ISSN No. 0976-5697

Volume 8, No. 5, May-June 2017

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Design and Analysis of Efficient Collaborative Filtering Based Optimization Approach

Krishan Kant Yadav
Research Scholar
Amity School of Engineering & Technology
Amity University, Gwalior, Madhya Pradesh

Dr. Sanjiv Sharma
Assistant Professor
Department of Computer Science Engg. & IT
Madhav Institute of Technology & Science, Gwalior

Dr. Manisha Singh
Associate Professor
Department of Applied Physics
Amity School of Engineering & Technology
Amity University, Gwalior, Madhya Pradesh

Abstract: The large amount of information that is currently being collected (the so-called "big data"), have resulted in model-based Collaborative Filtering (CF) methods to encountering limitations, e.g., the sparsity problem and the scalability problem. It is difficult for model-based CF methods to address the scalability-performance trade-off. Therefore, we propose a scalable clustering-based CF method in this paper that can help provide a balance by re-locating elements in the cluster model. The rapid development of information technology takes our shopping into the orbit of information. With the network construction of resources, the amount of shopping resources increases rapidly. Collaborative filtering (CF) is a technique commonly used to build personalized recommendations on the Web. Some popular websites that make use of the collaborative filtering technology include Amazon, Netflix, iTunes, IMDB. The most important issue which influences the collaborative filtering recommendation accuracy is the so-called data sparseness. Data sparseness causes the system difficulty in determining the nearest neighbors of the target user accurately. Clustering can solve this problem to some extent. Grouping a set of physical or abstract objects into classes of similar objects, this process is called as clustering. This paper presents the methods to generate recommendations using clustering-based collaborative filtering approach.

Keywords: *Collaborative Filtering (CF); DMT; and NBCF etc.*

I. INTRODUCTION

Today internet is another best medium for doing business. Now days numerous companies are adopting new business strategies to increase the business. Business over the internet allows prospective customers to get information about products, their features and comparison among them. E-commerce breaks the constraints of time and space as compared to the traditional business. Large companies around the world now understand that e-commerce is not a process of retailing and promotion over the Internet. They understand there is ample space for business growth. Business houses improve the competence on other in the existing market. Data mining sometimes also known as knowledge discovery is used for this purpose. On the other hand, web mining is a data mining technique that is applied on the World Wide Web.

The recent development in the Internet technology is very useful for the growth of enterprises. Almost every organization is on the internet and many of them have started their business through e-commerce. Organizations collect information about customers and their business transactions which are very beneficiary for their business growth. Data mining is a set of techniques used to extract unknown pieces of information from the large database repository. There are various data mining techniques (DMT) available to extract valuable and useful information for enterprises. Data mining techniques help e-commerce business in many ways.

The term Data Mining refers to the finding of relevant and useful information from large databases. Data mining or Knowledge Discovery in the database is the non-trivial extraction of hidden, previously unknown and much useful knowledge from the huge amount of databases. This includes many technical approaches, such as data clustering, data classification, generate new rules, analyzing changes and detecting anomalies.

Electronic commerce organizations use data mining to get hold of the dependable market and consumer response. For example, the information obtained may help organizations to undertake targeted promotion, marketing strategies, decide commercial positioning, and diminish the working expenditures.

II. LITERATURE SURVEY

In this study, a systematic and comprehensive survey in the field of Collaborative Filtering (CF) is carried out. Collaborative Filtering is an important information filtering technique for providing the particularly recommended approach to its users [Figure-1]. This survey is carried out with following objectives:

- A selective study of Recommender System.
- A comprehensive evaluation of the available filtering techniques.
- A study of collaboration techniques.
- A comprehensive study of collaboration algorithms.
- Identify hot issues and challenges in the area of recommendation algorithm in E-commerce research.

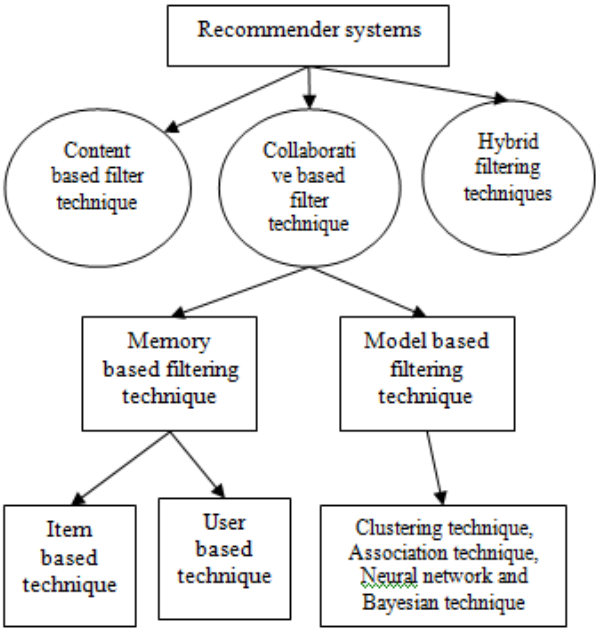

Figure1: Recommendation Techniques

A Recommender system (RS) is defined as a decision-making strategy for users under complex information environments [1]. A recommender system was coined from the perception of e-commerce as a device that helps customers search through the report of knowledge which is related to customer's preferences and interest [2]. Recommender systems was defined as a means of supporting and enhance the method of using recommendations of others to make decisions when there is no adequate information or understanding of the choices [3]. Recommender systems solve the problem of information overload that customers come upon by providing them with adapted, limited content and observe recommendation [4]. Many techniques have been developed for use in structured recommendation system, which can exploit either collaborative filtering, content-based or hybrid filtering [5].

Collaborative filtering approach is the most established and mainly realized technique that constructs a data base (user-product matrix) of favorite's products by its customers. It then matches customers with relevant preferences by calculating likeness among their profile to make the particular recommendations [6]. Collaborative filtering suggests things by classifying other customers with a similar flavor. It spawns their judgment to suggest item to the buyer. Collaborative recommender systems have been appreciated in many application areas. GroupLens is a news based architecture which engaged collaborative techniques in supporting customers to put articles from enormous news knowledge base [7].

Ringo is an online community information filtering system that uses collaborative filtering to construct customers profile based on their ratings on song albums [8]. Amazon.com uses area diversification process for enhancing their recommendation system [9]. The system uses collaborative filtering approach to conquering scalability problem by producing a table of related products offline through the use of the product to product matrix. Then the system recommends other items which are very similar to given the customer's purchase history. In another technique, that is in content-based technique match content resources to user characteristics. Content-based filtering approach finds the similarity between one user resources purchased to the user likely to be purchased. Content-based filtering approach works on the forecasting on the customer's information and they ignore contributions from other customers as with the case of collaborative approach [10, 11]. Another approach that uses content-based filtering to assist customers get information on the internet includes Letizia [12]. This particular system creates an interactive user interface that helps customers in browsing the World Wide Web. This system is capable of identifying the web browsing prototyping of a customer to forecast the web pages that they have interested [13], developed an intellectual system that attempts to forecast which web pages will interest a customer by using Bayesian classifier. The agent allows a customer to provide training examples by ranking different pages as either hot or cold. [14] developed a neural network model that shows the interests of a client in a Usenet news environment. From the success of these two filtering approaches, there are also some problems have been identified. Some limitations exist with the content based filtering approach that is limited content analysis, over-specialization and sparsity of data [7]. As well as collaborative technique suffers from cold-start, sparsity and scalability problems. In fact, these limitations reduce the quality of recommendations. Although there is hybrid filtering technique which is a combination of at least two or more filtering techniques use a different way to increase the accuracy and performance of the recommender system [15]. These approaches combine two or more filtering techniques in order to control their strengths while pointing out their resultant limitations [16]. They can be classified based on their operations into a mixed hybrid, weighted hybrid, switching hybrid, feature-combination hybrid, cascade hybrid, feature augmented hybrid and meta-level hybrid [17].

Since content-based filtering and collaborative filtering techniques are widely used today by implementing content-based and collaborative techniques differently and the results of their prediction adding the properties of content based and collaborative filtering approach. Finally, a general integrated model which includes equally content based and collaborative filtering characteristics could be developed [7]. The problem of the sparsity of data was coined by combining the ratings, features and demographic information about products in a cascade hybrid recommendation technique in [18]. A hybrid collaborative filtering technique was discussed to utilize mass taxonomic information designed for rigorous item categorization to address the data sparsity crisis of collaborative filtering recommendations, based on the construction of profiles via deduction of the tremendous topic score and issue diversification [19]. A combined technique is also suggested in [18]. This technique uses the content based profile of individual customer to find parallel customers which are used to formulate forecasting. In collaborative filtering was united with an information filtering representative. They developed a framework for incorporating the technique of content-based filtering and collaborative filtering [20]. A hybrid recommender procedure is employed by several applications as a result of new customer problem of content-based filtering approaches and average customer problem of collaborative filtering [21].

A simple and easy method for the combination of content-based and collaborative filtering was developed by [22]. A music recommendation system which joined cataloging information, play counts and social associations were developed in [23]. In order to find out the number of neighbors that can be repeatedly related on a social platform, it implanted social information into collaborative filtering algorithm [24].

A Bayesian hybrid produces model that incorporates customer ratings, customer and product features in a single integrated framework was developed by Condiff M.K. [25].

Memory based techniques (MBT) for our collaborative filtering (C.F.) categorize the likeness among two different customers by comparing their giving ratings on a particular item set. Since memory-based techniques suffer from two major problems i.e. Data Sparsity and problems with Scalability. With the use of data mining techniques the information filtering process can be performed former to the real recommendation process, as an outcome, the machine response time could be improved and making the framework more scalable.

III. PROBLEM STATEMENT

Recommender systems have appeared as powerful tools for helping customers find and evaluate products of interest. Since these methods use a range of methods to assist customers identify the particular products that best fit their requirements. Recommender systems have developed a particular interactive environment on the World Wide Web. They used analysis methods to the problem of serving customers as well as they which products they would like to buy from the E-Commerce database by producing a list of top recommended products for a given customer. The collaboration system only utilizes the user suggested product rating matrix to make future recommendation and forecasts, avoiding the need for collecting wide-ranging information about products and customers. Collaboration filtering can be easily accepted in many recommender systems without acquiring particular domain information.

Most of the researchers have considered Data Sparsity and Scalability as the main problem in collaborative filtering. A customer based collaborative filtering system is very useful but their common use has also exposed some possible issues such as: Data Sparsity and Scalability. In data Sparsity, there are several profitable recommender systems which are used to estimate large item sets. For example, the E-commerce website Amazon.com suggests books and CDnow.com recommends music albums. In such types of systems, a dynamic customer may have buy well under 1% of the products (1% of 2 million books is 20,000 books). So a recommender system based on nearest neighbor approaches may be not capable to build any product recommendations for a specific customer. In the term of Scalability, nearest neighbor approaches that developed with the number of users and the number of items, demand computations that are more time consuming as the number of customers and products increases.

3.1 Pros and cons of collaborative filtering methods-

Collaborative Filtering has major benefits over Content Based Filtering or knowledge based filtering, the content based filtering technique can perform in the area where there is not large content related with products and content is difficult for a computer system to analyze such as ideal or opinions. Content based technique also finds the problem in adding new customers so it need for adequate data structure classifier. Same in the case of Knowledge based filtering technique the knowledge acquisition id difficult. Collaborative filtering technique has the facility to provide unexpected recommendations, which means that it can recommend products that are appropriate to the customer even without the content being in the customer's profile [2]. However the collaborative filtering technique is successful but there are some problems exists as follows:

3.1.1 Cold-start problem– This refers to a condition where a recommender does not have sufficient information about a customer or a product in order to make appropriate forecast [26]. This is a big problem that diminishes the performance of our recommendation system. The profile of such new customer or product will be empty because that individual person has not rated any product, therefore his/her flavor is not known to the system.

3.1.2 Data sparsity problem- This problem occurs as a consequence of the lack of sufficient information that is when only a few of the total number of products available in a database are ranked by customers [27]. This problem all the time directs to a sparse customer-product matrix (CPM), inability to recognize successful neighbors and finally, it will generate weak recommendations. Data sparsity always leads to coverage problems, which is the percentage of products in the data base that recommendations can be prepared [28].

3.1.3 Scalability- This is another problem linked with recommendation methods because computation on average generates linearly with the number of customers and products [27]. A recommendation system that is competent when the amount of data set is restricted may be incapable to produce a suitable number of recommendations when the quantity of data set is increased. That's why it becomes essential to apply recommendation methods which are able of scaling up in a successful way as the number of data set in database increases. Some methods are there which solve the difficulty of scalability and speeding up recommendation generation are based on Dimensionality decrease technique, for example, Singular Value decomposition (SVD) method, which has the facility to produce reliable and competent recommendations.

3.1.4 Synonymy- This is the resemblance of the very similar products to have dissimilar names or entities. Many recommender systems find it is hard to make the difference between mostly related products such as the difference between baby clothes and baby wears. Collaborative Filtering systems find no match between these two products to be able to compute their similarity. There are many methods, such as automatic term extension. The formation of a vocabulary and Singular Value Decomposition (SVD), specially hidden Semantic Indexing are capable of solving the synonymy problem. The limitation of these techniques is that some supplementary expressions may have different meanings from what is proposed, which sometimes leads to immediate degradation of recommendation performance.

IV. OBJECTIVE OF STUDY

Collaboration filtering is a new and significant area of research. Following are the objectives of proposed work:

- To analyse the literature in the area of Collaboration Filtering Techniques.
- To investigate the outcome of various available recommendation algorithms.
- To identify the measures to improve the efficiency and effectiveness of the collaboration filtering technique.
- To propose an approach for collaboration filtering for recommendation algorithm.
- To compare the performance of proposed approach with existing approach.
- To evaluate the performance of proposed approach.

V. SCOPE OF STUDY

The recommender system evaluates the stored data as like or unlike data and determines a list of recommended products for the customer, the most prominent example of collaborative filtering is item-to-item collaborative filtering (i.e. customers who buy x also buy y), an algorithm popularized by Amazon.com's recommender system. Facebook, YouTube, MySpace, LinkedIn and other social networks use collaborative filtering to recommend new friends, society and other public links. Twitter uses lots of indication and remembrance techniques for recommending who to follow to its customers. An important feature of collaborative filtering technique is the capability to produce new adapted recommendations by evaluating information from the past activity of a particular customer, or the past of other customers considered to be of same taste to a known customer. This feature is used as customer profiling and supports the e-commerce website to recommend a customer according to their need [29].

Memory based collaborative filtering approach is consist of user based or item based top N recommendation. The Representative techniques are Neighbor-based collaborative filtering (NBCF) product-based/customer-based Collaborative Filtering algorithms with Pearson/vector cosine correlation. But there are some specific limitations of this approach as-

- Recommendations are dependent on human ratings.
- Recommendations performance decrease when data are sparse.
- Cannot recommend for new customers or products.
- It has limited scalability for large datasets.

Many recommender systems, in both educational and marketable locations, obtain ratings by customers openly state their choices for products. These likings are then used to calculate the customer likes for products they have not rated. From a system point of view, this is a straightforward technique and avoids potentially difficult deduction problems for customer choices.

VI. PROPOSED METHODOLOGY

Now these days, many research works have focused on collaborating technique. A real world application of collaboration approach to implement the proposed approach for collaborative filtering technique [30], the following methodology (in the form of Flow Chart) will be used:

Proposed algorithm:-

Step:1 Start
Step:2 Fetch data from database
Step:3 Generation of profiles
Step:4 Perform neighborhood selection
Step:5 Find the similarity between active user and other users by Euclidean distance (ED)

$$\text{euclidean (A, j)} = \sqrt{\sum_{i=1}^{z}\sum_{f=1}^{22} W_f * diff_{i,f}(A,j)^2}$$

Step:6 Provide rank to all profiles on the basis of ED
Step:7 If (ED>Threshold)
Then
Consider as most similar
Else
Dissimilar
Step:8 Create neighborhood of similar profiles of active user
Step:9 Apply Animal Migration optimization for best fitness value
Step:10 Evaluate fitness value of each user
Step:11 End

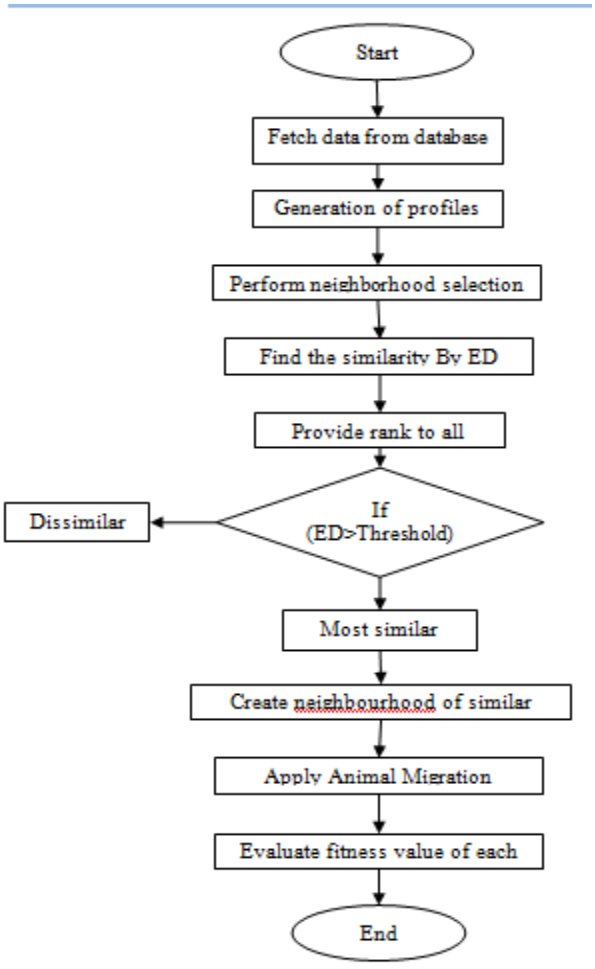

Fig 2: Flow Chart of proposed Algorithm

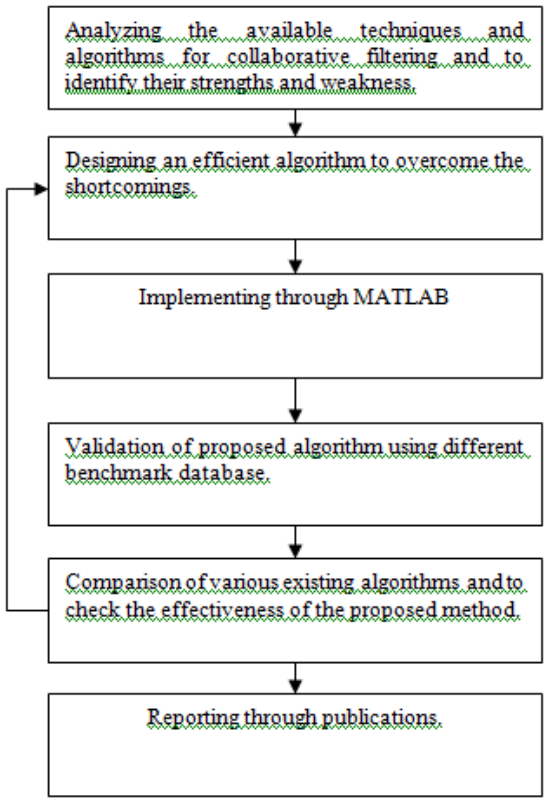

Fig 3: Approach of methodology

VII. ADVANTAGES OF AMO OVER PSO ALGORITHM

- AMO has the faster convergence speed and High convergence precision .
- PSO don't have the Genetic operators like crossover and mutation.
- AMO is suitable for the combinational problem.
- PSO is very poorly suitable for the combinational problem and particles update themselves with the interval velocity.
- PSO uses less number of functions so it is not efficient as compared to AMO.
- AMO is serial strategy.

VII. CONCLUSION

The Expected outcome of the proposed collaborative filtering technique will be an efficient system which will

1. Provide a collaborative filtering based novel, effective and capable recommendation system for the E-commerce domain.
2. Provide better recommendation to the customers on E-commerce website, with better accuracy and within less execution time.

The proposed technique is concerned with the study and analysis of collaborative filtering to improve the efficiency and usefulness of the solution for recommending products using collaborative filtering technique. The developed system is predictable to provide precious support to the customers in decision making at the time of choosing products at the E-commerce sites.

REFERENCES

[1]. Rashid A.M., Albert I., Cosley D., Lam S.K., McNee S.M., Konstan J.A. (2002). Getting to know you: learning new user preferences in recommender systems, Proc. of the international conference on intelligent user interfaces. 127-134.

[2]. Schafer J.B., Frankowski D., Herlocker J., Sen S. (2007). Collaborative filtering recommender systems. In: Brusilovsky P, Kobsa A, Nejdl W, (ed.) The adaptive Web, LNCS 4321, Berlin Heidelberg, Germany, pp 291-324.

[3]. Resnick P., Varian H.R. (1997) Recommender systems. Common ACM, 40(3), 56-58.

[4]. Acilar A.M., Arslan A. (2009). A collaborative filtering method based on Artificial Immune Network, Exp Syst Appl,36(4), 8324-32.

[5]. Jalali M., Mustapha N., Sulaiman M., Mamay A. (2010). WEBPUM: a web based recommendation system to predict user future movement. Exp Syst Applicat, 37(9), 6201-12.

[6]. Herlocker J.L., Konstan J.A., Terveen L.G., Riedl J.T. (2004). Evaluating collaborative filtering recommender systems. ACM Trans Inform Syst, 22(1), 5-53.

[7]. Celma O., Serra X. (2008). FOA Fing the music: bridging the semantic gap in music recommendation. Web Semant: Sci Srev Agent's World Wide Web, 16(4), 250-56.

[8]. Chen L.S., Hsu F.H., Chen M.C., Hsu Y.C. (2008). Developing recommender systems with the consideration of product profitability for sellers. International Journal Information Science, 178(4), 1032-48.

[9]. Ziegler C.N., Mc Nee S.M., Konstan J.A., Lausen G. (2005). Improving recommendation lists through topic diversification, Proc. of the 14th international conference on World Wide Web, 22-32.

[10]. Min S.H., Han I. (2010). Detection of the customer time-variant pattern for improving recommender system. Exp Syst Applicat, 37(4), 2911-22.

[11]. Celma O., Serra X. (2008). FOA Fing the music: bridging the semantic gap in music recommendation. Web Semant: Sci Srev Agent's World Wide Web, 16(4), 250-56.

[12]. Lieberman H., Letizia. (1995). An agent that assists web browsing. Proc. of the 1995 international joint conference on artificial intelligence, Montreal, Canada. 924-29.

[13]. Pazzani M.J. (1999). A framework for collaborative, content-based and demographic filtering. Artific Intell Rev, No. 5(6), 393-408.

[14]. Jennings A., Higuchi H. (1992). A personal news service based on a user model neural network. IEICE Trans Inform Syst, 75-D (2), 198-209.

[15]. Murat G., Sule G.O. (2010). Combination of web page recommender systems. Exp Syst Applicat, 37(4), 2911-922.

[16]. Al-Shamri M.Y.H., Bharadway K.K. (2008). Fuzzy-genetic approach to recommender systems based on a novel hybrid user model. Expert Syst Appl, 5(3), 1386-399.

[17]. Mican D., Tomai N. (2010). Association rule based recommender system for personalization in adaptive web-based applications, Proc. of the 10th international conference on current trends in web engineering (ICWE'10), Berlin, Springer-Verlag. 85-90.

[18]. Ghazantar M.A., Pragel-Benett A. (2010). 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.

[19]. Ziegler C.N., Lausen G., Schmidt-Thieme L. (2004). Taxonomy-driven computation of product recommendations, Proc. of the 13th international conference on information and knowledge management (CIKM'04), Washington, DC, USA. 406-15.

[20]. Sarwar B.M., Konstan J.A., Herlocker J.L., Miller B., Riedl J.T. (1998). Using filtering agents to improve prediction quality in the grouplens research, collaborative filtering

system, Proc of the ACM conference on computer supported cooperative work, NY (USA). ACM New York, 345-54.

[21]. Averjanova O., Ricci F., Nguyen Q. (2008). Map-based interaction with a conversational mobile recommender system. In: The Second International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies, UBICOMM '08, pp. 212–218.

[22]. Cunningham P., Bergmann R., Schmitt S., Traphoner R., Breen S., Smyth B. (2001). Websell: Intelligent sales assistants for the World Wide Web, Proc. Content Based Recommender (CBR) in E-commerce, Vancouver BC, 104-09.

[23]. Konstan I., Stathopoulos V., Jose J.M. (2009). Social networks and collaborative recommendation. Proc. of the 32nd international ACM conference (SIGIR'09), ACM, New York, NY, USA. 195-202.

[24]. Lee D.H., Brusilovsky P. (2010). Social networks and interest similarity: the case of CiteULike, Proc. of the 21st ACM conference on Hypertext and Hypermedia (HT'10), ACM, New York, NY, USA. 151-56.

[25]. Condiff M.K., Lewis D.D., Madigan D., Posse C. (1999). Bayesian mixed effects models for recommender systems, Proc of ACM SIGIR workshop of recommender systems: algorithm and evaluation, 12, 324-37.

[26]. Burke R. (2007). Hybrid web recommender systems. In: Brusilovsky P, Kobsa A, Nejdl W, (ed). The adaptive web, LNCS 4321. Berlin Heidelberg, Germany, Springer, pp 377-408.

[27]. Park D.H., Kim H.K., Choi I.Y., Kim J.K. (2012). A literature review and classification of recommender systems research. Expert Syst Appl, 39(11), 10059-0072.

[28]. Su X., Khoshgoftaar T.M. (2009). A survey of collaborative filtering techniques. Advanced Artificial Intelligence.12, 04-19.

[29]. Schafer J.B., Konstan J., Reidl J. (1999). Recommender system in e-commerce, Proc. 1st ACM conference on electronic commerce, 158-166.

[30]. Hatami, M., and Pashazadeh, S. (2014). Improving results and performance of collaborative filtering-based recommender systems using cuckoo optimization algorithm, International Journal of Computer Application, 88, 46-51.

