

# Investigation of Bi-Max Algorithm for On-Line Purchase Recommender System using Social Networks

B. Akshaya, S. K. Akshaya, S. Gayathri and P. Saravanan\*

School of Computing, SASTRAUniversity, Tirumalaisamudram, Thanjavur – 613401. Tamilnadu, India;akshayabadri@gmail.com,akshayask05@gmail.com, s.gayu95@gmail.com, sharan.doit@gmail.com

## Abstract

**Objectives:** Recommender systems in an E-commerce scenario, aim at improving a product's visibility to a customer. Existing recommendation systems incorporate traditional algorithms and have not been built to consider many behavioral patterns of the user. This opens a huge scope for improvement. **Methods/Statistical Analysis:** The proposed research work is set to an emphasis on the user's social network, location, history, patterns based on history (time series), relationship between users, similarity between users and similarity between the items that are in the subset of recommendations to be made. The raw data collected is first fed into a Bi-clustering algorithm called the Bi-Max. Then, Pearson's Coefficient is used to find the degree of similarity and filter out similar users based on a set threshold. Further filtering is done based on user networks and location of the user based on their latitudinal and longitudinal data obtained. **Findings:** The similar product has identified based on the degree of similarity. Similar users have been filtered out from the large set of users. It is much more likely that a given user will find a product that they would be interested in. **Applications/Improvements:** The proposed system has been used for the online purchase of any product. The performance of the proposed system was compared with the traditional approach of K-Means algorithm and the results show increased efficiency produced the recommendation by the proposed model.

**Keywords:** Bi-Max Algorithm, Online Purchase, Recommender System, Social Network, Time Series

## 1. Introduction

Internet nowadays is sprouting up with a variety of services; this is probably the consequence of its rapid expansion and use. Data related and collected from these services has become too enormous to be handled, processed and inferred from by the traditional mining and knowledge discovery approaches. Data Mining refers to the act of extracting useable information from a collection of data. This information can be applied for enhancing many fields. One such field is E-commerce. It is one of the most impactful trends in our contemporary world. There are many methodologies and systems being implemented to boost E-commerce. One such system is the recommender system. It aims at improving a product's visibility to a customer and accomplished by displaying a set of 'recommended items' exclusive to each user which

consists of items that they would most likely want to buy. These items are decided based on certain criteria that are set using the user's demographic data. Recommendation systems can be developed on many dimensions.

Usually recommendation systems focus on recommendations by the usage of collaborative filtering by considering the similar patterns among the other users and then combining it with the available usage statistics of the current user session. This methodology lacks the usage of the domain knowledge and its interpretations in the comparison of the usage patterns of other users while presenting the recommendations.

The next step down the processing lane is analyzing the history of a given user to improve prediction. Time series is the series of measurements of variables collected over time. Also, when dealing with real time scenarios like e-commerce, often it becomes impossible

\* Author for correspondence

to recommend items for new users because of the lack of data or information about them. This discrepancy is called the cold start problem. The proposed work also tries to address the cold start problem by using the user's demographic data, especially location. The focus of our project here also remains the location of the user.

The remaining sections in this manuscript have been organized as follows. Section 2 focuses on the work related to recommender systems and its challenging. The proposed work and algorithms are discussed in Section 3. The performance of the proposed method is evaluated in Section 4. Finally, Section 5 concludes the article with future work.

As a result of these discrepancies, a clustering based collaborative filtering approach is adapted for recommendation. It targets recruiting similar services among the elements of similar clusters to propose the actions collaboratively. By principle, these actions are approached in two stages. Dividing into clusters of small sizes and performing some collaboration filtering on the obtained clusters. Because each cluster most likely has less number of services than the entire dataset, the online run time of the approach is expected to decrease<sup>1</sup>.

Inclusion of contextual information to identify the suitable resources from a collection has gathered major research interest. This contextualization is observed as a model for constructing intelligent systems to predict the necessities of users. Also the system act much more efficiently in reply to their actions<sup>2</sup>. Another approach mainly focuses on including the domain knowledge and usage statistics knowledge in personalizing and comparing similar user patterns for recommendation systems<sup>3</sup>.

The multi-objective method of recommendation<sup>4</sup> uses location and social information for check-in recommendation. The features have been categorized into rate based, name based, weight based to predict the future check-in positions.

It is clustering approach for identifying subsets (mostly for genes and samples), in such a way that when one among these items is used as a measure to cluster the other, a set of stable and significant partitions are obtained. This method preferably uses an iterative clustering for more efficient search and analysis. By picking out relevant subsets of the data and shedding some focus on these selected subsets, many partitions and coefficients of correlations are found which were unnoticeable and

insignificant when the entire data set was used for the analysis<sup>5</sup>.

Being one of the most thriving techniques for recommender systems, Collaborative filtering advises items on the basis of the nearest neighbors from the target user. Hence, the performance of this recommendation system depends majorly on the technique implemented for similarity measures. Thus, two different user-preference models are implemented to display the user interest in and to focus the average ratings a customer has credited to different genre. The results of the new similarity measure, out performs the traditional similarity measures<sup>6</sup>. Recommendation systems that apply collaborative filtering use databases about user preferences that predicts the products or items that new user might like.

Collaborative filtering fails when it comes to data scarcity problems, because collaborative is a very traditional and inefficient method and that is proved inefficient in the paper. So, to tackle this, trust based recommender systems can be used that improves the effectiveness of the recommendation. Such a system is built based on social relation and user's profile that contains the trust information. So, both the social information and trust factors are considered which increases the efficiency. Tags are the keywords that are a component of the browsing history. Tags are formed based on a particular user's browsing history, and a user-item matrix based on these tags is formed. These tags can be used as a feature and thus similarity between users and similarity between items can be formed. The system is then compared with 2 other traditional methods. So, finally this method delivers an effective performance compared to other methods<sup>7</sup>.

Collaborative Filtering can promote items and products of relevance to a target buyer from an enormous collection of items available. It can be categorized widely based on the algorithm adapted for its implementation into two broad categories. They are memory based algorithms and model based algorithms. Cross domain recommendation issues have been resolved using collaborative filtering<sup>8</sup>.

Model based algorithms consume a certain amount of time to be built as a model but they tend to give recommendation for online items quickly. Memory based algorithms on the other hand are time-consuming but do not require as much pre-building time as the model

based algorithms but they are much less effective than them. On looking at the draw backs and the attributes of these two categories of algorithms, a novel algorithm is proposed so as to try to overcome these short comings. It is called Community-Based User domain Collaborative Recommendation Algorithm (CUCRA)<sup>9</sup>. This idea comes from the principle that recommendations are to be made among the users with similar preferences. So, the first step in the processing is to construct a User-User social network depending on the users' preference data. Second step of the processing involves finding communities that have similar user preferences by the usage of a community detecting algorithm. Furthermore, the items are recommended to potential buyers or the users by the application of collaborative filtering on these communities<sup>10</sup>.

Since items are recommended to users in derived communities instead of recommending to the entire existing social network, this method has profound online performance. The community driven method is applied to a collaborative tagging and filtering system. The experimental results show that the recommendation efficiency is relatively good, and the online time-complexity reduces<sup>11</sup>.

As more people are flocking towards the side of excessive socializing through online networks, they end up sharing their opinions excessively on these forums. Also, this increases their impact factor on the people who they have connected to, from these online forums. This bring about a major challenge for the existing traditional recommender systems which have to strive to solve the cold start issues for users with very less social information and the information and they should also work at solving data dispersion problem of datasets<sup>12</sup>.

Traditionally the selection of content in a scenario must be done manually. Recent implementations as a consequence of growth and expansion of recommender systems, programmed accumulation of content based social data warehouse from social media and networks. Social ties refer to the friendships of various users that result in some kind of networks and sub-networks. This data when used to design the system can result in a system that affects the degree of affectation of the user with respect to the recommendation.

There is no evidence to prove that such friendships would affect the recommendation largely. A positive effect is there on the value of the recommendation

based on structural equation model results. The value of recommendation depends on the standing of the user recommending and the media source effect<sup>13</sup>.

## 2. Proposed Work:

The sales information are recorded by every user over time and predict the product the user most likely would buy based on a similarity analysis between items (item-item similarity). Time series data are usually large in size, they are high dimensional and required to be updated constantly.

The information stacked up in databases of virtual networks happens to be outsized, complex and mixed in nature. Hence, both competent and sharp processing has been required in their mining. Also, the analysis of user-user interaction information and exchange during time oriented social activities. This paper, presents a bi-clustering algorithm for handling large logs of data sets that record the online activities of members of a virtual community. This approach is useful as it is aimed to extract as much relevant knowledge as possible about user activities such as navigation and behavior patterns, activities performed along with the parameters related to such activities.

It is also tied to integrate three major social factors namely a person's interest towards a service or item, similarity between persons and their inclination towards a service or product, and the impact factor of a user. The probabilistic matrix factorization has been used for recommendations. The first factor works at making the recommender system to recommend the items as per the tastes, especially for already existing users. Furthermore, for cold start users, the impact factor will increase the essential association among aspects in the covert space. By comparing this system with already implemented traditional approaches, this approach seems to increase the working efficiency of a recommender system<sup>14</sup>.

The system introduces community based social interactions as a new dimension for recommending products to a potential buyer or target and it also present a social recommendation system using collaborative filtering and community detection methodologies. This is successfully done using two steps

- A community detection algorithm to analyze the friendship relations among users by analyzing user-user social graph.

- User-item based collaborative filtering for rating prediction.

These approaches were developed using map-reduce functions.

This greatly increases scalability, coverage and handles the cold start issue of recommender systems that are based on collaborative filtering. The resulting system is then compared with the performance of the traditional collaborative filtering based recommendation system. It is then concluded to increase the effectiveness of the recommendations.

User ratings not only include abundant data for learning user preferences, but also textual reviews following the ratings. But it so happens that most existing recommendation systems take rating credentials for granted and thus ignore the high quality of information that can be extracted from them in combination with the reviews. In this paper, in order to exploit user profiles' information embedded in both ratings and reviews exhaustively, a Bayesian model is proposed that links a traditional Collaborative Filtering (CF) technique with a topic models iteratively. By employing a topic model with the review text and aligning user review topics with "user attitudes" over the same distribution, this method achieves greater accuracy than the traditional approach on the rating prediction task.

Also, with review text information involved, latent user rating attitudes are interpretable and "cold-start" problem can be alleviated. This property qualifies our method for serving as a "recommender" task with very sparse datasets<sup>14</sup>.

Rapid expansion of E-commerce has made it very easy for the many users to share their opinions and views on the many services and item available in these websites or applications. Processing this amount of assorted and wide-ranging information leaves the users in great confusion on how to come to a conclusion about the relevance and their need for a product or service. Usage of spatial, temporal and social information of the users that are available on these websites has become an integral part of most approaches implemented for the knowledge discovery process in the field. And this helps build a better recommendation system which is efficient in handling excessive data. In this work, a position and network attentive recommender system improved with multi intention filtering is projected and implemented. The results tend to support the proposed system as being more objective and efficient.

### 3.1 Clustering the Data Collected

Clustering is invariably the most widely accepted first step to process raw data among all mining methodologies. It involves grouping data into clusters (a bundle of similar data) such that the data elements belonging to the same cluster are much for alike in behavior than the elements present in another cluster. It is basically an iterative knowledge discovery process. Clustering can be implemented by many different algorithms all of which focus on one or more attributes of the data collected. Its fundamental ideology is that forming clusters of users using an efficient algorithm, finding the degree of similarity between these users and further filtering recommendations based on location, user networks, time series behavior pattern etc can build a strong system. A combined framework has been developed to observe the increase in quality of recommendations<sup>15</sup>.

Bi-clustering<sup>16</sup> is one among the most recently proposed clustering methodologies. It proposes to cluster rows and columns of a matrix simultaneously. For a given input matrix (m x n) it generates "bi-clusters", a subset of rows which exhibit similar behavior across a subset of columns, or vice versa. Usually, most clustering algorithms depend on global similarity of rows and columns of a data matrix. But, the similarity of the data may sometimes be restricted to certain experimental conditions. Many proposals have used cosine similarity<sup>17</sup>. In our proposal, the data can be highly dimensional and dynamic. Bi-clustering algorithms focus on solving such issues. Goal of a successful bi-clustering algorithm is to identify "homogeneous" sub-matrices.

The algorithm chosen here to implement the bi-clustering is called the Bi-Max algorithm. The basic requirement for a Bi-Max algorithm to work is that the input matrix should be a binary matrix. So for the numeric data, a threshold has been set based on requirement and converts the numeric values to binary depending on whether they cross or do not cross the threshold. The basic intention of the BIMAX algorithm is to find all bi-clusters consisting of 1s in the binary matrix. Bi-Max uses a recursive divide and conquers strategy to enumerate all the bi-clusters in a matrix.

From the input, a binary matrix M, using threshold values has been created. Set of rows and set of columns R and C respectively have created with

$$R \in 2^{\{1, \dots, m\}} \text{ and } C \in 2^{\{1, \dots, n\}}$$

such that,

$$\forall i \in R, \forall j \in C : M[i, j] = 1$$

For other  $(R', C')$  that meets the first condition,  $(R \subseteq R' \wedge C \subseteq C') \Rightarrow (R = R' \wedge C = C')$

Divide-Conquer method has been used for clustering in Bi-Max algorithm. Figure 1 considers input data sets consists of 0s and 1s. Then each row has applied the divide and conquers strategy to form sub-matrices and further each sub-matrix processed independently. To find the sub-matrix of row chosen  $r^*$ , column  $C = \{1..n\}$  are partitioned into set of 1s and set of 0s as follows.

$$C_U = \{ c : M[r^*, c] = 1 \}$$

$$C_V = C - C_U$$

Then, the set of rows partitioned into three more sets.

$R_U$  : Set of rows have 1s in only  $C_U$

$R_W$  : Set of rows have 1s in  $C_U$  and  $C_V$

$R_V$  : Set of rows have 1s in only  $C_V$

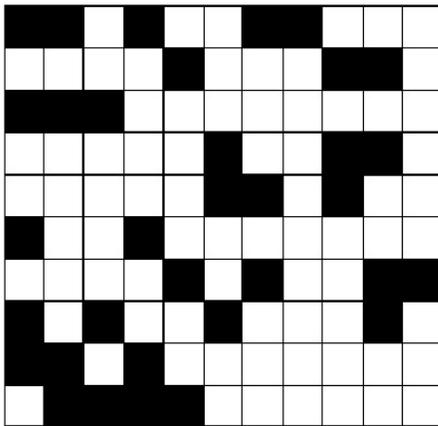


Figure 1. Distribution of Success Value in Binary Matrix.

After applying the technique the matrix elements are rearranged as showed in Figure 2.

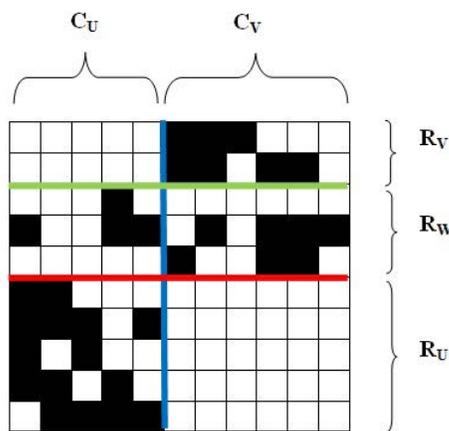


Figure 2. Distribution of Success Value in Biclustering.

### 3.2 Similarity Measurement

Another important methodology used to process data is finding the similarity between focused aspects of the data. In this, one or more set of characters or attributes are chosen from the data and used as a measure to find similarity. This can be done by calculating Pearson's coefficient, as follows:

For any two consumers A and B, the ratings on the selected product k are noted by  $A_k$  and  $B_k$ . Then the correlation between A and B is calculated by,

$$r(A, B) = \frac{\sum_k (A_k - \bar{A})(B_k - \bar{B})}{\sqrt{\sum_k (A_k - \bar{A})^2 (B_k - \bar{B})^2}}$$

Where  $\bar{A}$  is the mean value of ratings by consumer A and

$\bar{B}$  is the mean value of ratings by consumer B.

Based the previous ratings the prediction is calculated as follows for any consumer X on the item i.

$$P(A_i) = \frac{\sum_r B_i - r(A, B)}{n}$$

Similarity can be performed as a step succeeding clustering. The clusters can be analyzed and cross analyzed to filter out data which have high correlation coefficient between them. This process is called collaborative filtering.

There are many various approaches to finding similarity. The wide spread use of Pearson's coefficient as a measure is due to its balance between computational complexity and statistical significance. It takes into account all the data pertaining to the selected attribute of every user and is still not a very time consuming calculation.

The value of the correlation coefficient usual varies -1 to 1. If the value of the coefficient is -1 it means the two sets of data are negatively correlated, that is when one increases, the other decreases. If the value of the coefficient is 1 it means the two sets of data are positively correlated, that is when one increases, the other also increases. If the value of the coefficient is zero then it means that the two sets of data are independent. Any value between these boundaries show the degree to which the can be negatively or positively correlated.

### 3.3 Social-Network Based Filtering

It so happens that in most situations users take the opinions of their own friends and acquaintances more seriously than other users they haven't met before. This method tries to take advantage of that factor. Social networks

are often formed due to similarities in taste, or some common existing ground between two or more users. This can lead to fact that they have similar requirements or are more likely to be interested in same products. This can be done using simple filtering techniques like comparing the output of the previous step to the already existing friendship matrix that has been formed from the collected data. This can increase the precision of the prediction made.

### 3.4 Location Based Filtering

Often recommendations are made based on a given users location. This enhances the quality of the recommendations made. This is done to avoid meaningless situations like recommending a Tamil movie to someone in Kerala or recommending a sweatshirt to someone who lives in Chennai. Such recommendations bring down to product interest rates between two users. Location filtering can be made by determining the latitudinal and longitudinal data of a give set of users and elimination them based of different criteria depending on the kind of product.

For example, in movies, the distances among different parts of a country even neighboring states have to consider. But for recommending clothes it's almost accurate to just see if they are the same country. Location based filtering is thus essential in most recommender systems but is often not incorporated. For distance value has been calculated using the latitudinal and longitudinal information as follows.

Longitude difference  $dlon = lon_2 - lon_1$

Latitude difference  $dlat = lat_2 - lat_1$

Where  $lon_1$  and  $lat_1$  are the longitude and latitude of first location respectively and  $lon_2$  and  $lat_2$  are the longitude and latitude of second location respectively

Then the distance  $d = R * c$

Where,

$$a = \left(\sin\left[\left(\frac{dlat}{2}\right)\right]\right)^2 + \cos(lat_1) * \cos(lat_2) * \left(\sin\left[\left(\frac{dlon}{2}\right)\right]\right)^2$$

$$c = 2 * a \tan^2\left(\sqrt{a}, \sqrt{1-a}\right)$$

### 3.5 Implementing Time-Series Analysis

Often excess filtering of possible recommendations can lead to unavailability of quality recommendation. This can be eliminated by considering the target user's own product interest pattern and how it varies over a specific period of time. This kind of pattern recognition is called

time series analysis and this is used to enhance prediction from the point of user's history. Short time series analysis has been identified<sup>18</sup> as dynamic measurement and statistical predictions are not allowed. Efficient Dynamic Time Warping (EDTW)<sup>19</sup> has been proposed to measure the similarity which efforts in linear time. Time series also used in prediction based system. The expression value of gene can be predicted using time series analysis<sup>20</sup>. Similarity measurements also calculated using time series analysis based on clustering<sup>21</sup>.

In this work, the history of the user's most similar friend has considered and the maximum rated item has identified. Once the friend's favorite item is found, an item-item similarity matrix has been build, which analysis the similarity between different items based on tags. Then the similar items are predicted and ensured the recommended display of good variety.

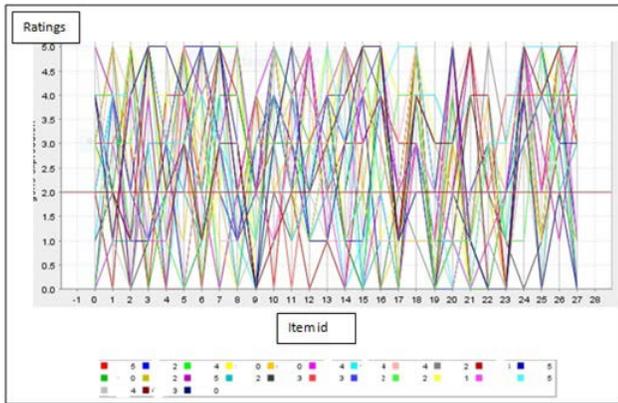
For the entire analysis process a large amount of has been stored to access the user's history. It has required recording the value of certain variables, or a fixed variable over a period of time. Mathematical or statistical method like regression has been used to predict the possible future value. This has been done to avoid stale predictions or recommendations.

### 3.6 Addressing Cold-Start Issue

When a new user joins to an existing network, they might not have any friends or history to make sensible predictions. This issue is called cold-start problem. It can be addressed by using an algorithm Modified Intuitionistic Possibilistic Fuzzy Geographically Weighted Clustering-Cold Start (MIPFGWC-CS)<sup>21</sup>. This algorithm uses the user's demographic information to find similar users and thus comes up with recommendations.

## 4. Results and Discussion

K-means is a greedy search method to cluster one dimensional data effectively. It is not very effective in case of large 2 dimensional data. The current graph in Figure 3 has worked on a 3600, 2D dimension data in which the rows of the matrix represent users and the columns, the ratings. As one can see from the graph, k-means results in a large number of small clusters. Moreover, for a large data k-means cannot bring about just 2 clusters.



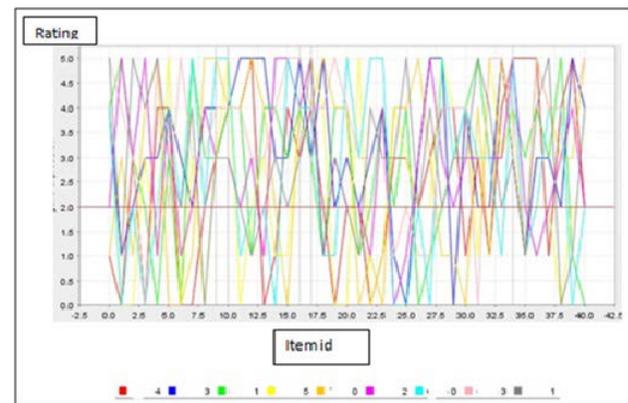
**Figure 3.** Ranking Distribution of Items using K-Means Algorithm.

The various colour codes show it effectively. The main aim of clustering is to group similar data together, and in this case, it refers to group similar users based on rating information on various items. Moreover k-means uses Euclidean distance for similarity measurement and grouping of the data for the cluster formation.

Bi-clustering is a method in which input is taken in the form of a matrix and similar sub-matrices also, known as bi-clusters are formed. Unlike a normal clustering approach like that of k-means, bi-clustering has acquired its name from the fact that, the rows and columns are clustered. In the research work, a bi-max bi-clustering algorithm has been adopted that aims at finding the maximum bi-clusters so that a large group of similar users can be identified. Figure 4 shows the results of bi-max algorithm and the color code clearly segregates the various bi-clusters formed. In this, case nine bi-clusters have been formed. The work has made use of bi-max bi-clustering algorithm for clustering the dataset which is present in the form Figure 1 dimensional matrix, in which the rows represent user id and the columns represent the various item id. The reason for adopting this bi-clustering approach instead of a k-means approach is that, k-means cannot effectively with a large number of dataset and moreover it results in a large number of small clusters. It is tedious to work with such large clusters.

On the other hand, Bi-max bi-clustering algorithm produces effective bi-clusters or bi-matrices taking the dataset as input. The graph clearly shows the demarking. The k-means expression graph produces twenty-five clusters, whereas the bi-max algorithm provides only nine bi-clusters. So, the comparison is based on the fact

that the number of clusters or bi-clusters determines the easiness to work with them.



**Figure 4.** Ranking Distribution of Items using Bi-Max Algorithm.

## 5. Summary and Conclusion

The research work aims at giving an optimized recommendation by filtering the users at each stage and zeroing in on a particular set of items. To achieve these objectives, similarity finding, the social network-based filtering, combining time series information with item-item correlation and clustering using Bi-Max algorithm have been implemented. The stages have been applied to a large dataset, and the results exhibited that Bi-Max algorithm ranks the product with the highest accuracy using social network information compared to traditional recommender systems. In this research, the cold start problem has also been addressed and solved it using similarity and location of the users.

## 6. References

1. Verbert K, Manouselis N, Ochoa X, Wolpers M, Drachsler H, Bosnic I, Duval E. Context-aware recommender systems for learning: a survey and future challenges. *Learning Technologies, IEEE Transactions*. 2012 Oct; 5(4):318–35.
2. Hu R, Dou W, Liu J. ClubCF: A clustering-based collaborative filtering approach for big data application. *Emerging topics in computing, IEEE Transactions*. 2014 Sep; 2(3):302–13.
3. Venu Gopalachari M, Sammulal P. Personalized collaborative filtering recommender system using domain knowledge. In *Computer and Communications Technologies (ICCCCT), International Conference, IEEE; 2014 Dec 11*. p.1–6.

4. Ozsoy MG, Polat F, Alhadj R. Multi-objective optimization based location and social network aware recommendation. In collaborative computing: Networking, applications and work sharing (CollaborateCom), International Conference, IEEE; 2014 Oct 22. p. 233–42.
5. Xhafa F, Caballé S, Barolli L, Molina A, Miho R. Using bi-clustering algorithm for analyzing online users activity in a virtual campus. In Intelligent Networking and Collaborative Systems (INCOS), 2nd International Conference, IEEE; 2010 Nov 24. p. 214–21.
6. Cheng Q, Wang X, Yin D, Niu Y, Xiang X, Yang J, Shen L. The new similarity measure based on user preference models for collaborative filtering. In information and automation, IEEE International Conference; 2015 Aug 8. p. 577–82.
7. Jin J, Chen Q. A trust-based Top-K recommender system using social tagging network. In Fuzzy Systems and Knowledge Discovery (FSKD), 9th International Conference, IEEE; 2012 May 29. p. 1270–74.
8. Vinayak S, Sharma R, Singh R. MOVBOOK: A personalized social network based cross domain recommender system. Indian Journal of Science and Technology. 2016 Aug; 9(31):1–10.
9. Qian F, Zhang Y, Zhang Y, Duan Z. Community-based user domain model collaborative recommendation algorithm. Tsinghua Science and Technology. 2013 Aug; 18(4):353–9.
10. Qian X, Feng H, Zhao G, Mei T. Personalized recommendation combining user interest and social circle. Knowledge and data Engineering, IEEE Transactions. 2014 Jul; 26(7). p. 1763–77.
11. Lalwani D, Somayajulu DV, Krishna PR. A community driven social recommendation system. In Big Data (Big Data), IEEE International Conference; 2015 Oct 29. p. 821–26.
12. Oechslein O, Hess T. The value of a recommendation: The role of social ties in social recommender systems. In System Sciences (HICSS). 47th Hawaii International Conference, IEEE; 2014 Jan 6. p. 1864–73.
13. Jiang M, Song D, Liao L, Zhu F. A Bayesian recommender model for user rating and review profiling. Tsinghua Science and Technology. 2015 Dec; 20(6):634–43.
14. Son LH. Dealing with the new user cold-start problem in recommender systems: A comparative review. Information Systems; 2014. p.1–10.
15. Jin J, Chen Q. A trust-based Top-K recommender system using social tagging network. In Fuzzy Systems and Knowledge Discovery (FSKD). 9th International Conference, IEEE; 2012 May 29. p.1270–74..
16. Prelic A, Bleuler S, Zimmermann P, Wille A, Bühlmann P, Gruissem W, Hennig L, Thiele L, Zitzler E. A systematic comparison and evaluation of biclustering methods for gene expression data. Bioinformatics. 2006 May 1; 22(9):1122–9.
17. Bigdeli E, Bahmani Z. Comparing accuracy of cosine-based similarity and correlation-based similarity algorithms in tourism recommender systems. In Management of Innovation and Technology. ICMIT 2008. 4th IEEE International Conference; 2008 Sep 21. p. 469–74.
18. Kopytov VV, Petrenko VI, Tebueva FB, Streblianskaia NV. An improved Brown's method applying fractal dimension to forecast the load in a computing cluster for short time series. Indian Journal of Science and Technology. 2016 May; 9(19):1–9.
19. Vasimalla K, Challa N, Naik SM. Efficient dynamic time warping for time series classification. Indian Journal of Science and Technology. 2016 Jun; 9(21):1–7.
20. Yoon HJ, Wang BH, Lim JS. Prediction of time series microarray data using neurofuzzy networks. Indian Journal of Science and Technology. 2015 Oct; 8(26):1–5.
21. Devi DMR, Thambidurai P. Similarity measurement in recent biased time series databases using different clustering methods. Indian Journal of Science and Technology. 2014 Jan; 7(2):1–10.