International Journal of Innovative Research in Advanced Engineering (IJIRAE) ISSN: 2349-2163
Issue 6, Volume 2 (June 2015) www.ijirae.com

# FUZZY NEIGHBOUR PREDICTION BASED ON OBJECT TYPICALITY

Ferozkhan AB[1*], Mohammed Yacoob B.A[2] , Francina Sophiya D [3]
[1,2,3] *Department of Master of Computer Applications*
[1,2,3] *C.Abdul Hakeem College of Engineering &Technology, Tamilnadu, India*

***Abstract - Collaborative based recommendation techniques are used in E-commerce nowadays to predict the ratings of user more accurately as possible. Measuring user similarity based on user ratings leads to inaccurate results. Now we borrow the idea of measuring similarity based on neighbors of users not based on ratings and cluster the items in to item groups and forming user groups corresponding to each item group then find the user preference based on neighbor list.***

***Keywords— Bigerror, Recommendation inaccuracy, sparsity, fuzzy group, coclustering***

## I. INTRODUCTION

Collaborative filtering based recommendations are popular methodology for recommender systems. Collaborative filtering are classified in to user based collaborative filtering and item based collaborative filtering. User based method measures similarity based on users likes and dislikes and item based method measures similarity between items, for eg user who bought item x also bought item y. Although these methods are widely used there are some inadequacies have been identified including

### Bigerror

People require recommender systems to predict users' preferences or ratings as accurately as possible. However, some predictions provided by current systems may be very different from the actual preferences or ratings given by users. These inaccurate predictions, called bigerror predictions, may reduce the trust of users on the recommender system. With the above-mentioned issues, it is clear that a good mechanism to find "neighbors" of users is very important. A better way to select "neighbors" of users or items for collaborative filtering can facilitate better handling of the challenges. Measuring users' similarity based on such a low-level representation of users i.e., similar items of users) can lead to inaccurate results in some cases. For example, suppose Bob has only rated five typical war movies with the highest ratings while Tom has rated other five typical war movies with his highest ratings. If we use traditional CF methods to measure the similarity between Bob and Tom, they will not be similar at all, for the reason that there are no corated items between Bob and Tom. However, such a result is intuitively not true. Even though Bob and Tom do not have any corated items, both of them are fans of war movies and they share very similar preference on war movies. Thus, we should consider them to be similar to a high degree.

### Sparsity

The data sparsity problem is the problem of having too few ratings, and hence, it is difficult to find out correlations between users and items. It occurs when the available data are insufficient for identifying similar users or items.It is a major issue that limits the quality of CF recommendations. Besides, the sparser the user rating data is, the more seriously traditional CF methods suffer from such a problem.

## II. FUZZY BASED APPROACH

In reality, people may like to group items into categories, and for each category there is a corresponding group of people who like items in that category. Cognitive psychologists find that objects (items) have different typicality degrees in categories in real life. For instance, people may consider that a sparrow is more typical than a penguin in the concept of "bird," and "Titanic" is a very typical romance movie, and so on. Similarly, different people may have different degrees of typicality in different user groups (i.e., sets of persons who like items of particular item groups). For instance, Raymond is a very typical member of the concept "users who like war movies" while not so typical in the concept "users who like romance movies." The typicality of users in different user groups can indicate the user's favor or preference on different kinds of items. The typicality degree of a user in a particular user group can reflect the user's preference at a higher abstraction level than the rated items by the user.

Thus, in this paper, we borrow the idea of object typicality from cognitive psychology and propose a typicality-based CF recommendation approach. The mechanism of typicality-based CF recommendation is as follows: First, we cluster all items into several item groups. For example, we can cluster all movies into "war movies," "romance movies," and so on.

Second, we form a user group corresponding to each item group (i.e., a set of users who like items of a particular item group), with all users having different typicality degrees in each of the user groups. Third, we build a user-typicality matrix and measure users' similarities based on users' typicality degrees in all user groups so as to select a set of "neighbors" of each user. Then, we predict the unknown rating of a user on an item based on the ratings of the "neighbors" of at user on the item.

A distinct feature of the typicality-based CF recommendation is that it selects the "neighbors" of users by measuring users' similarity based on user typicality degrees in user groups, which differentiates it from previous methods. To the best of our knowledge, there has been no prior work on using typicality with CF recommendation. It provides a new perspective to investigate CF recommendations. We conduct experiments to validate this method and compare it with previous methods.

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

- *It generally improves the accuracy of predictions when compared with previous recommendation methods.*
- *It works well even with sparse training data sets, especially in data sets with sparse ratings for each item.*
- *It can reduce the number of big-error predictions.*

It is more efficient than the compared methods

**Content-Based Recommender Systems**

The inspiration of this kind of recommendation methods comes from the fact that people have their subjective evaluations on some items in the past and will have the similar evaluations on other similar items in the future. The descriptions of items are analyzed 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. For example, Mooney and Roy [15] adopt text categorization methods in LIBRA system to recommend books.

**Collaborative Filtering**

CF recommendation methods predict the preferences of active users on items based on the preferences of other similar users or items. For the reason that CF methods do not require well-structured item descriptions, they are more often implemented than CB methods and many collaborative systems are developed in academia and industry. There are two kinds of CF methods, namely user-based CF approach and item-based CF approach .

**User to User**

The basic idea of user-based CF approach is to provide recommendation of an item for a user based on the opinions of other like-minded users on that item. The user-based CF approach first finds out a set of nearest "neighbors" (similar users) for each user, who share similar favorites 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.

**Item to Item**

The basic idea of item-based CF approach is to provide a user with the recommendation of an item based on the other items with high correlations. Unlike the user-based CF, the item-based CF approach first finds out a set of nearest "neighbors" (similar items) for each item. The item based CF recommender systems try to predict a user's rating on an item based on the ratings given by the user on the neighbors of the target item. For example, Sarwar et al.[17] discuss different techniques for measuring item similarity and obtaining recommendations for item-based CF; Deshpande and Karypis [18] present and evaluate a class of model based top-N recommendation algorithms that use item-to-item or item set-to-item similarities for recommendation. For both user-based CF and item-based CF, the measurement of similarity between users or items is a significant step. Pearson correlation coefficient, cosine-based similarity, vector space similarity, and so on are widely used in similarity measurement in CF methods. There are some hybrid methods such as [13]. Besides, Huang et al. [1] try to apply associative retrieval techniques to alleviate the sparsity problem. Hu et al. [19] explore algorithms suitable for processing implicit feedbacks. Umyarov and Tuzhilin [20] propose an approach for incorporating externally specified aggregate ratings information into CF methods.Recently, latent factor model has become popular. A typical latent factor model associates each user u with a user-factor vector pu, and each item i with an item-factor vector qi. The prediction is done by taking an inner product. The more involved part is parameter estimation. Some recent works such as [21] have suggested modelling directly only the observed ratings to avoid overfitting through an adequate regularized model

### Hybrid Recommender Systems

Several recommender systems (e.g., [28] and [29]) use a hybrid approach by combining collaborative and content based methods, 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 combine their predictions by a combining function, such as a linear combination of ratings or a voting scheme or other metrics. Melville et al. [28] 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. [13] propose an effective missing data prediction (EMDP) by combining item-based CF and user-based CF.

## III. TYPICALITY-BASED COLLABORATIVE FILTERING

In this section, we propose a typicality-based collaborative filtering approach , in which the "neighbors" of users are found based on user typicality in user groups instead of co-rated items of users. We first introduce some formal definitions of concepts . The mechanism of this new approach is then described below.

### Preliminaries

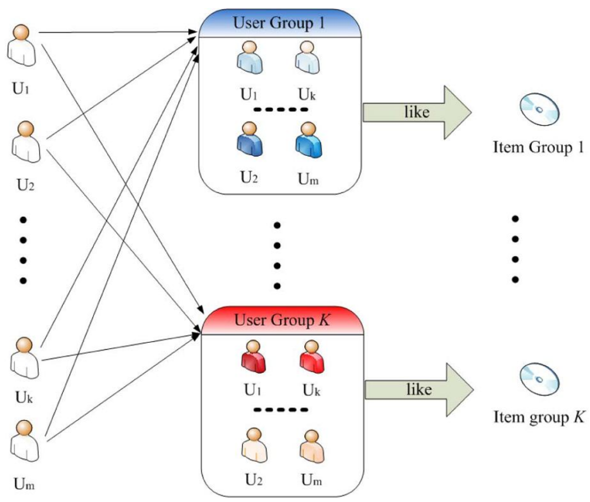
Assume that in a CF recommender system, there are a set U of users, and a set O of items. Items can be clustered into several item groups and an item group is intuitively a set of similar items. For example, movies can be clustered into action movies, war movies, and so on. Each movie belongs to different movie groups to different degrees. The choice of clustering method is application domain dependent, and is out of the scope of this paper. For instance, based on the keyword descriptions of movies, we can use Topic Model based clustering [30], [31] for the task of obtaining movie groups and the degrees of movies belonging to movie groups. In other application domains, other clustering approaches can also be used. In this paper, we will not discuss clustering methods further. The formal definition of an item group is given in the following.

Definition 3.1. An item group denoted by ki is a fuzzy set of objects, as following:

$$K_i = \{ O^{wi,1}{}_{1\,,} O^{wi,2}{}_{2} O^{wi,3}{}_{3\ldots\ldots} O^{wi,h}{}_{h}\}$$

where h is the number of items in ki, Ox is an item, and wi;x is the grade of membership of item Ox in ki.

Users who share similar interests on an item group could form a community, and we name such a community as a user group. Users have different typicality degrees in different user groups. In other words, for each item group ki, we define a corresponding user group (i.e., a fuzzy set of users who like objects in ki) to some degrees. For instance, Bob and Raymond are very interested in war movies but not so interested in romance movies, while Amy and Alice like romance movies very much but do not like war movies. Thus, Bob and Raymond are typical users in the user group of users who like war movies, but not typical users in the user group corresponding to romance movies; while Amy and Alice are typical users in the user group of users who like romance movies but not typical in that of war movies. We consider a user group gi corresponding to an item group ki as a fuzzy concept "users who like the items in ki." Note that users may have different typicality degrees in different gi. The following is the formal definition of a user group.



International Journal of Innovative Research in Advanced Engineering (IJIRAE) ISSN: 2349-2163
Issue 6, Volume 2 (June 2015) www.ijirae.com

A user group gi is a fuzzy set of users, as follows:

$g_i =\{ U_1^{vi,1,} U_2^{vi,1 \ldots\ldots} U_3^{vi,1}\}$

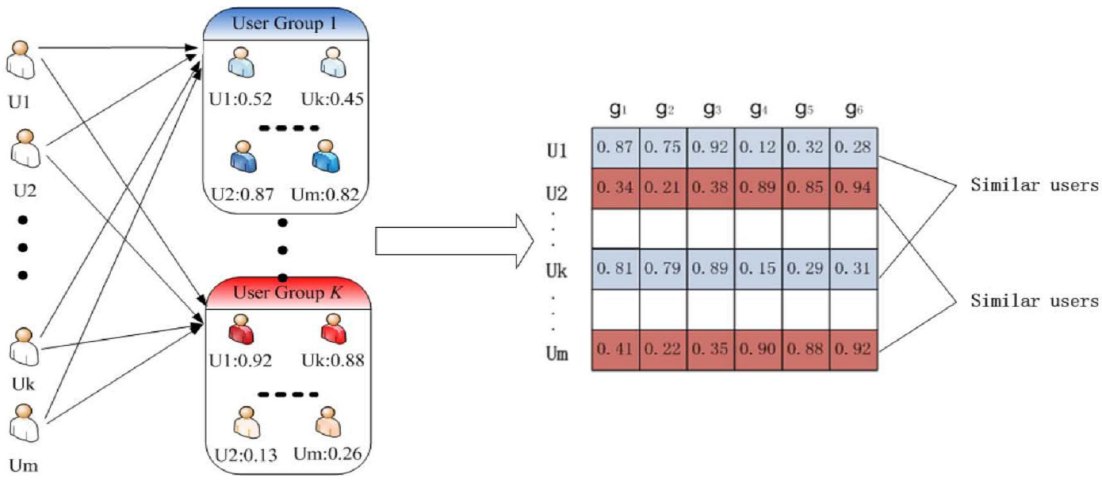
Users possess different typical degrees in different user groups: the darker a user is in Fig. 1, the more typical it is in that user groups. For examples, U1 and Uk are typical in user group gk but not typical in g1, while U2 and Um are typical in g1 but not typical in gk. For the reason that users have different typicality degrees in different user groups.

. Using this approach we can find that Bob and Tom are very typical users in the user group "users who like war movies" and not typical users in other user groups (because they have not rated any other kinds of movies in their histories). Then, we can find that Bob and Tom are very similar by comparing their typicality vectors. Such a result is intuitively more reasonable than that of traditional CF.

| | $i_1$ | $i_2$ | · · · | $i_k$ | · · · | $i_n$ |
|---|---|---|---|---|---|---|
| $U_1$ | 5 | ? | ... | 3 | ... | 4 |
| $U_2$ | ? | ? | ... | 4 | ... | 5 |
| ⋮ | ... | ... | ... | ... | ... | ... |
| $U_k$ | 2 | 5 | ... | ? | ... | 3 |
| ⋮ | ... | ... | ... | ... | ... | ... |
| $U_m$ | 5 | 4 | ... | 2 | ... | ? |

As introduced above, the typicality of an object in a concept depends on the central tendency of the object for the prototype of the concept. In other words, if an object is more similar to the prototype of a concept, it has a higher typicality degree in the concept. Generally, an item is represented by a set of properties, which, following our previous work [34], we shall call item property vector. For example, keywords, actors, directors, and producers are properties of a movie and these properties can form an item property vector to represent a movie. For each item group kj, we can extract a prototype to represent the item group.

## IV MECHANISM OF DISCOVERING SIMILAR USERS



We regard an item group as a fuzzy set and there is only one prototype to represent an item group (i.e., a cluster of similar items). Thus, the frequency of instantiation of the unique prototype for the item group is 1, and the typicality degree of an item in an item group only depends on the central tendency. Based on the works in [36] and [9], the central tendency of an object to a concept is affected by the degrees of the internal similarity and external dissimilarity. Internal similarity is the similarity of the object property vector of the item and the prototype property vector of the item group. External dissimilarity is the similarity of the object property vector of the item and prototype property vectors of other item groups.

## V. CONCLUSION AND FUTURE WORKS

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 method, a user is represented by a user typicality vector that can indicate the user's preference on each kind of items. A distinct feature of this method is that it selects "neighbors" of users by measuring users' similarity based on their typicality degrees instead of corated 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. We conduct experiments to evaluate this approach and demonstrate the advantages of this approach. In this, there are some preprocessing procedures, such as constructing user prototype by clustering and measuring user typicality in user groups. The cost of these preprocessing 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 preprocessing offline, and then adopt some parallel computing methods (e.g., MapReduce) to speed up the computing. There are several possible future extensions to our work.In this method we do 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.,MapReduce) 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 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.