate the effectiveness of the algorithm [1, 12]. The number of users is registered in this system. The profile of the user consists of name, age, gender, and type of movie in which they are interested and the set of questions which follow fuzzy rule. The answer of users is again stored in the database in order to know the interest of user clearly.



mlanguage	mname	dname	stype	msll	scaste	fname
tamil	valai	Vishnuvardhan	Action	<Binary data>	ajith	ajith valai movie ...
tamil	paruthiveeran	ameer	Love	<Binary data>	karthik	paruthiveeran_0...
English	fast and furious 6	Justin Lin	Action	<Binary data>	vin diesel	fast_and_furiou...
English	g i joe 2	Jon M. Chu	Action	<Binary data>	rock	gjoel-poster-2....
English	Finding Nemo	Lee Unkrich	Cartoon/Animation	<Binary data>	nemo	FIAT\TAB\09\EE...
Tamil	Varuthapadtha...	Rajesh.M	Comedy	<Binary data>	siva karthikeyan	2013082413042...
tamil	vishwaroopam2	kamal hassan	Action	<Binary data>	kamal hassan	Vishwaroopam 2...
tamil	Aathabagavan	ameer	Action	<Binary data>	ravi	Aathi-bhagavan...
tamil	kutty puli	sasi kumar	Love	<Binary data>	sasi kumar	Kutti-Puli-Movie-...
Tamil	Rajarani	Atlee	Comedy	<Binary data>	Arya,nayanthar...	raja-rani_13772...
tamil	kanchana	Ragava Lavrance	Thrilling / Suspense	<Binary data>	lakshmi rai	80_New-Kancha...
English	Little rascals	Penelope Spheeris	Comedy	<Binary data>	jack	215pic-Little_ras...
tamil	sivaji 3D	Shankar	3D	<Binary data>	rajnikanth	images (5).jpg
Tamil	Thalaiva	Vijay	Action	<Binary data>	Vijay,amala paul	Vijay-Thalaiva-M...
* NULL	NULL	NULL	NULL	NULL	NULL	NULL

Figure 4.1 Movie Data set

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2015

B. Cosine Similarity

The similarity between two users' d1 and d2 can be computed using cosine similarity.

$$\text{Sim}(d1,d2)=\frac{v(d1).v(d2)}{|v(d1)||v(d2)|}$$

The cosine similarity provides the high rate of accuracy. It will compare the interest between two users using this similarity approach. The questionnaires are created by the administrator. This questionnaire helps to check the similarity between two users [10]. The exact movies based on user interest are recommended to the user.

C. Prediction Rating

Recommender system using a weighted arithmetic mean to calculate the predicted rating of an item "i" for some user "u". The rating gives by review experts. Generate the questionnaires using fuzzy rules (sugeno). The user can select the movie by rating. Predict the rating score that is user is likely to give to a movie that she or he has not seen before [2]. The high rated movies among all were recommended to the user based on the similarity and interest [12, 15].

S - Similarity
R - Rating
N - Overall number of item

$$P_{u,i} = \frac{\sum_{\text{all similar items}, N}^N (S_i, N * R_{u, N})}{\sum_{\text{all similar items}, N}^N (|S_i, N|)}$$

D. NB Tree Approach

NB (Naive Bayesian) Tree approach is like a tree structure that can generate the user link formation based on the given rating shown in Table 1. It can help to separate the movie based on the user's interest, age, gender, language. This is layered approach consist of primary and secondary levels in each level the users are separated based on their interest [4].

Table.1 Internal Structure of NB Tree

Movie Language	Movie Name	Director Name	Story type
Tamil	Paruthiveeran	Ameer	Love
Tamil	Kuttipuli	Sasikumar	Love
Tamil	Vishvaroopam2	KamalHassan	Action
Tamil	Sivaji3D	Shankar	3D
Tamil	Kanchana	Raghava Lawrence	Thrilling & suspense

V. EXPERIMENTAL AND RESULT ANALYSIS

The purpose of testing is to discover errors rebuild the functions without affecting the system behavior. Testing is the process of trying to notice every conceivable fault or weakness in a work product. It provides a system to ensure the functionality of components, sub assemblies, assemblies and/or a finished product it is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an intolerable manner. There are various types of investigation. Each investigation type addresses a precise testing requirement.

The design of the model was tested shown in Fig 5.1 there are many structure alignment problems occurred while executing it. The structure of the tree was not in a accurate format that we are predictable. The pixel sizes of the movie posters are dissimilar from one another. So this problem was conquering by taking the posters of accurate sized pixels which range in the certain limit value. The coding for all the design form are made and debugged before the execution. The local server port address was added into the program to execute the recommendation system in the server. The database link for the movie details and user details are also linked with the coding part. The alert message in a small dialogue box was never being delayed for the user activity. When the user enters wrong details on the specified textbox means the system will respond quickly with the alert for the error caused. We encountered the problem on delay for the

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2015

occurrence of this alert message. Field testing is performed manually and functional tests are written in detail. All the test cases mentioned above passed successfully. No defects encountered.

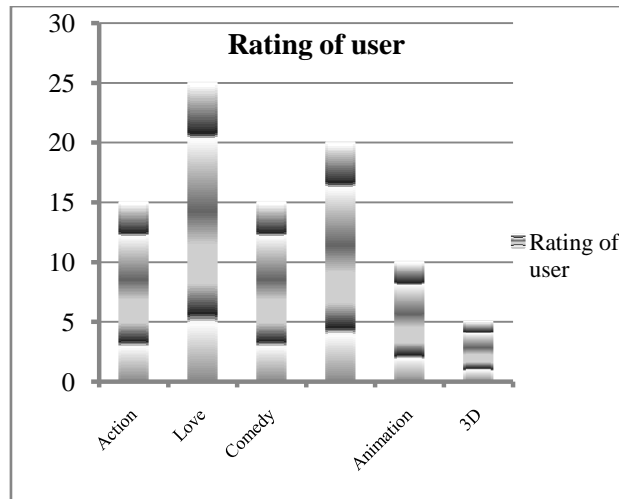


Figure 5.1 Rating of User

VI. CONCLUSION

This system helps the user to recommend movies based on their interest with number of choices. The satisfaction of user is much important that will be attained by using this recommendation system. The similarity measures between the personal profile of users and the movie profile is accurate by using the cosine similarity approach. The fuzzy rule create set of questionnaires to user at the time of registration this helps to understand the interest of the user more accurately.

The prediction rating method is used in this system which is to recommend the user with the top rated movies. The naïve Bayesian algorithm used to construct the naïve Bayesian tree which has number of layers. The user link formation is made based on this approach. The final stage of this system will recommend the high rated movie to the user based on their interest. The recommendation system on movies made by the personal interest of user. This provides complete satisfaction while watching the movie. If it is implemented as a mobile application means it will facilitate well for each and every people and also decreases cold start problem. So we planned to implement this concept as a mobile application that also supports for big data.

REFERENCES

- [1] Bing Liu , “Recommender system” cs583 (book) , 2007.
- [2] Bo Huan, G Qing-Pu Zhang, Yun-Quan Hu, “Research on Credit Risk Management of the State-owned Commercial Bank”, Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, Guangzhou, 18-21 August 2005.
- [3] Georgios Pitsilis , Svein J. Knapkog , “Social Trust as a solution to address sparsity-inherent problems of Recommender systems” New York, NY, USA. *IEEE / WIC / ACM international conference* ,October 25, 2009.
- [4] Helge Langseth, Anders Kofod-Petersen, “Design of a Bayesian Recommender System for Tourists Presenting a Solution to the Cold-Start User Problem” (book) – June 2010.
- [5] Inay Ha , Kyeong- Jin Oh , Geun-Sik Jo “Social Filtering using social relationship for movie recommendation”, 4th International Conference, Proceedings, ICCCI 2012, Ho Chi Minh City, Vietnam, November 28-30, 2012.
- [6] Luca Canini, Sergio Benini, and Riccardo Leonardi, “Affective Recommendation of Movies Based on Selected Connotative Features” *IEEE Transactions on Circuits and Systems for Video Technology* , Volume 23 Issue no 4, April 2013.
- [7] Luis M.de Campos , Juan M. Fernandez-Luna , “A collaborative recommender system based on probabilistic inference from fuzzy observations”, published in journal fuzzy sets and systems, volume 159, issue no. 12 , June -2008.
- [8] Manish Singh fargo, “Social network based recommendation system”, IEEE –October 2013.
- [9] Robert Bell and Chris Volinsky, “Matrix factorization for recommender system” AT&T lab, IEEE , published in computer Volume:42 , Issue:8, august 2009.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2015

- [10] Seth Sorensen, "Accuracy of Similarity Measures in Recommender Systems" published in IEEE Intelligent Systems22(3):48–55, IEEE – October 2011.
- [11] Sonal Patil , Ankush Mahajan , "Design and Architecture for Web Graph Mining Base Recommender System for Query, Image and Social Network using Query Suggestion Algorithm and Heat Diffusion Method", IOSR Journal, research paper, IEEE – January 2014.
- [12] Sudip Bhattacharjee, "Accurate recommendations on online movie rating" Submitted to "The 2010 Winter Conference on Business Intelligence", IEEE- November 2009.
- [13] P.N. Tan, M. Steinbach , Addison-Wesley & V. Kumar, "Introduction to Data Mining (book)",2005.
- [14] M. Tkalcic, U. Burnik, and A. Kosir, "Using affective parameters in a content-based recommender system for images",published in IEEE Journal User Modeling and User-Adapted Interaction (Volume 20 Issue 4) , October 2010.
- [15] Xueming Qian, He Feng, Guoshuai Zhao, and Tao Mei, "Personalized Recommendation Combining User Interest and Social Circle" , IEEE Transactions Knowledge and Data Engineering (volume 26 Issue 7) ,July 2014.