ISSN 2347 - 3983
International Journal of Emerging Trends in Engineering Research (IJETER), Vol. 3 No.6, Pages : 140- 146 (2015)
Special Issue of NCTET 2K15 - Held on June 13, 2015 in SV College of Engineering, Tirupati
http://warse.org/IJETER/static/pdf/Issue/NCTET2015sp29.pdf

TYPICALITY BASED - COLLABORATIVE FILTERING RECOMMENDATION USINGCLUSTERING

A.Ganesh [1],P.M Jhansi Rani[2]
[1]*Assistant Professor,Dept of CSE, SVCE, Tirupathi.*
[2]*Dept of Computer Science &Engineering, SVCE, Tirupathi.*
[1]achari.ganesh@gmail.com
[2]pasupakamjanu08@gmail.com,

Abstract

Collaborative filtering is a good mechanism used in recommender system, which is used to find the similar items in a group. The similar favour items can be identified by using the collaborative filtering based on items and the users. However there are some drawbacks in previous filtering techniques which leads to less accuracy, data sparsity and prediction errors. In the huge collection of data the recommendation can be accurately obtained by using clusters. Typicality is used to know the neighbours of the users in a cluster for better recommendations of items.

Keywords:Encryption,Authentication,privacy,multi-keyword,Efficiency.

I. INTRODUCTION

The recommender system actually used to find the similar items which are favour or suggested to use. We have plenty of applications to suggest items but as amount of data increases it becomes much hard to perform. Collaborative filtering makes easy way of recommendation using item based and user based approaches. The accuracy can be continued in recommendations of items when the clustering mechanism is followed. Ratings of a particular group of users and the item based grouping makes the recommendation more ease. Items in large data setare rated by predictions and we can differ it from the actual ratings by using Mean Absolute Error (MAE).

Typicality based involves in finding neighbours instead of co-rated items of users. Recommender system has content based, collaborative and hybrid types for typicality finding. For user and item based collaborative filtering the measurement of similarity items or users is primary step to do this we have vector space similarity, cosine based similarity, pearson correlation coefficient techniques.

Ratings from the user is considered and maintained as group from 1 to 5 low to high and recommended for other users. With respect to the ratings we use neighbours for recommending typicality items. Cluster helps to make recommendations easy way.

ISSN 2347 - 3983
International Journal of Emerging Trends in Engineering Research (IJETER), Vol. 3 No.6, Pages : 140- 146 (2015)
Special Issue of NCTET 2K15 - Held on June 13, 2015 in SV College of Engineering, Tirupati
http://warse.org/IJETER/static/pdf/Issue/NCTET2015sp29.pdf

II.BACKGROUND AND RELATED WORK

Prototype View and Typicality

In the prototype view of concepts, a concept is represented by a best prototype abstracted by the property list that consists of the salient properties of the objects that are classified into this concept. The salient properties defining the prototype include both necessary and unnecessary properties. It has been found that typicality of an instance can be determined by the number of its properties which it shares with the concept prototype. For example, the property "can-fly" will probably appear in the prototype of the concept "bird" because most birds can fly. So birds that can fly will be judged as more typical than thosethat cannot. A prototype of a concept is considered as the best example of the concept, and is abstracted to be a feature list. Although the prototype view can explain many different aspects of how concepts and properties are represented in human's mind, there are also situations in which it fails to give a thorough explanation. For example, there is virtually no prototype to represent the concept "animal." It cannot explain the co-occurring relations among properties of an instance, either.

There are some works on measuring object typicality in computer science. Rifqi proposes a method to calculate object typicality in large databases, which is later extended by Lesot et al. In their works, the typicality of an object for a category depends on its resemblance to other members of the category, as well as its dissimilarity to members ofother categories. Au Yeung and Leung [12] have formalized object typicality in a model of ontologies, in which the typicality of an object in a concept is the degree of similarity matching between the object property vector and the prototype vector of the concept. All these works focus on developing methods to calculate object typicality in concepts. There has been no work on integrating typicality in collaborative filtering recommendation.

III. EXISTING SYSTEM

In existing approach the user's preferences at low level is only captured which leads to inaccurate results. Difficulty to find correlations between users and items when very few ratings are given and it limits the quality of collaborativefiltering recommendations. User based and item based collaborative filtering is not accurate to pose on the available data. Item and user groups are not correlated which makes inaccurate data recommends for users.

	i_1	i_2	...	i_k	...	i_n
U_1	5	?	...	3	...	4
U_2	?	?	...	4	...	5
⋮	...	...	...	...	...	...
U_k	2	5	...	?	...	3
⋮	...	...	...	...	...	...
U_m	5	4	...	2	...	?

Fig1: user rating matrix in traditional Collaborative Filtering.

Disadvantages of existing system:

- It is difficult to find out correlations between users and items.

ISSN 2347 - 3983
International Journal of Emerging Trends in Engineering Research (IJETER), Vol. 3 No.6, Pages : 140- 146 (2015)
Special Issue of NCTET 2K15 - Held on June 13, 2015 in SV College of Engineering, Tirupati
http://warse.org/IJETER/static/pdf/Issue/NCTET2015sp29.pdf

- It occurs when the available data are insufficient for identifying similar users or items.

IV. PROPOSED SYSTEM

In this paper we have collaborative filtering using clustering. At first all items are grouped as several groups, next we form a user group corresponding to each item group, at last we build user typicality matrix and measure users similarities based on users. The neighbour's selection by measuring user's similarity based on user typicality in user groups can be done by using the collaborative filtering recommendation.

Proposed system reduces the number of big error predictions, improves accuracy of predictions and works with sparse training data sets.

V. TYPICALITY BASED COLLABORATIVEFILTERING

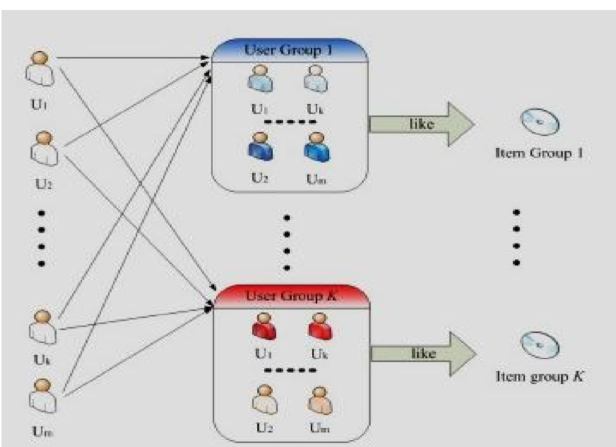

Fig2: The relations among users, user groups and item groups.

There are a set U of users, and a set O of items. Items can be clustered into several item groups and an item group is intuitively a set of similar items. For example, movies can be clustered into action movies, war movies, and so on. Each movie belongs to different movie groups to different degrees. The choice of clustering method is application domain dependent.

$K_i = \{O_1^{wi,\,1}$, O2wi, 2, Owe, n}

A user group gi is a fuzzy set of users

gi={U1vi,1,U2vi,2.....,Umvi, m}

	g_1	g_2	g_3	g_4	g_5	g_6
U_1	0. 87	0. 75	0. 92	0. 12	0. 32	0. 28
U_2	0. 34	0. 21	0. 38	0. 89	0. 85	0. 94
⋮	...	...	...	...	...	...
U_k	0. 81	0. 79	0. 89	0. 15	0. 29	0. 31
⋮	...	...	...	...	...	...
U_m	0. 41	0. 22	0. 35	0. 90	0. 88	0. 92

Fig3: User typicality in proposed system.

VI. RECOMMENDER SYSTEMS

There have been many works on recommender systemsand most of these works focus on developing new methods of recommending items to users .The objective of recommender systems is to assist users to

ISSN 2347 - 3983
International Journal of Emerging Trends in Engineering Research (IJETER), Vol. 3 No.6, Pages : 140- 146 (2015)
Special Issue of NCTET 2K15 - Held on June 13, 2015 in SV College of Engineering, Tirupati
http://warse.org/IJETER/static/pdf/Issue/NCTET2015sp29.pdf

find out items which they would be interested in. Items canbe of any type, such as movies, jokes, restaurants, books, news articles, and so on. Currently, recommendation methods are mainly classified into collaborative filtering (CF), content based (CB), and hybrid methods. For the reason that we are focusing on proposing a new CF method, we will introduce the related works about CF methods in more details.

Content-Based Recommender Systems

The descriptions of items are analysed to identify interesting items for users in CB recommender systems. Based on the items a user has rated, a CB recommender learns a profile of user's interests or preferences. According to a user's interest profile, the items which are similar to the ones that the user has preferred or rated highly in the past Will be recommended to the user. For CB recommender systems, it is important to learn users' profiles. Various learning approaches have been applied to construct profiles of users.

Example: LIBRA SYSTEM

Collaborative Filtering

For the reason that CF methods do not require well-structured .There are two kinds of CF methods, namely User-based CF approach and item-based CF approach. user-based CF approach first finds out a set of nearest "neighbors" (similar users) for each user, who share similar favourites or interests. Then, the rating of a user on an unrated item is predicted based on the ratings given by the user's "neighbors" on the item.

Hybrid Recommender Systems

Several recommender systems and use a hybrid approach by combining collaborative and content basedmethods, so as to help avoid some limitations of content-based and collaborative systems. A naive hybrid approach is to implement collaborative and CB methods separately, and then combines their predictions by a combining function, such as a linear combination of ratings or a voting scheme or other metrics. Melville et al. use a CB method to augment the rating matrix and then use a CF method for recommendation.
Some hybrid recommender systems combine item-based CF and user-based CF. For example; Ma et al. Proposean effective missing data prediction (EMDP) by combining item-based CF and user-based CF.

Experiments show that typicality-based CF method has the following several advantages:

- It generally improves the accuracy of predictions when compared with previous recommendationmethods.
- It works well even with sparse training data sets,especially in data sets with sparse ratings for eachitem.
- It can reduce the number of big-error predictions.
- It is more efficient than the compared methods.

VII. METRICS

Statistical accuracy can be measured by using Mean Absolute Error (MAE) metric its a measure of deviation of

ISSN 2347 - 3983
International Journal of Emerging Trends in Engineering Research (IJETER), Vol. 3 No.6, Pages : 140- 146 (2015)
Special Issue of NCTET 2K15 - Held on June 13, 2015 in SV College of Engineering, Tirupati
http://warse.org/IJETER/static/pdf/Issue/NCTET2015sp29.pdf

recommendations from real user rated ratings which is commonly used and easy to interpret. It's computed by averaging all the sums of the absolute errors of the *n* corresponding rating prediction pairs, and defined as follows

$$MAE = \frac{\sum_{i=1}^{n} [\mathrm{fi} - \mathrm{hi}]}{\mathrm{n}}$$

*n*is the number of rating-prediction pairs, f_i is an actual user-specified rating on an item, and h_i is the prediction for a user on an item given by the recommender system.

Lower MAE value indicates that the recommendation method can predict users rating more accurately means the smaller the better accuracy.

	γ	χ=0.1	χ=0.3	χ=0.5	χ=0.7	χ=0.
n=5	0.8	0.8106	0.771	0.7478	0.7436	0.736
n=10	0.7	0.8115	0.7771	0.7546	0.7451	0.739
n=15	0.6	0.8117	0.7774	0.7563	0.7502	0.739
n=20	0.6	0.8125	0.7757	0.7568	0.7481	0.73
n=25	0.5	0.8136	0.777	0.7576	0.7515	0.73
n=30	0.5	0.8129	0.7726	0.7536	0.7438	0.734
AVG		0.8121	0.7751	0.7544	0.747	0.737

Table: sensitivity on *n*on MAE with different test ratios

The difference between proposed and previous user-based collaborative filtering is that it finds a user's neighbours based on their typicality degrees in all user group, instead of based on users' ratings on items in previous methods. For item-clustering-based CF, they are based on clustering items, while it is based on users' typicality. That is, item-clustering-based CF is item-based recommendation while proposed system is user-based recommendation. Current hybrid methods are based on combining both collaborative filtering and content-based methods, for example, using some aggregation to aggregate the recommendation results of CF method and content-based method, while present CF is a neighbour-based recommendation. Latent factor methods use latent factors or concepts to find neighbours instead of pure rating. The idea behind such models is to characterize both items and users by vectors of factors inferred from item rating patterns.

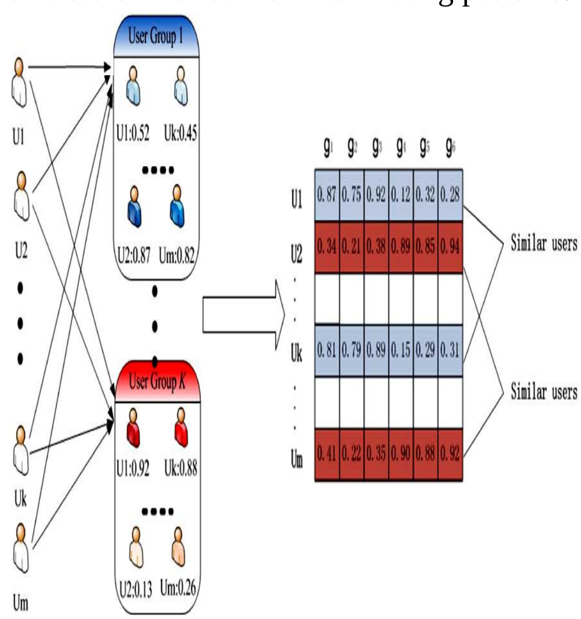

Fig4: Mechanism of discovering similar users in collaborative filtering

VIII. CONCLUSION AND FUTURE WORK

In this paper, we investigate the collaborative filtering recommendation from a new perspective and present a novel typicality-based collaborative filtering recommendation method. In this a user is represented by a user typicality vector that can indicate the user's preference on each kind of items. Its distinct feature is that it selects "neighbours" of users by measuring users' similarity based on their typicality degrees instead of co-rated items by users. Such a feature can overcome several limitations of traditional collaborative filtering methods. It is the first work that applies typicality for collaborative filtering. there are some pre-processing procedures, such as constructing user prototype by clustering and measuring user typicality in user groups. The cost of these pre-processing procedures depends on the particular clustering method used. In real life applications, these procedures can be processed offline. While users' prototypes are constructed, the remained recommendation process which is based on user typicality will be efficient. For large scale applications, we can also first conduct the above pre-processing offline, and then adopt some parallel computing methods (e.g., Map Reduce) to speed up the computing.

There are several possible future extensions to our work. In collaborative filtering technique, wedo not specify how to cluster resources so as to find out item groups and the corresponding user groups. One possible future work is to try different clustering methods and see how the recommendation results are affected. How to using parallel computing methods (e.g. Map Reduce) to handle the large scale applications is also one of the possible future works.

REFERENCES

[1] 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.
[2] 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.
[3] K.M. Galotti, Cognitive Psychology In and Out of the Laboratory, third ed. Wadsworth, 2004.
[4] G.L. Murphy, The Big Book of Concepts. MIT Press, 2002.
[5] L.W. Barsalou, Cognitive Psychology: An Overview for Cognitive Scientists. Lawrence Erlbaum Assoc., 1992.
[6] S. Schiffer and S. Steele, Cognition and Representation. Westview Press, 1988.
[7] D.L. Medin and E.E. Smith, "Concepts and Concept Formation," Ann. Rev. of Psychology, vol. 35, pp. 113-138, 1984.
[8] W. Vanpaemel, G. Storms, and B. Ons, "A Varying Abstraction Model for Categorization," Proc. Cognitive Science Conf. (CogSci'05), pp. 2277-2282, 2005.
[9] L.W. Barsalou, "Ideals, Central Tendency, and Frequency of Instantiation as

ISSN 2347 - 3983
International Journal of Emerging Trends in Engineering Research (IJETER), Vol. 3 No.6, Pages : 140- 146 (2015)
Special Issue of NCTET 2K15 - Held on June 13, 2015 in SV College of Engineering, Tirupati
http://warse.org/IJETER/static/pdf/Issue/NCTET2015sp29.pdf

Determinants of Graded Structure in Categories," J. Experimental Psychology: Learning, Memory, and Cognition, vol. 11, no. 4, pp. 629-654, Oct. 1985.
[10] M. Rifqi, "Constructing Prototypes from Large Databases," Proc. Int'l Conf. Information Processing and Management of Uncertainty (IPMU '96), pp. 301-306, 1996.

Mr.A.Ganesh M.Tech. He received Bachelor of Technology (CSE) degree in 2008 from JNTUH, Hyderabad and Master of Technology (Computer Networks) degree in 2012 from JNTUA, Ananthapur. His research interests are Computer Networks, Network Security, and Cloud Computing.

Mrs P M Jhansi Rani M.Tech. She received Bachelor of Technology (CSE) degree in 2013 from JNTUA, Anantapur and Master of Technology (Computer Science and Engineering) degree in 2015 from JNTUA, Ananthapur. Her research interests is Cloud Computing.