IJESRT
INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

DEVELOPMENT OF INTEREST BASED COLLABORATIVE FILTERING RECOMMENDATION ENGINE

Mr. Suraj Ashok Shinde *, Mr. Vipul Vinayak Bag
* Department of Computer Engineering, N.K.Orchid College of Engineering, Maharashtra, India

ABSTRACT

Recommender systems are used on the web for recommending products to users. Most of the electronic commerce sites have such systems. Collaborative filtering is an important and popular technique for recommender system. In this paper, an expert system for movie recommendation is presented with new approach. This system is implemented using co clustering method. Category of the movie and ratings given by users are used to give recommendations. Users are interested in grouping items into categories and for each category; there can be corresponding user group who like items in that category. Finding interest of user in particular item group and grouping users of similar interest is distinguishable feature, which differentiate our approach from the previous works.

KEYWORDS: Recommender system, Collaborative filtering, User group, Item group.

INTRODUCTION

With the development of the internet and its associated information explosion, users are faced with the problem of too much choice. Right from looking for a movie to looking for good investment options, there is too much information available. Thus, to guide users, companies have deployed recommendation systems. These recommendation systems provide users the useful data, employing some information filtering techniques. There are two main methods for information filtering, one is content based and other is collaborative filtering. Collaborative filtering is one of the most effective technique, due to its simplicity [1] There are two methods in collaborative filtering, user based and item based [3, 4]. The basic idea of user based collaborative filtering is to recommend items to users, based on opinions of other likeminded users, while item based collaborative filtering provides recommendation to user based on other items with high co-relation.

Recommendation systems are special types of expert systems. They combine the knowledge of the expert in a given domain, with the user's preferences, to filter the available information and provide the user with the most suitable information. Large scale recommendation system faces the problem when large amount of data is present. When the available data is small available traditional algorithms works well but when data sets increases, the traditional algorithms may face difficulty. Ratings given by user to item only show user's preference for that item at low level. In some scenarios measuring user's similarity based on such low level representation can give inaccurate results. Assume user X has rated three action movies with the highest ratings while user Y has rated other three action movies with his highest ratings. If traditional collaborative filtering methods are used to find out similarity between user X and user Y they will not be similar , as there is no co rated item between user X and User Y. However, this type of result is not true. As both the users are not having co rated items but both of them are interested in action movies. Thus they should be considered as similar users and can be grouped together.

In this paper, idea of calculating user interest in particular item group is used and an interest-based collaborative filtering recommendation approach is proposed. The mechanism of interest-based collaborative filtering recommendation is as follows: First, all items are clustered into several item groups. For example, we can cluster all movies into "comedy movies," "drama movies," and so on. Second, a user group corresponding to each item group is formed, with all users having different interest degrees in each of the user groups. Third, we build a user-interest matrix and measure users' similarities based on users' interest degrees in all user groups so as to select a set of "neighbors" of each user. Then, unknown rating of a user on an item is predicted based on the ratings of the "neighbors" of user on the item. A distinct feature of the interest-based collaborative filtering recommendation is that it selects the "neighbors" of users by measuring users' similarity based on user interest degrees in user groups, which differentiates it from previous methods. To the best of our knowledge, there has been no prior work on using interest degree of user with collaborative filtering recommendation. This system provides a new perspective to investigate collaborative filtering recommendations.

RELATED WORK

A large number of recommendation systems have been developed and are in use. These systems use a variety of methods such as content based, collaborative filtering, hybrid, etc. Tapestry was the first commercial recommender system, which was designed to recommend documents to users. When it comes to collaborative filtering most important step is to find set of similar users. Currently, almost all collaborative filtering methods measure users' similarity based on co-rated items of users. One such system was developed by Webster, Harris and Her locker (2004). The system collects ratings from users and these ratings are used to make recommendations. Even though these recommendation methods are widely used, a number of deficiencies have been identified. Recommender Systems continue to exist as an active area of research. Applications have been pursued in diverse domains ranging from recommending web pages to music, books, movies and other consumer products.

An item belongs to the group if it is more similar to the prototype (property) of the group. L.W. Barsalou [8] measures two factors which affect the items interest in a group. First factor is central tendency, which is the degree of an items' "family resemblance." The more an item is similar to other members of the same concept, the more it belongs to that group. Second factor is frequency of instantiation of a cluster of similar items in a group. It is an estimate of how often people experience, or considers, objects in the cluster as members of a particular group. Items of a cluster with higher frequency of instantiation in a group are more familiar to user, and thus they belong to the group. There are some works on measuring item interest in computer science. M. Rifqi [9] proposes a method to calculate object interest in large databases. In his work, the interest of an object for a category depends on its resemblance to other members of the category, as well as its dissimilarity to members of other categories.
Problem in existing systems is that it is hard to find out correlations between users and items. It happens when the available data are not sufficient for recognizing similar users or items [1]. It is an important issue that restricts the quality of collaborative filtering recommendations. There may be difference between ratings given by the user and predictions given by the system. These incorrect predictions may reduce the faith of users on recommender system. In existing systems ratings given by the user are used to provide the recommendation [3]. Actually people may like to group items into categories and for each category there is corresponding group of users who likes items in that category. Taking this concept, co clustering approach can be used to group items and its corresponding users. The relation between user groups and item groups are shown in Fig. 1. Users are having different interest degrees in different groups, darker the user is shown in the group more it belongs to that group. For example, U_1 and U_p have more interest in user group N than user group 1, while U_2 and U_q have more interest in user group 1.

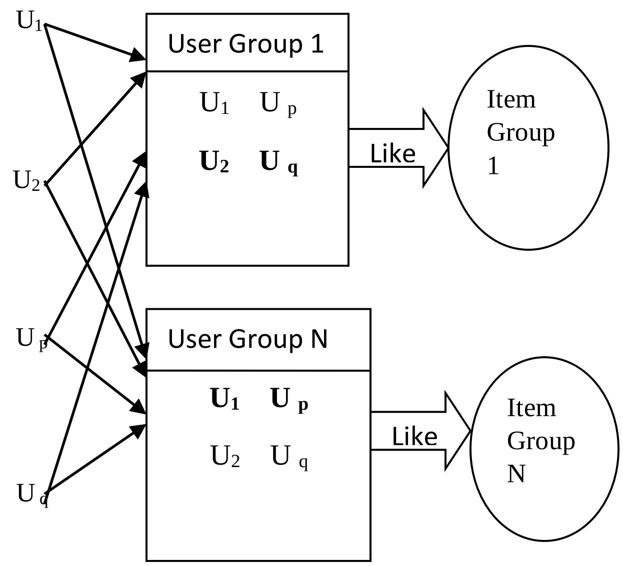

Fig. 1The Relation between User Groups and Item Groups

PROPOSED MODELLING

Interest Based Collaborative Filtering

In interest based collaborative filtering neighbors of user can be found based on user's interest in user groups. If we consider a collaborative filtering recommender system having item set and user set, items can be clustered into various item groups. For example, movies can be clustered according to their genre as comedy, action, drama and so on. Users having similar interest on an item group can form their group as user group.

An item group denoted by M_i is a set of items, as following:

$$M_i = \{I_1^{x_{i,1}}, I_2^{x_{i,2}} \dots In^{x_{i,n}}\}$$

Where n is the number of items in Mi, I_l is an item, and $x_{i,l}$ is the membership of item I_l in Mi.

A user group P_i corresponding to an item group Mi shows "users who like the items in Mi."Users may have different interest degrees in different Pi.

$P_i = \{N_1^{yi,1}, N_2^{yi,2} \ldots N_m^{yi,m}\}$

Where m is the number of users in the group Pi, N x is a user,
and y i , x is the interest degree of user N x in user group Pi.

For the reason that users have different interest degrees in different user groups, a user is represented by a user interest vector defined below:

$N_j = (y_{1,j}, y_{2,j}, \ldots, y_{n,j})$

Where n is number of user groups and $y_{i,j}$ is the interest degree of user N_j in user group N_i.
Thus, for all users, a user interest matrix can be obtained. A user interest matrix, denoted by Q_T , is a matrix with i^{th} row being user interest vector of user N_i:

$$Q_T = \begin{pmatrix} \overrightarrow{N1} \\ \ldots \\ \overrightarrow{Nm} \end{pmatrix} = \begin{pmatrix} y_{1,1}, y_{2,1}, \ldots, y_{n,1} \\ \ldots \\ y_{1,m}, y_{2,m}, \ldots, y_{n,m} \end{pmatrix}$$

Where m stands for number of users, n for number of user groups and $\overrightarrow{N_i}$ is the user interest vector of user N_j.
Users have different interest degrees in different user groups. For each item group we can define its related user group. Given a set $M=\{M_1, M_2, \ldots, M_h\}$ of items and set $P=\{P_1, P_2, \ldots, P_m\}$of users, set $R=\{r_1, r_2, \ldots, r_n\}$of item groups can be formed. For each item group there can be corresponding user group. Then, a user interest vector Q_T can be built for each user, from which user-interest matrix cab be obtained. After obtaining users' similarity based on their interest degrees in user groups, a set of "neighbors" can be obtained for each user. Then, we can recommend item to an active user based on the genre of the movie and ratings by "neighbors" of that user. Selecting "neighbors" of users by measuring users' similarity based on their interest degrees is a well defined characteristic, which distinguish our approach from previous collaborative filtering approaches

	P_1	P_2	P_3	P_4	P_5	P_6
N_1	0.84	0.72	0.90	0.15	0.34	0.25
N_2	0.37	0.24	0.41	0.92	0.88	0.97
	…	…	…	…	…	…
N_k	0.81	0.76	0.87	0.18	0.31	0.28
	…	…	…	…	…	…
N_m	0.44	0.25	0.38	0.93	0.91	0.95

Fig. 2 An Example of User Interest Matrix

Fig. 2 shows an example of user interest matrix. User N_1 and N_k are similar users as both possess similar interest degree in all user groups P_1 to P_6. Similarly N_2 is similar to N_m. In traditional collaborative filtering methods, similarity depends on co rated items by the users. Fig. 3 shows example of user rating matrix in traditional collaborative filtering.

	I_1	I_2		I_k		I_m
N_1	4	?		3		5
N_2	?	?		4		4
	…	…	…	…	…	…
N_k	2	5		?		3
	…	…	…	…	…	…
N_m	4	5		2		?

Fig. 3 User Rating Matrix in Traditional Collaborative Filtering

Item Interest Measurement
As mentioned above, the interest of an item in a concept depends on the central tendency of the item for the prototype (property) of the concept. In other words, if an item is more similar to property of a group, it has higher interest degree in the group. Items can be represented by set of properties. For example, genre, actor, director, producers etc. For each item group, property can be extracted to represent the item group.

All other items having same property can be added to that group. The property vector can be represented as follow:

 ISSN: 2277-9655
(I2OR), Publication Impact Factor: 3.785

$\vec{P}_{K\,j} = (P_{K\,j,\,1}, P_{K\,j,\,2}, \ldots, P_{K\,j,\,m})$

Where m represents number of properties of item group K_j. Interest of item I_y in an item group K_j, denoted by $W_{j,y}$, depends on the similarity between item I_y and the property of K_j, i.e. ,

$$W_{j,y} = \text{Similarity}\left(\vec{P}_{K\,j}, \vec{P}_{Iy}\right) \quad (1)$$

Where $\vec{P}_{K\,j}$ is property vector and $\vec{P}_{Iy}$ is item property vector, and Similarity is the cosine similarity function. An item group is regarded as a fuzzy set and is represented by cluster of similar items thus the interest degree of item in an item group depends on the similarity value obtained from the equation (1). Interest degree of item in an item group is affected by the degrees of internal similarity and external dissimilarity. Similarity of the item and property of the item group represent the internal similarity. External dissimilarity is similarity of item and properties of other item groups.

User Interest Measurement

Existing recommender systems are having data sets, which contains little information related to users' interests and ratings given by users on items are used to describe users' interests. User group represents users who like the items in the corresponding item group. For this purpose category of the item and rating given by the user to item are considered to calculate user interest in particular item group.

Mapping between Users

Category of item and ratings given by user are used for co clustering approach. For this purpose Nearest Neighborhood method is used. This method contains parameters as (n, ms, data).n stands for neighborhood size i.e. the number of users in the data model. ms denotes minimal similarity required for neighbors and data represents generic Boolean preference data model which is used for creating a new generic data model from the given users and their preferences.

Map<String, Integer> map1 = user Genre (uid1) (1)

Map<String, Integer> map2 = user Genre (uid2) (2)

Mapping is done as per the genre of item .This mapping process is used to get all genre for single user. Same mapping is applied for other user.

Map<Integer, Integer> map3 = user Rating (uid1) (3)

Map<Integer, Integer> map4 = user Rating (uid2) (4)

Item id and ratings given by user to are used for co cluster. Results obtained from above equations are passed to the cosine similarity for calculating similarity between users.

Cosine Based Similarity

In interest based collaborative filtering, user is represented by user interest vector. Similarity between users X_i and Y_i is calculated by computing cosine of the angle between them.

$$\text{Similarity} = \cos(\theta) = \frac{\sum_{i=1}^{n} X_i \times Y_i}{\sqrt{\sum_{i=1}^{n}(X_i)^2} \times \sqrt{\sum_{i=1}^{n}(Y_i)^2}}$$

Table 1: Mapping of Users as Per Genre (Category) of Movies and Ratings

Input Parameters	**Value**
Genre(Category)	Action, Adventure, Animation, Comedy, Crime, Drama, Family, Fantasy, Horror, History, Musical, Romance, Sci-Fi, Sports, Thriller, War.
Rating	1 to 10

Recommendations

After obtaining similarity between users, generic Boolean preference based user recommender is used to get recommendations. This recommender uses values of mapping between user group and item group and neighborhood value obtained from similarity formula.

RESULTS AND DISCUSSIONS

Data Set

Any expert system depends mainly on an extensive dataset. To get reliable result it is important that we have good dataset. Most of movie recommendation systems depend on user given ratings rather than the properties of the movies. Our system is mainly based on the properties of the movie. Hence, it was important for us to have dataset that will have all information about properties of the movie. To evaluate our recommendation method, movie dataset is used. Properties of movies are extracted from the internet movie database (IMDB).

Evaluation Metrics

To measure statistical accuracy, Coverage metric is used. It measures the percentage of items for which recommender system can make predictions. As an example, if system can predict 7500 out of 10000 ratings on items to be predicted, the coverage is 0.75. This shows that, larger the coverage, more the method will be able to predict ratings on unrated items. Larger the coverage values, better the recommendation method.

Performance Comparison on coverage

For interest based collaborative filtering, nearest neighborhood based clustering method is used. Using this method item groups and corresponding user groups are formed. This system is compared with several classic baseline methods, which includes a user based collaborative filtering with Pearson Correlation Coefficient (UBCF), an item based collaborative filtering with Pearson Correlation Coefficient (IBCF) and Typicality based collaborative filtering method (Tyco).The Data set is divided into two parts, one is training set and other is test set. Recommendation predictions are obtained based on the training and test set, which are used to evaluate accuracy of interest based collaborative filtering (InBCF).To form the training and test sets, user-movie rating pairs are randomly chosen. One variable named train/Test ratio is used, to indicate percentage of data used as the training and test sets. As mentioned above, coverage gives percentage of items for which system can make predictions. As per Fig. 4, interest based collaborative filtering (InBCF) can obtain highest coverage with all train/test ratios. IBCF and UBCF achieve coverage around 0.4 with train/test ratio 0.1 and around 0.8 with train/test ration of 0.3.For InBCF, it can obtain stable coverage. This indicates that InBCF can predict more ratings on unrated items compared to IBCF and UBCF. Using InBCF good coverage values can be obtained even with low train/test ratio. If train/test ratio is small, it is difficult for traditional collaborative filtering methods to find out similar neighbors of user or item. Thus, in case of traditional collaborative filtering methods recommendation accuracy is low. However, for interest based collaborative filtering users are having different interest degrees in different groups. The neighbors of users are having different interest degree in different user groups and usually number of user groups is not large.

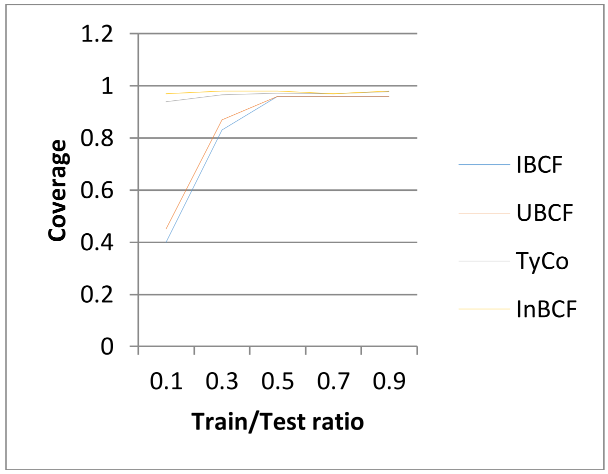

Fig. 4 Comparison on Coverage

CONCLUSION

In this paper, collaborative filtering recommendation is implemented with new approach. Co clustering method is used to measure users' similarity. A distinct feature of interest based collaborative filtering is that it selects similar users based on their interest degree in particular item group. This system can overcome several limitations of traditional collaborative filtering methods.

REFERENCES

[1] Yi Cai , Ho-fung Leung, Qing Li , Senior Member, IEEE, Huaqing Min, Jie Tang, and Juanzi Li "Typicality-based Collaborative Filtering Recommendation" IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 26 NO. 3, March 2014

[2] A. Umyarov and A. Tuzhilin, "Improving Collaborative Filtering Recommendations Using External Data," Proc. IEEE Eighth Int'l Conf. Data Mining (ICDM '08), pp. 618-627, 2008.

[3] Maddali Surendra Prasad Babu "An Implementation of the User-based Collaborative Filtering Algorithm", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 2 (3) , 2011, 1283-1286

[4] B. Sarwar, G. Karypis, J. Konstan, and J. Reidl, "Item-Based Collaborative Filtering Recommendation Algorithms," Proc. 10th Int'l Conf. World Wide Web (WWW '01), pp. 285-295, 2001.

[5] Z. Huang, H. Chen, and D. Zeng, "Applying Associative Retrieval Techniques to Alleviate the Sparsity Problem in Collaborative

Filtering," ACM Trans. Information Systems, vol. 22, no. 1, pp. 116- 142, 2004.
[6] G. Adomavicius and A. Tuzhilin, "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions," IEEE Trans. Knowledge and Data Eng., vol. 17, no. 6, pp. 734-749, June 2005.
[7] K.M. Galotti, Cognitive Psychology In and Out of the Laboratory, third ed. Wadsworth, 2004.
[8] L.W. Barsalou, Cognitive Psychology: An Overview for Cognitive Scientists. Lawrence Erlbaum Assoc., 1992
[9] Chong-Ben Huang, Song-Jie Gong, Employing rough set theory to alleviate the sparsity issue in recommendersystem, In: Proceeding of the Seventh InternationalConference on Machine Learning and Cybernetics(ICMLC2008), IEEE Press, 2008, pp.1610-1614.
[10] Sarwar B, Karypis G, Konstan J, Riedl J. Item-Based collaborative filtering recommendation algorithms. In: Proceedings of the 10th International World Wide Web Conference. 2001. 285-295.
[11] Manos Papagelis, Dimitris Plexousakis, Qualitative analysis of user-based and item-based prediction algorithms for recommendation agents, Engineering Application of Artificial Intelligence 18 (2005) 781-789.
[12] Hyung Jun Ahn, A new similarity measure for collaborative filtering to alleviate the new user coldstarting problem, Information Sciences 178 (2008) 37-51.
[13] SongJie Gong, The Collaborative Filtering Recommendation Based on Similar-Priority and Fuzzy Clustering, In: Proceeding of 2008 Workshop on Power Electronics and Intelligent Transportation System (PEITS2008), IEEE Computer Society Press, 2008, pp. 248-251.
[14] SongJie Gong, GuangHua Cheng, Mining User Interest Change for Improving Collaborative Filtering, In:Second International Symposium on Intelligent Information Technology Application(IITA2008), IEEE Computer Society Press, 2008, Volume3, pp.24-27.
[15] Breese J, Hecherman D, Kadie C. Empirical analysis of predictive algorithms for collaborative filtering. In: Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence (UAI'98). 1998. 43~52.
[16] M. Rifqi, "Constructing Prototypes from Large Databases," Proc.Int'l Conf. Information Processing and Management of Uncertainty (IPMU '96), pp. 301-306, 19

AUTHOR BIBLIOGRAPHY

Mr. Suraj Shinde

He received B.E degree in information technology from Pune University, Maharashtra, India and pursuing the M.E. degree in Computer Science and Engineering in Nagesh Karajagi Orchid College of Engg. & Technology, Solapur, India. He is doing his dissertation work under the guidance of Mr. Vipul V Bag, Head of Department CSE Nagesh Karajagi Orchid College of Engg. & Technology, Solapur, Maharashtra, India.

Mr. Vipul Bag

He is working as Associate Professor in Department of Computer Science and Engineering in NK Orchid College of Engineering and Technology, Solapur, Maharashtra, India. He has 16 years of teaching experience. He has co-authored over 20 International Journal Publications. He is pursuing PhD from SGGSIET, Nanded, Maharashtra, India. The current research interests are Recommendation systems, Data Mining and Machine Learning.