Scientific Journal of Impact Factor (SJIF): 5.71
e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 5, Issue 03, March -2018

ITEM BASED COLLABORATIVE FILTERING FOR RECOMMENDATION SYSTEM

Shradha Basagre[1], Amruta Kamble[2], Supriya Zende[3], Gouri Tavate[4], Vaibhavi Kharge[5]

[1]*Computer Science and Engineering, Sanjay Ghodawat Group of Institutions, Kolhapur, Maharashtra, India*
[2]*Computer Science and Engineering, Sanjay Ghodawat Group of Institutions, Kolhapur, Maharashtra, India*
[3]*Computer Science and Engineering, Sanjay Ghodawat Group of Institutions, Kolhapur, Maharashtra, India*
[4]*Computer Science and Engineering, Sanjay Ghodawat Group of Institutions, Kolhapur, Maharashtra, India*
[5]*Computer Science and Engineering, Sanjay Ghodawat Group of Institutions, Kolhapur, Maharashtra, India*

Abstract: *Recommendation system is used to recommend information to users as per their expectations and provide personalized services through analyzing the behaviors. Buying trend in India shifting from time consuming shop visits to highly flexible online shopping. There is increasing number of customers for online shopping and their likes and dislikes are also different so it is very challenging to generate recommendation system. These are producing high quality recommendations, performing many recommendations per second for millions of users and items. Previous collaborative filtering systems the amount of work increases with the number of participants in the system. New recommender system technologies are needed that can quickly produce high quality recommendations,even for very large-scale problems. To address these issues we have explored item-based collaborative filtering techniques. We are proposing item based collaborative filtering technique in which,Item-item collaborative filtering, or item-based, or item-to-item, is a form of collaborative filtering for recommender systems based on the similarity between items calculated using people's ratings of those items.*

Keywords: *Recommendation Engine, Collaborative Filtering, Item Based Similarity.*

I. INTRODUCTION

Nowadays, more and more people are using smart phone, personal computer, tablets and other intelligent terminals. People uses social networking sites such as Amazon, eBay etc. Traditional e-commerce had some problems such as, the provider had limited target customers as they can advertise and reach few prospective customers, provider could not attract customers to buy more products which customer does not intend to buy. Sometimes, customers spend more time in searching for their expected products. Luckily, the behaviors of users can be tracked and recorded on the social networks and e-commerce sites. This makes it easier to analyze the preference of users and provide better service. For this purpose, we can use recommendation system.

A recommendation system is illustrated as the techniques used to predict the rating one individual will give to an item or social entity. There are two approaches first is content based and second is collaborative filtering which involves characteristics of an item. Second collaborative filtering approaches, which considers user's past behavior to make choices. In collaborative filtering, users are chosen who will make recommendations because they provide similar ratings history with the appropriate user. One partner who has similar ratings to the target user may not be a reliable predictor for an item. So, the past record of the user of making anauthentic recommendation also needs to be take into consideration which is dictated by trustworthiness of a partner. To maintain track of past records of a recommend reputation systems comes into the picture those who assign reputation ratings to the users.

Recommendation system is used to recommend information to users as per their expectations and provide personalized services through analyzing the behaviors. The thought behind the recommendation systems for e-Commerce is to build correlation between the products (items), users (visitors/customers) and make decision to choice the most suitable product to exact user. For this, Recommendation systems may use data mining, machine learning algorithms. During a learning phase, the system builds the model which is an abstraction of the relationship between the items and users.

Recommendation system on an ecommerce website can generate product recommendations, create personalized emails and merchandise products on your site, Drive more Traffic to you website, Deliver Relevant Content to users, Engage Shoppers in your website, Convert Shoppers to Customers, Increase Average Order Value, Increase Number of Items per Order, Control Merchandising and Inventory Rules, Reduce Workload and Overhead, Provide Reports, Offer Advice and Direction.

Perhaps the biggest issue facing recommender systems is that they need a lot of data to effectively made recommendations.

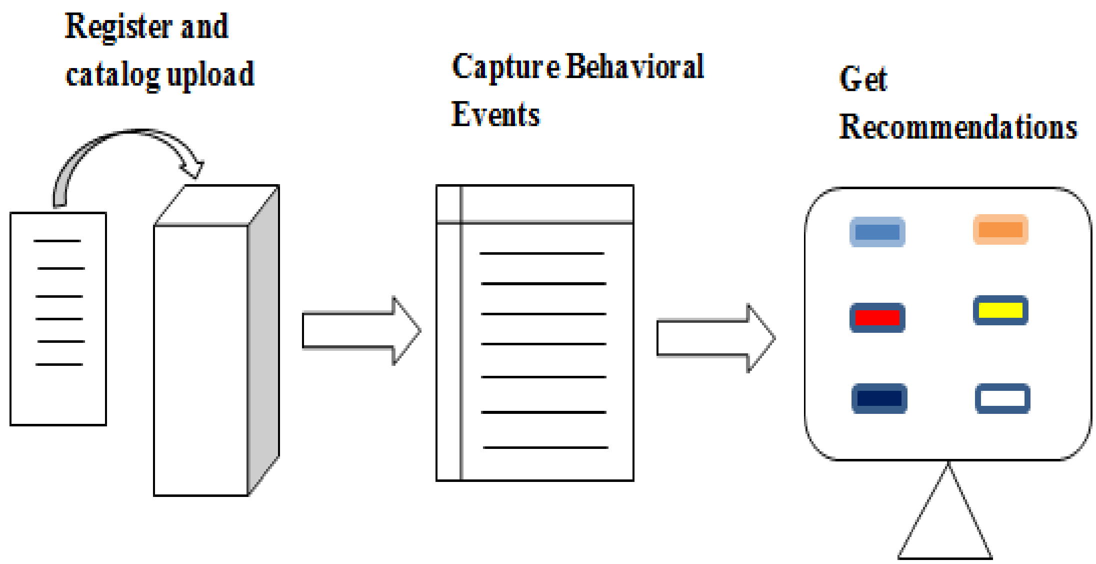

Figure 1: Flow of recommendation system

When there's not enough information to build a firm profile for a user, the recommendation could not be provided properly. In past, user - based collaborative filtering was popular but there was a problem of lack of large item sets. So, the accuracy of recommendations may be poor.

The main idea here is to analyze the user-item representation matrix to identify relationship between various items and then to use these relations to calculate the estimation score for a given user-item pair. The perception behind this approach is that a user would be concerned in purchasing items that are similar to the items the user be fond of earlier item and would tend to avoid items that are similar to the items the user didn't like earlier. These techniques don't require identifying the neighborhood of similar users when a recommendation is requested; as a result they tend to produce much faster recommendations.

The collaborative filtering [1] has become most widely used method to recommend items for users. It makes recommendation as per similar users with the active user or the similar items with the items which are rated by active user. The collaborative filtering includes memory based method and model based method [2]. The memory-based method first calculates the similarities among users and then selects the most similar users as the neighbors of the active user. Finally, it gives the recommendations as per the neighbors. However, the model-based method first constructs a model to describe the behavior of users and, therefore, to predict the ratings of items. The memory-based method can give considerable recommended accuracy, but the computing time will grow rapidly with the increasing of users and items.

We are going to use Collaborative filtering algorithm. The root of collaborative filtering is to computecorrespondences among users or items. The common traditional correspondence measures, such as Pearson correlation coefficient [3], cosine [4], mean squared difference [2], are insufficient to capture the effective similar users, especially for cold user who only rates a small number of items.

II. RELATED WORK

Collaborative filtering (CF), as a kind of personalized recommendation technique, has been widely used in many domains [1–3,5,6]. However, collaborative filtering also suffers from afew of issues, for instance, cold start problem, scalability and so on. These problems seriously reduce the user experience. This paper emphases on how to improve the prediction precision. Collaborative filtering recommends items to users as per their preferences. Therefore, a previous database of user's consideration must be available. However, the database is always very spare, that is, user only rates a few number of items. Up to now, there are many researchers who have concentrated on the prediction precision and proposed some solutions. To improve the accuracy, many researchers have proposed some new similarity measures.

An [14] proposed a new similarity for collaborative filtering that is called PIP (Proximity-Impact-popularity). We considered the drawbacks of Pearson correlation coefficient [3] and cosine correspondence[4]. This new correspondence considered three aspects: proximity, impact and popularity of the user ratings. But, this correspondencedeliberates only the local information of the ratings and does not consider the global preference of user ratings. Traditional Pearson correlation coefficient does not consider the size of the set of common users. To overcome this problem, weighted Pearson correlation coefficient has been projected[9]. It considers the idea of taking the confidence which can be placed on the neighbor.

The confidence will increase with the number of common rated items. Jamali and Ester [8] familiarized a correspondence measure based on the sigmoid function. This method canweak inthe similarity of small mutual items among users. The adjusted cosine similarity measure [7] was proposed to make up the shortage of traditional cosine similarity; however, it did not deliberate the preference of user ratings. Bobadilla et al. [9] proposed a new metric which combined theJaccard measure and mean squared difference [2]. It assumed that these two measures could complement each other. Additional new metric, which is called MJD (Mean–Jaccard–Difference), was suggested to solve the cold user problem.

This metric contains three steps: first the selection of correspondence measures, the new metric has six related measures after this step. Then, the weights of each related measure will be calculated by neural network learning. Finally, the estimation can be obtained with respect to the new metric. Recently, a singularity based similarity measure (SM) was also presented. This measure assumed that the results obtained by applying traditional correspondence measures could be improved by taking relative information. This paper first deliberated the rating as positive and negative. Then, it calculated the originality values of each user and each item. It replaced the similarity with singularity value. The experiments verified the effectiveness of this approach. Moreover, Bobadilla et al.[10] introduced a significance based similarity measure. This measure first calculates three kinds of significances, which is the significance of an item, the significance of a user to recommend to other users and the significance of an item for a user. Then the traditional Pearson correlation coefficient or cosine similarity will be used to evaluate the correspondences among users according to the consequence.

Data smoothing technique is another most used method to improve the recommend performance in collaborative filtering. Various sparsity measures [11] were used to enhance accuracy of collaborative filtering. These lightly measures were calculated based on local and global comparisons. Then, an estimating parameter scheme for weighting the various sparsely measures was proposed.

The experimental results demonstrated that the proposed estimate parameter outperforms the schemes for which the parameter was kept constant on accuracy of prediction ratings. A et al. [12] proposed a partial missing data prediction algorithm, in which the information of both users and items was taken into account. In this algorithm, similarity threshold for users and items was set respectively and the missing data will be predicted if and only if, the intersection of the neighbors of user and the neighbors of item is not empty.

The iterative prediction method [13] clusters the user and item respectively by using spectral clustering algorithm. Then, the iterative prediction technique is used to convert user-item sparse matrix to dense one based on the explicit ratings. Beyond that, dimensionality reduction technique, such as principle component analysis (PCA) [14] and singular value decomposition (SVD), is commonly used to alleviate the problem. Combined the SVD and item-based recommender in CF. It utilized the results of SVD to fill the missing ratings and then used the traditional item-based method to recommend.

This combination method can increase the accuracy of system. Moreover, hybrid methods are also proposed. [15] Investigated a hybrid recommendation method which was based on two-stage data processing–dealing with content features describing items and handing user behavioral data. This hybrid method combined random indexing (RI) technique and SVD to pre-process the content features.

The experiments improved the recommendation accuracy without increasing the computational complexity. Probabilistic matrix factorization [16] is also combined in social recommendation to solve data scarcity. Moreover, cluster-based smoothing method [17], support vector machine (SVM) [18], BP neural networks [19] and zero-sum reward and punishment mechanismare also applied to smooth the missing ratings for the solution of accuracy in collaborative filtering.

III. PROPOSED SYSTEM ARCHITECTURE:

Generally, a pattern recognition based recommender system consists of two phases; the first phase is clustering followed by classification task. In the first phase, the system is delivered with enough learning so that the classification precision of the system is quite higher at the preferred level. After the system learns, it generates a set of recommendations with appropriate rankings.

In our system, we first formed clusters to acquire knowledge about web users and the classification technique was used later for enhancing the learning capability and to generate recommendations. A web user may have multiple interests for which he needs to be put into multiple clusters. Hence, we have used a similarity upper approximation based clustering algorithm. In order to capture sequential behavior of the user, we utilized S3M [16] similarity measure while forming clusters. Soft clusters allow elements to appear in more than one cluster.

This means a data point can represent the attributes of more than one cluster. Once the clusters are formed, we utilized the singular valued decomposition to classify the web user sessions. In Figure. 2, we have outlined the general architecture of the system.

The first step is the collection of web data through web logs. After collecting web logs, there-processing is done, followed by the clustering stage. In the clustering module, each sequence is considered as a data point and all the points are clustered into several groups using a rough set based clustering algorithm that generates soft clusters allowing multiple interests of the users. After clustering, for any new user forwhomrecommendationhas to be generated, Top clusters are identified based on the similarity between user and cluster centers.

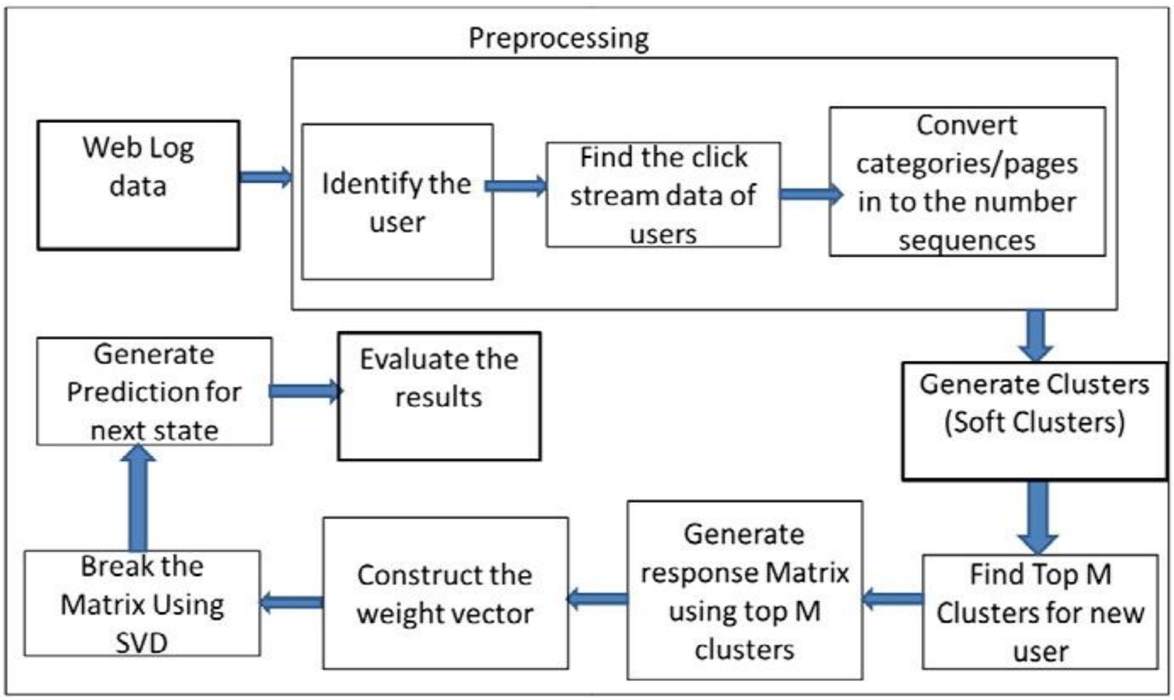

Figure 2: Proposed System Architecture for collaborative filtering[22]

IV. DETAILED RECOMMENDATION SYSTEM:

4.1 Item Based Collaborative Filtering:

Collaborative filtering methods can be extended and developed in consideration of both external and internal item-totem relations. Regarding these concerns, we reviewed challenges in collaborative filtering methods and possibilities for utilization of social network analysis methods in the previous section. Based on the reviews, a new recommendation method, item-network-based collaborative filtering, is proposed and four steps in the process are described. Through the application of our method to sample data, we expect that the consideration of item-to-item relations can potentially remedy the external relation dependency problems.[17]

The Comparison between item-based collaborative filtering and item-network-based collaborative filtering is shown in following diagram:

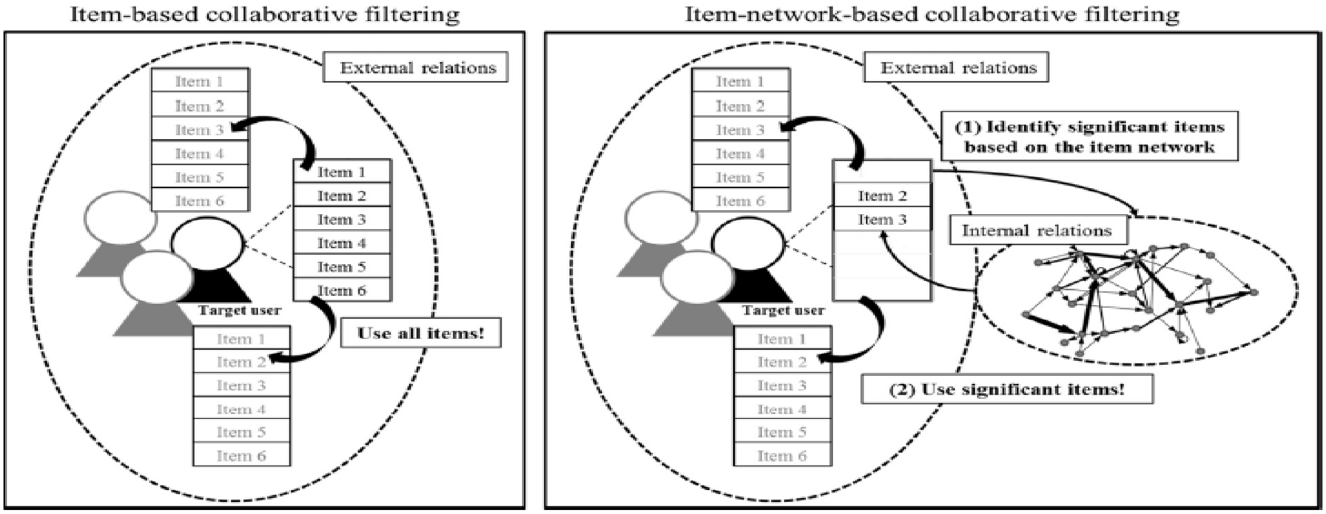

Figure 3: Comparison between item-based collaborative filtering and item-network-based collaborative filtering.[21]

4.2 ITEM-BASED COLLABORATIVE FILTERING ALGORITHM:

In this section, we discuss a class of item-based recommendation algorithms for producing predictions to users. Unlike the user-based collaborative filtering algorithm, the item-based approach looks into the set of items the target user has rated and computes how similar they are to the target item *I* and then selects *k* most similar items {*i*1, *i*2, . . . , *ik*}. At the same time their corresponding similarities {*si*1, *si*2, . . . ,*sik*} are also computed. Once the most similar items are found, the prediction is then computed by taking a weighted average of the target user's ratings on these similar items. We describe these two aspects namely, the similarity computation and the projection generation in details here.

4.1.2 Item Similarity Computation:

The most difficult step in this item-based filtering algorithm is computing similar items with items and then recommending most equivalence items. The basic idea in similarity computation between two items *i* and *j* is to first separate

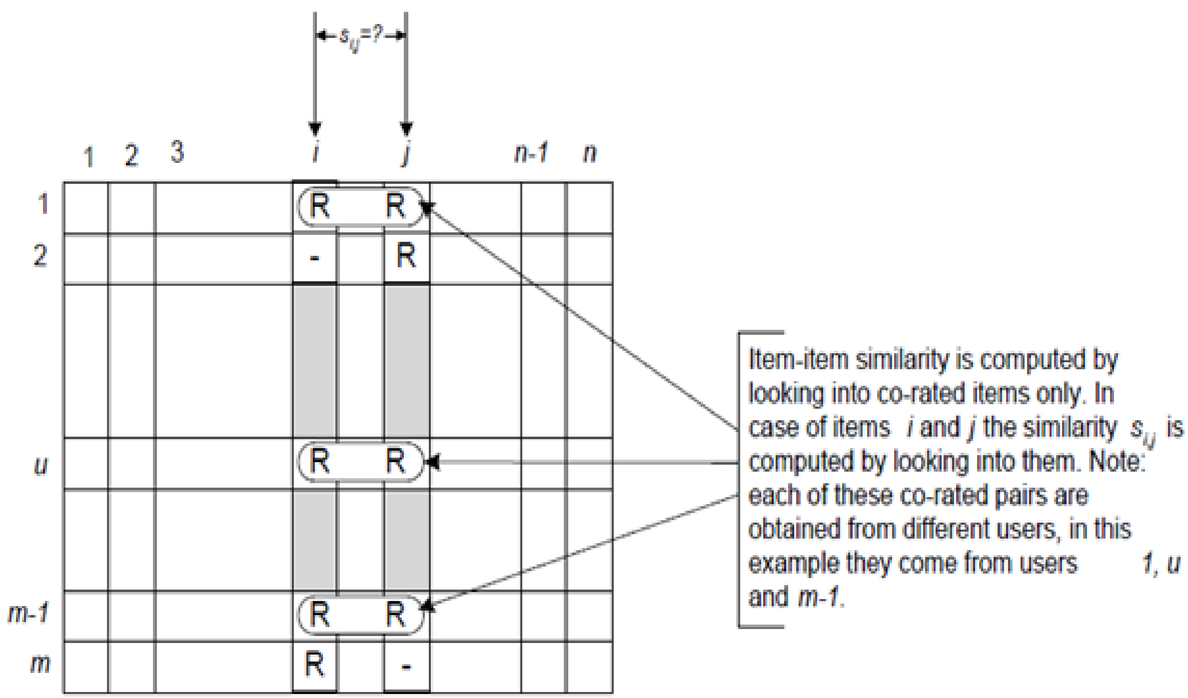

Figure 4: Isolation of the co-rated items and similarity computation[20]

The users who have calculated both of these items and then to apply a parallel computation technique to determine the similarity *si, j* . Figure 4 illustrates this process, here the matrix rows represent users and the columns represent items. There are a number of different ways to compute the similarity between items. Here we present three such methods. These are cosine-based similarity, correlation-based similarity and adjusted-cosine similarity[20].

4.2.2 Cosine-based Similarity:

In this case, two items are view of as two vectors in the *m* dimensional user-space. The similarity between them ismeasured by computing the cosine of the angle between these two vectors. Formally, in the $m \times n$ ratings matrix in Figure 2, similarity between items *i* and *j*, denoted by *sim(i, j)* is given by[20].

$$sim(i,j) = \cos(\vec{i},\vec{j}) = \frac{\vec{i}\cdot\vec{j}}{\|\vec{i}\|_2 * \|\vec{j}\|_2}$$ ……………………………….eq(1)

Where "·" denotes the dot-product of the two vectors.

4.2.3 Correlation-based Similarity:

In this case, correspondence between two items *i* and *j* is measured by calculating the *Pearson-r* correlation *corr i, j*. Tomake the correlation calculationprecise we must first separate the co-rated cases (i.e., cases where the users rated both *i* and *j*) as shown in Figure 2. Let the set of users who both rated *i* and *j* are represented by *U* then the correlation correspondence is given by

$$sim(i,j) = \frac{\sum_{u\in U}(R_{u,i}-\bar{R}_i)(R_{u,j}-\bar{R}_j)}{\sqrt{\sum_{u\in U}(R_{u,i}-\bar{R}_i)^2}\sqrt{\sum_{u\in U}(R_{u,j}-\bar{R}_j)^2}}.$$ ……………………….eq(2)

Here *Ru,I* signifies the rating of user *u* on item *i* , .*R*is the average rating of the *i* -th item[20].

4.3 EVALUATION METRICS:

Accuracy is an important indicator for the evaluation of recommended system performance. As one of the most commonly used methods, the Mean Absolute Error (MAE) is adopted as a metric in this paper to compare the prediction quality of our proposed approach with other collaborative filtering methods. Supposing the top-*N* prediction rating set for the active user is $p_1,p_2,\ldots,p_N$, and corresponding actual rating set $q_1,q_2,\ldots,q_N$, the MAE can be defined as follows:

Where, *N* is the number of the items recommended to the active use. The lower the MAE denotes more accuracy in the prediction for user interest of the recommendation system.

V. CONCLUSION

Recommendation algorithms provide an effective form of targeted marketing by creating an excellent shopping experience for each customer. These systems help users find items they want to buy. Recommendation systems benefit users by enabling them to find items they like. Recommendation systems are rapidly becoming tool in E-commerce on the Web.In this paper we are evaluating algorithm for CF-based recommender systems.

VI. ACKNOWLEDGMENT

We are thankful to Mrs. Pallavi Kulkarni for her valuable inputs and guidance for this work.

VII. REFERANCES

[1] Acilar AM, Arslan A. A collaborative filtering method based on Artificial Immune Network. ExpSystAppl 2009;36(4):8324–32.

[2] Rashid AM, Albert I, Cosley D, Lam SK, McNee SM, Konstan JA et al. Getting to know you: learning new user preferences in recommender systems. In: Proceedings of the international conference on intelligent user interfaces; 2002. p. 127–34.

[3]Goldberg K, Roeder T, Gupta D, Perkins C. Eigentaste: a constant time collaborative filtering algorithm. Inform Retrieval J 2001;4(2):133–51.

[4]Billsus D, Pazzani MJ. User modeling for adaptive news access. User Model User-adapted Interact 2000;10(2–3):147–80.

[5] Adomavicius G, Tuzhilin A. Toward the next generation of recommender system. A survey of the state-of-the-art and possible extensions. IEEE Trans Knowl Data Eng 2005;17(6):734–49.

[6] Ziegler CN, McNee SM, Konstan JA, Lausen G. Improving recommendation lists through topic diversification. In: Proceedings of the 14th international conference on World Wide Web; 2005. p. 22–32.

[7] Cunningham P, Bergmann R, Schmitt S, Traphoner R, Breen S, Smyth B. ‘‘WebSell: Intelligent sales assistants for the World Wide Web. In: Proceedings CBR in ECommerce, Vancouver BC; 2001. p. 104–9.

[8] Schwab I, Kobsa A, Koychev I. Learning user interests through positive examples using content analysis and collaborative filtering. Draft from Fraunhofer Institute for Applied Information Technology, Germany; 2001.

[9] Melville P, Mooney-Raymond J, Nagarajan R. Content-boosted collaborative filtering for improved recommendation. In: Proceedings of the eighteenth national conference on artificial intelligence (AAAI), Edmonton, Canada; 2002. p. 187–92.

[10] Linden G, Smith B, York J. Amazon.com recommendation: itemto-item collaborative filtering. IEEE Internet Compute 2003;7(1): 76–80.

[11] Wasfi AM. Collecting user access patterns for building user profiles and collaborative filtering. In: Proceedings of the 1999 international conference on intelligent user, Redondo Beach, CA; 1999. p. 57–64.

[12] Berry MJA, Linoff G. Data mining techniques: for marketing, sales, and customer support. New York: Wiley Computer Publishing; 1997.

[13] Ghazantar MA, Pragel-Benett A. A scalable accurate hybrid recommender system. In: the 3rd International conference on knowledge discovery and data mining (WKDD 2010), IEEE Computer Society, Washington, DC, USA.

[14] Kuzelewska U. Advantages of information granulation in clus- _ tering al-gorithms. In: Agents and artificial intelligence. NY: Springer; 2013. p. 131–45.

[15] Burke R. Hybrid web recommender systems. In: Brusilovsky P, Kobsa A, Nejdl W, editors. The adaptive web, LNCS 4321. Berlin Heidelberg, Germany: Springer; 2007. p. 377–408.

[16]P. Kumar, B.S. Raju, P.R. Krishna, A new similarity metric for sequential data, International Journal of Data Warehousing and Mining (IJDWM) 6 (4) (2010) 16–32.

[17]TaehyunHa,SangonLee Item Network Based Collaborative Filtering; A Personalized Recommendation Method Based On Users Item Network.

[18]BadrulSarwar, George Karypis, Joseph Konstan, And John Riedl, Item Based Collaborative Filtering Recommendation Algorithms.

[19]Guanwen Yao, Lifeng Car, User Based And Item Based Collaborative Filtering Recommendation Algorithm Design.

[20]BadrulSarwar, George Karypis, Joseph Konstan, and John Riedl. Item-based Collaborative Filtering Recommendation Algorithms
[21]Taehyun Ha, Sangwon Lee.Item-network-based collaborative filtering: A personalized recommendation method based on a user's item network.
[22] Rajhans Mishra, Pradeep Kumar, Bharat Bhasker.A Web Recommendation System Considering Sequential Information

