Review Paper on Next Generation Collaborative Filtering of Data Based on Typicality Method

Rameshwar Gaikwad[1], Prof. Dr. Kishor. R. Kolhe[2]

[1]Department of Information Technology, Pune, India
[2]Department of Information Technology, Pune, India

Abstract: *Collaborative clustering increases the researcher's interest since from last decade for sufficient recommender systems. There are different methods presented for CF which is based on two different categories such as user based CF and item based CF. in user based CF, main task is to find out set of users who have similar favor patterns to a given user and recommend to the user those items that other users in the same set like. In item based CF method, main task is to provide a user with the recommendation on an item based on the item with high correlations. But exiting methods are having limitations like data sparsity, recommendation accuracy and prediction error. Here, we have proposed novel efficient Expectation Maximization (EM) Clustering and typicality-based collaborative filtering recommendation (EmTyCo). EmTyCo method is based on existing Tyco method by using efficient Expectation Maximization Algorithm for clustering purpose.*

Keywords: Collaborative clustering

1. Introduction

Recommendation systems found their application in the field of e-commerce and internet where items suggest to a group of user on the basis of their requirement based on their area of interest. A recommendation system is an information filtering system that built a model from the characteristic of an item according to the rating or prediction, given by a user to an item. Recommendation system has an important component in social media sites (such as Amazon, IMDB, Movie Lens); social sites giants such as Amazon have been greatly gained from the capability of their recommenders in accurately delivering the correct item to the correct user [17]. Collaborative filtering (CF) is an important and popular technology for recommender system. CF methods are classified into user based CF and item-based CF. The basic idea of user-based CF approach is to find out a set of users who have similar favor patterns or interest to a given user and the basic idea of item-based CF approach is to find out a set of items having highest correlation with the given item. 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 the category [18]. Cognitive psychologists find that objects (items) have different typicality degrees in categories in real life [19], [20], [21]. But these collaborative filtering methods have facing some problems like- 1.Data Sparsity. 2. Recommendation accuracy. 3. Big-error Prediction. 4. Scalability.

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 makes it differ from previous methods. To the best of our knowledge, there is no work on combining typicality into collaborative filtering recommendation. TyCo provides a new perspective to investigate CF recommendations. We conduct experiments to validate the proposed method and compare it with previous methods. Experiments show that TyCo has several advantages. It can obtain higher recommendation quality, work better even with sparse training data sets and produce less big-error predictions than previous methods.

2. Related Work

A. Prototype View and Typicality

According to cognitive psychology, object typicality is a measure of the goodness degrees of objects as exemplars in a concept [18]. Object typicality depends on properties shared by most of the objects of the concept, which usually include properties that are not necessary for defining the concept [7]. In the prototype view [17], a concept is represented by a best prototype or a property list which has all the salient properties of the objects that are classified to this concept. An object is considered more typical in a concept if it is more similar to the prototype of the concept. In [3], Barsalou measures a factor named central tendency which affects the object typicality in a concept. Central tendency is the degree of an objects family resemblance'. The more an object is similar to other members of a concept, the more typical it is in the concept. There are some works on measuring object typicality in computer science. Rifqi [19] proposes a method to calculate object typicality in large databases, which is later extended by Lesot et al. [13]. In their works, the object typicality of a member for a category that it belongs to depends on its resemblance to the other members of the category, and its dissimilarity to members of other categories. Cai and Leung [5] [4] have formalized object typicality by constructing several vectors in ontologies. The typicality of an object in a concept in their model 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 is no work on integrating typicality in collaborative filtering recommendation.

B. Collaborative Filtering

These kinds of recommendation methods predict the preferences of active users on items based on the preferences of other similar users or items. For the reason that collaborative filtering methods do not require well-structured item descriptions, they are more often implemented than content-based methods [7] and many collaborative systems are developed in academia and industry.

C. Hybrid Recommendation Systems

Several recommendation systems use a hybrid approach by combining collaborative and content-based methods, which helps to avoid some limitations of content-based and collaborative systems. A naive hybrid approach is to implement collaborative and content based 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.

3. Implementation: Proposed System

Subsection Problem Statement The main problem found in our existing system is a user with the recommendation on an item based on the other items with high correlations (i.e., neighbors of the item). It is significant to find users (or items) neighbor in the similar set, but it is difficult to find correlation between users and items.

A. Model

Third, we build a matrix known as user-typicality and measure similarities of users 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 particular user on an item based on the ratings of the neighbors of that user on the item. We propose a technique for an error-correction to suggest similar terms for the query keywords and return answers of the similar terms. To help users formulate high quality queries, as users type in keywords, we suggest keywords that are topically (popularly) relevant to the query keywords and we propose a query expansion-based technique to recommend users relevant keywords.

B. Algorithm Used

There are many algorithms that can be used to create item clustering. In this paper, we choose the k means algorithm as the basic clustering algorithm. The number k is an input to the algorithm that specifies the desired number of clusters. Firstly, the algorithm takes the first k items as the centers of k unique clusters. Each of the remaining items is then compared to the closest center. In the following passes, the cluster centers are re-computed based on cluster centers formed in the previous pass and the cluster membership is re-evaluated.
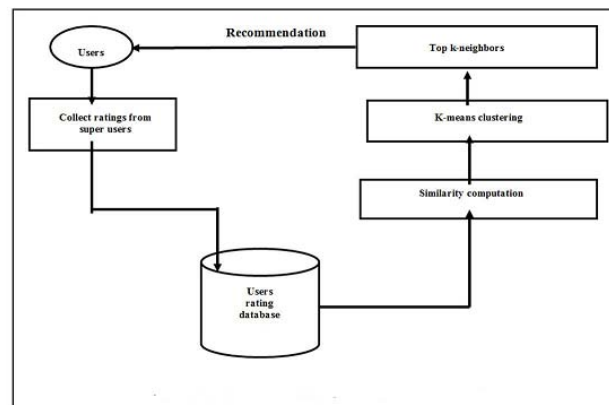

Figure 1: Architecture of proposed system using collaborative Filtering

In this paper, we borrow the idea of object typicality from cognitive psychology and propose a typicality-based CF recommendation approach named OTCO. The basic mechanism of object typicality-based CF recommendation is as follows: First, we cluster all items into multiple item groups. 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), in each of the user groups with all users having different typicality.

No	Algorithm Name	Advantage over Existing Algorithm
1	Cluster Model Algorithm	1. Cluster models divide the customer base into many segments and treat the task as a classification problem. 2. The algorithm's goal is to assign the user to the segment containing the most similar customers. 3. It then uses the purchases and ratings of the customers in the segment to generate recommendations. 4. Using a similarity metric, a clustering algorithm groups the most similar customers together to form clusters or segments.
2.	K-means clustering algorithm	1. Fast, robust and easier to understand. 2. Gives best result when data set are distinct or well separated from each other.

Figure 2: Algorithm Used

Specific algorithm as follows: Input: clustering number k, user-item rating matrix Output: item-center matrix Begin Select user set U=U1, U2, Um;

Select item set I=I1, I2, In;

Choose the top k rating items as the clustering CI=CI1, CI2,

CIk;

The k clustering center is null as c=c1, c2, ck; do

for each item IiI

for each cluster center CIjCI calculate the sim(Ii, CIj); end for

sim(Ii, CIx)=maxsim(Ii, CI1), sim(Ii, CI2), c, sim(Ii, CIk);

cx=cxIi end for

for each cluster cic for each user IjI CIi=average(ci, Ij); end for

end for

while (CU and c is not change) End

C. Explanation

The ratings are collected from the super user and are represented in the form of user-item matrix. The ratings of the super user are compared with other users in the rating database and their similarities are computed using Pearsons correlation coefficient. Using the similarity values we cluster the users based on k-means clustering approach and find the top k-neighbors for producing recommendations. The recommendations from the top k neighbors are the products that the super user has not accessed yet that are given high ratings by their top neighbors.

4. Conclusion

In this paper, we investigate the recommendation system from a new perspective and present a novel recommendation method based on object typicality. A distinct feature of the typicality-based recommendation method is that it predicts ratings based on user typicality and item typicality. The higher typicality degrees of users and items in the corresponding user and item groups, the higher recommendation scores. To the best of our knowledge, it is the first work which applies typicality to handle recommendation problem. Evaluation shows that the typicality-based method outperforms previous recommendation methods on recommendation quality.

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] Y. Cai and H. F. Leung. Multi-prototype concept and object typicality in ontology. In Proceedings of the 21st International Florida Artificial Intelligence Research Society Conference, pages 470475. AAAI Press, 2008.

[5] W. Vanpaemel, G. Storms, and B. Ons, A Varying Abstraction Model for Categorization, Proc. Cognitive Science Conf. (CogSci 05), pp. 2277-2282, 2005

[6] M.-J. Lesot, L. Mouillet, and B. Bouchon-Meunier, Fuzzy Prototypes Based on Typicality Degrees, Proc. Intl Conf. Eighth Fuzzy Days 04, 2005.

[7] C. Li and G. Biswas. Unsupervised learning with mixed numeric and nominal data. IEEE Transactions on Knowledge and Data Engineering, 14:673690, 2002.

[8] R. Xu and D. Wunsch. Survey of clustering algorithms. IEEE Trans. Neural Network, 16(3):645678, 2005

[9] J. Wang, A. P. de Vries, and M. J. T. Reinders. Unifying user-based and item-based collaborative filtering approaches by similarity fusion. In SIGIR 06, pages 501508, New York, NY, USA, 2006. ACM.

[10] B. Sarwar, G. Karypis, J. Konstan, and J. Reidl. Item-based collaborative filtering recommendation algorithms. In WWW 01, pages 285295, New York, NY, USA, 2001. ACM