ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645
Indian Journal of Science and Technology, *Vol 11(22), DOI: 10.17485/ijst/2018/v11i22/116392, June 2018*

Collaborative Filtering Recommender System (CFRS): Comparative Survey on Cold-Start Issue

S. Vairachilai*

Department of Computer Science and Engineering, Faculty of Science and Technology,
The ICFAI Foundation for Higher Education (IFHE), Hyderabad – 501203, Telangana, India;
vairachilai@ifheindia.org

Abstract

Objectives: To analyze the issue of cold-start (user cold-start and item cold-start) in Collaborative Filtering Recommender System (CFRS) and to compare its solution with various approaches are summarized in this paper. **Methods/Statistical Analysis:** The manuscript discussed about the cold-start issue in which the recommender system cannot recommend items to the new user because no ratings made by the new user (user cold-start) as well as for the newly added items, the system cannot be able to provide recommendations to the user because the system has no ratings for the newly added item (item cold-start). The solutions for cold-start issue are analyzed based on the model based approach, demographic data, ask-to-rate technique, and Social Network Analysis (SNA). **Findings**: The comparative review of the aforementioned approaches provides the detail about how to implement the model based approach, how to collect the demographic data from the new user, how to apply the ask-to-rate technique and how to make use of the SNA concept to solve the cold-start issue in CF recommender system. **Application**: The recommender system on Amazon helps the user to purchase books, Compact Disks (CDs), Netflix helps the user to choose CDs to purchase/rent and Epinions, helps the users to decide to purchase based on user reviews.

Keywords: Ask-To-Rate Technique, Cold-Start, Collaborative Filtering Recommender System, Demographic Data, Model Based Approach, Social Network Analysis

1. Introduction

With an increasing market size, electronic commerce is a driving force for a business to enable a firm or individual for online shopping or marketing over an electronic network, typically the internet. Because it reduces the transaction cost, low energy cost and provides access to the global market. At the same time, the explosive growth of the Internet has increased the volume of information. Thus, e-commerce applications to confront the information overload problem in which users are finding the right information at the right time is difficult. From the business point of view, recommender systems have the potential to solve the information overload problem. The recommender system task is to recommend items and also help the user in selecting/purchasing items from an overwhelming set of choices. The collaborative filtering recommender system is one of the most successful approaches in an e-commerce website. Collaborative filtering is a method that provides personalized recommendations, based on preferences expressed by a set of users and calculates the similarity between customer preference ratings to identify like-minded customers and predict their product preferences.

Although the collaborative filtering recommender system successful, it undergoes a major issue such as

**Author for correspondence*

cold-start problem. The enough information is not available for a new item or user, the recommender system suffers into the item/user cold-start problem. The new item cold-start problem take place when the item is new and it does not have any ratings. The new item cannot be recommended unless a user has rated it before. The new user cold-start problem take place when the user is new or user has not yet rated enough number of items. Thus, the system cannot gather the information about new users and the system cannot recommend any item to the new user. To increase the usefulness of collaborative recommender systems, it could be desirable to eliminate the cold start problem.

The paper is organized as follows: The model based solution to the cold-start problem is explained in section.

2. The demographic data based solution to the cold-start problem is explained in section 3. Theask-to-rate technique based solution to the cold-start problem is explained in section 4. The social network analysis based solution to the cold-start problem is explained in section 5. Finally, the conclusion and future works are explained in section 6.

2. Solution to Cold-Start Problem - Model Based Approach

Model Based Recommender System: In Model-based CF technique, first build a model based on the rating database. Then, the model generates a recommendation without using the complete database every time. The author solved new user cold-start problem based on SCOAL (Simultaneous Co-Clustering and Learning) algorithm[1]. This method identifies the closest cluster for each new user based on demographic information. They used the MovieLensdataset to evaluate the performance of this approach. The recommendation process is shown in Figure 1.

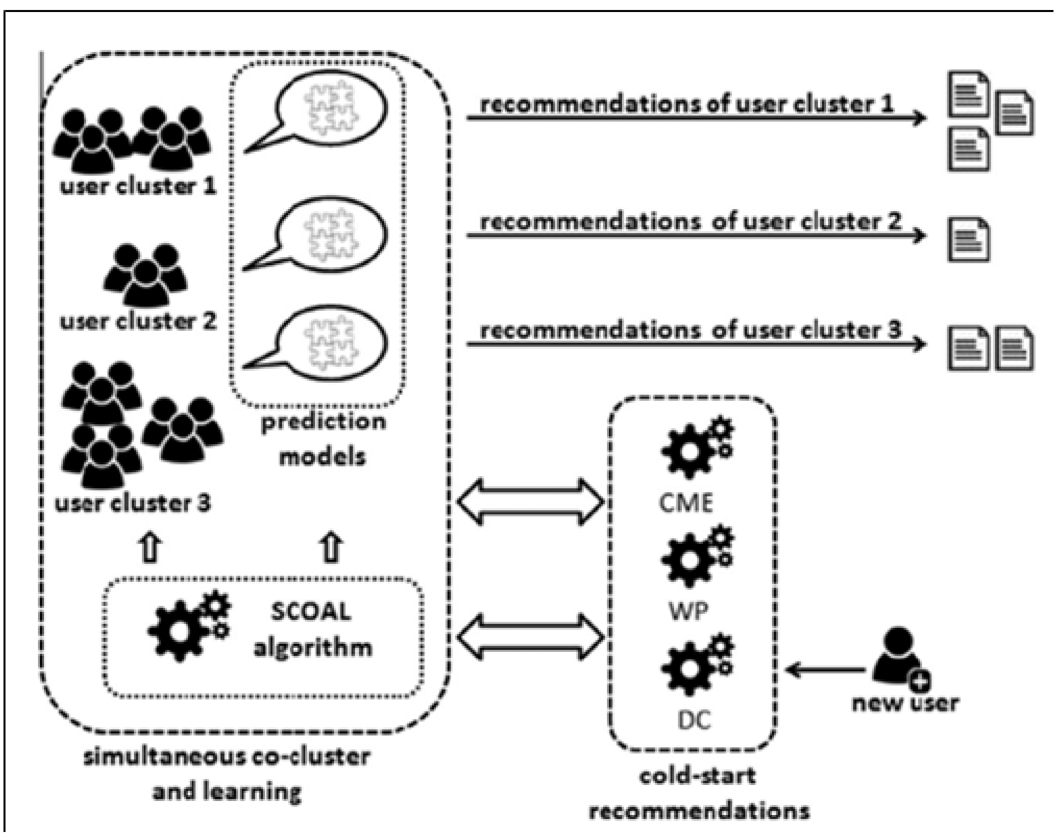

Figure 1. SCOAL algorithm based approach to the cold - start problem[1].

Recommendation Process

Step 1. SCOAL algorithm, cluster the users based on the common interests (demographic information).

Step 2. Build a separate prediction model for each cluster.

Step 3. Assign the new user into an appropriate cluster based on demographic information.

Step 4. Calculate the similarity between the new user and each of its neighbors.

Step 5. Recommend the top rated items to the new user.

The author proposed a solution to the user cold-start problem based on the weighted graph model in a social network using the concept of critical node[2,3].

Recommendation Process

Step 1. Identify the sets of most important critical nodes based on network efficiency and trust indices.

Step 2. Specify the conditions under which critical nodes should be controlled using connected components.

Step 3. Extracts set of most powerful articulation points which are able to play a critical node's role.

Step 4. Effective critical nodes that ensure the best community representation and management of the new user in the system.

The author proposed dynamic browsing tree model to solve the new item problem in e-commerce website based on additional information such as browsing time, a number of clicks, and interaction with users and so on. In this model, an item category tree is a reverse tree which composed with all items and their categories and an item is called a leaf in the tree[4]. The fresh degree indicates the weight value of the leaf. If the user clicks the certain item or browses certain items, the RS creates a link between the root of an item category tree and the leaf of that item. Then, the fresh degree of the leaf will be updated. The Interest Matching Degree (IMD) refers to the matching degree between new item and a user's interest. The sample

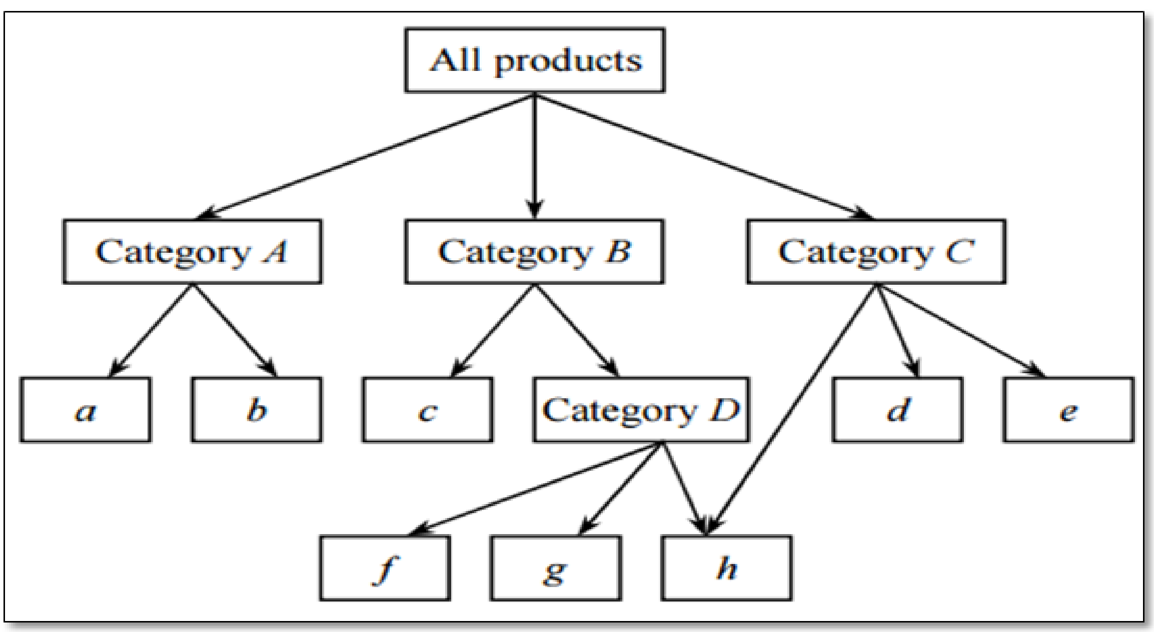

Figure 2. Item category tree[4].

item category tree is shown in Figure 2. In the Figure 2 a-h represent the items that belong to the particular type of item category.

Recommendation Process

Step 1. Transform the user browsing records into Dynamic Browsing Tree (DBT) based on the product category.

Step 2. Design a degree decay operatorwhich is based on user access time.

Step 3. Compute the category similarity between new item and every leaf of DBT.

Step 4. Compute the matching degree between new item and dynamic browsing trees of all users based on IMD.

Step 5. Recommend the new item to the users who have higher IMD than designated threshold value.

3. Solution to Cold-Start Problem - Demographic Data Based Approach

Demographic Recommender Systems: The demographic recommender systems are based on demographic characteristics of customers. In demographic recommender systems, the users are classified based on their individual attribute. The RS recommends a list of items that have good feedback from the customers that are demographically similar to the target customers. This approach is preferred when user rating history is not available for generating recommendations. This approach is fast, simple and straight forward to make results based on a few observations.

The author proposed a solution to the user cold-start problem based on the demographic information stored in the user profiles[5]. The author analyzed the demographic attributes of customers and found that a set the attributes can produce more accurate recommendations. The

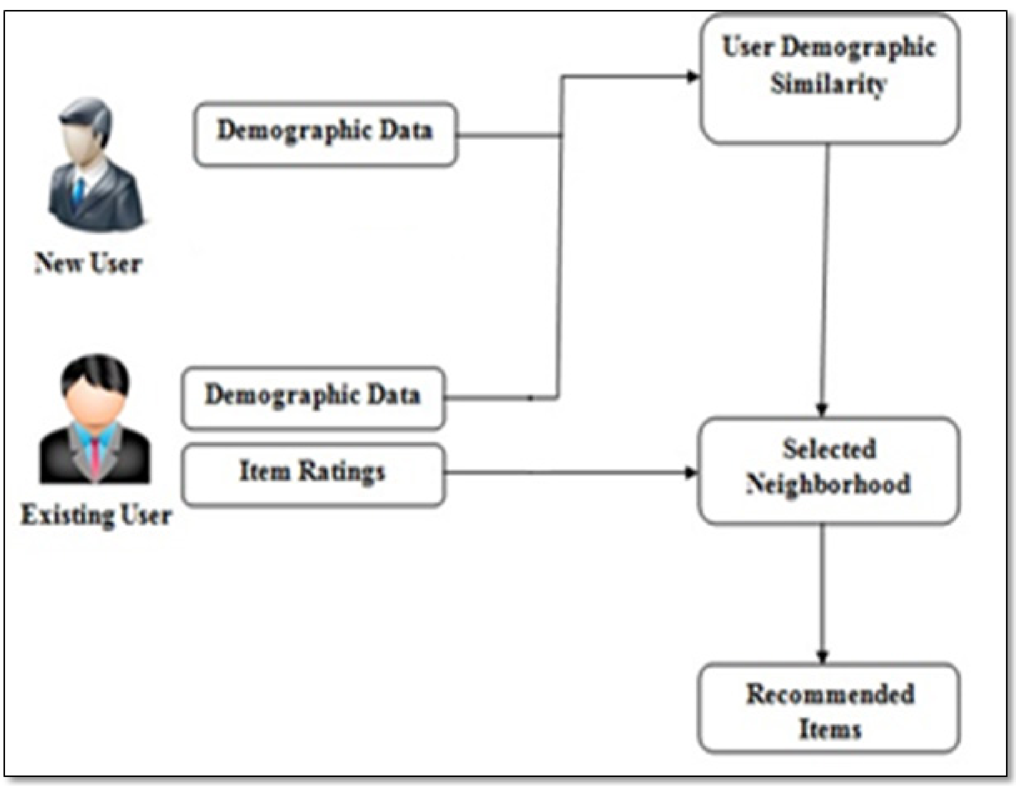

Figure 3. Demographic data based approach[5].

underlying assumption of this method is the users with similar demographic attribute(s) will rate items similarly. The demographic based recommendation process is shown in Figure 3.

Recommendation Process

Step 1. Collect the demographic data explicitly from the user such as age, gender, location, occupation, religion, marital status, hobbies, and country.

Step 2. Calculate the similarity between the existing user and the new user based on their demographic attributes. The RS form a neighborhood user based on the similarity value.

Step 3. Recommend the items to the new user based on a demographic attribute(s) based similarity value.

In addition, this demographic recommender system validates the recommendations by checking the following conditions:

1. The new user age is less than or equal to nine years old.
2. More than one user occupation attributes as none or other.
3. The attribute value is highly sparse such as zip code attribute.

If the aforementioned conditions exist in the recommender system then it will not be able to offer recommendationsto the newuser. The author used the Movielens dataset to evaluate the performance of this approach.

The author recommended regression model to solve the user and the item cold-start problem based on legally accessible demographic information such as age, gender, residence and occupation and information about items such as product name, cost, manufacturer, genre, production and year[6]. In this approach, RS learns the user-item and item-item affinity and uses it for predictions. In this approach, 50% of the users are selected as new users and 50% of existing users. Similarly, 50% of the items are selected as new and 50 % items are existing items. The data partition for user and item is shown in Figure 4. The author used two standard movie datasets such as Movielens and EachMovie to evaluate the performance of this approach.

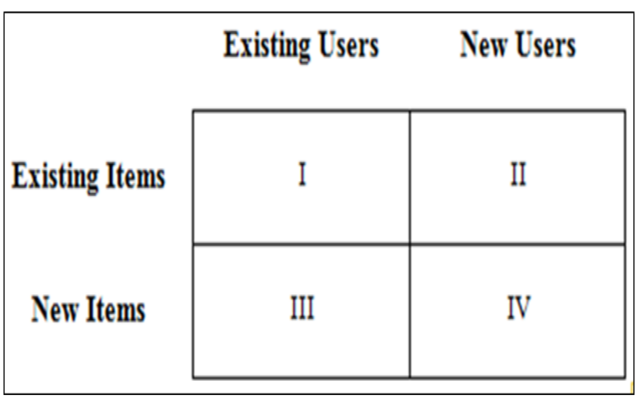

Figure 4. Data partition[6].

Recommendation Process

Step 1. Split the user rating into four partitions: 1. New Users, 2. New Items, 3. Existing Users, and 4. Existing Items. The Partition I is the training data set and Partition 2, 3 and 4 are tested dataset.

Step 2. Collect the demographic based information of the user and the item.

Step 3. Predictive feature-based regression model learns the user-item and item-item affinity and uses it for predictions.

Step 4. Build a predictive model for user/item pairs by leveraging all the available information on users and items to solve the cold-start problem in a recommender system.

The author used the collaborate filtering algorithm named as Number of Common Terms/Term Frequency (NCT/TF) to solve the user cold start problem in recommender systems[7]. The basic assumption of this approach is, if two people demographic attributes are similar, then they may have a common interest and prefer the same types of items. They used two standard movie datasets such as Movielens and EachMovie to evaluate the performance of this approach.

Recommendation Process

Step 1. Calculate the demographic vector based on user information such age, sex, occupation, and self-description during the registration process and create the user profile.

Step 2. Generate users attribute keywords set from user's registered information.

The user information can be classified into two:

1. *Specific information:* It includes user's occupation, age, marital status, hobbies, country, income level, gender, job and so on.
2. *Vague Information:* It includes a self-introduction and self-evaluation.

Step 3. Calculate the user's attribute similarity on demographic data between two users based on the common keywords and its weights.

Step 4. Generate a new similarity measure by combining the traditional similarity measures such as Cosine Correlation Coefficient (CCC) and Pearson Correlation Coefficient (PCC).

Step 5. Recommend the list of items to the new user from the similar neighbours.

4. Solution to Cold-Start Problem: Ask-to-Rate Technique Based Approach

Ask-To-Rate Technique: In ask-to-rate based RS, the new user is asked explicitly to rate the selected items until having an enough number of rated items. In RS, asking a series of questions about their tastes or by recommending some items in order to get any rating that will make the user displeasure. It's also tedious and cumbersome activities.

The author proposed the ask-to-rate technique to solve the user cold-start problem in RS. In this technique, initially the system crawls and fetches news articles from news portals and then it stores the collected articles and metadata information in the centralized database[8]. When the new user enters into the system, she/he enters into the prompting phase in which a series of news articles are suggested to them. If the new user is not given a rating for the suggested articles then the system continues with the suggestions based on the rated list. The list is stored in the database. If the user has given a rating for one article then the system suggests for subsequent ratings. Then it will check that the articles that belong to the same article cluster or another cluster. This procedure continues until enough user ratings have been gathered. Finally, the recommender system goes into its normal recommendation process.

Recommendation Process

Step 1. Crawl and fetch the news articles from news portals. The collected articles and metadata information are stored in the centralized database.

Step 2. Perform the preprocessing on the fetched article's content. It includes cleaning of articles, noun extraction, stopword removal (eg: Common words such as 'a', 'an', 'the', or 'of') and stemming (e.g. 'cook', 'cooking', 'cooked' are all should be considered same).

Step 3. Cluster the articles based on the clustering algorithm.

Step 4. Forward the generated article cluster and cluster assignments for labelling. The labels are considered as the output.

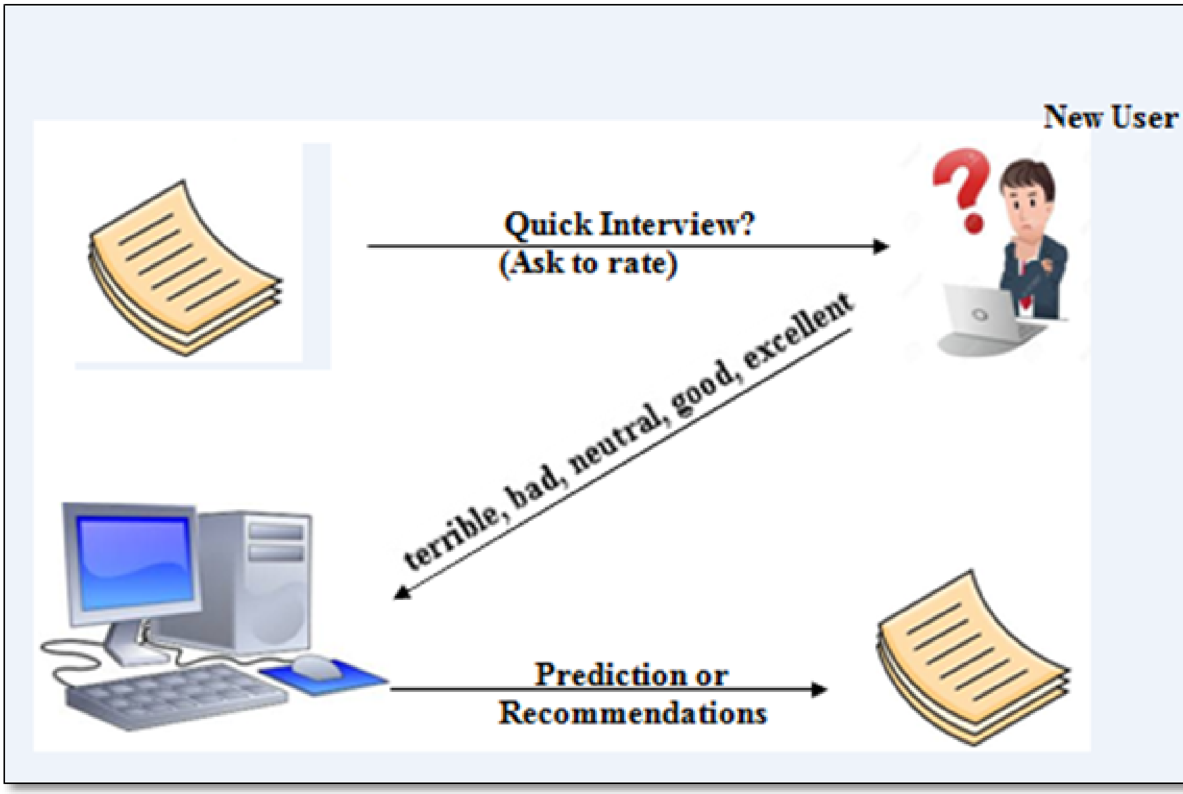

Figure 5. The new user prompting process.

Step 5. Recommend the news articles to the new user based on the labels.

The author used entropy and variance statistical method in the rating data in order to generate more accurate predictions for new users[9]. The author used the prompting approach (Active WebMuseum) that recommends paintings and artworks to the museum visitors. This method is based on nearest-neighbor collaborative filtering concept. This approachmake use of variance and entropy method in the rating data in order to generate more accurate predictions for a new user to solve the cold-start issue in recommender systems. They used the Movielens dataset to evaluate the performance of this approach. This prompting approach is shown in Figure 5.

The variance of the target item computed using the Eq. (1).

$$v(o) = \frac{\Sigma_{u \epsilon U_o}\left(r_{u,o} - \overline{r_o}\right)^2}{|U_o|} \quad (1)$$

Where the variance of the target item o, is refers to all users who have rated the target item, is the rating given by the user for the targeted item andis the average rating for the target item.

Recommendation Process

Step 1. Collect the new user information by asking the user to rate the more items during the short interview and make a profile of the user.

Step 2. Show the item set to the new user based on popularity (items should be famous to the users) and contention (items should be pinpointing of user's trend).

Step 3. Express the user preferences in a five scale rating such as 1 to 5 (terrible, bad, neutral, good, and excellent). If the new visitor has not been rated for the painting, then the ratings are predicted by the other user ratings.

The author proposed a prompting approach to solve the user cold-start problem in RS[10]. The user effort and recommendations accuracy to be considered while eliciting the ratings from a new user during the initial registration phase: 1. *User effort*: The sign-up process should not look difficult to the new user. The elicitation technique should ask for ratings that the user is likely to be able to provide. If the technique is not effective and the process is tedious, the new user will be frustrated, and 2. *Recommendation accuracy:* To list the items to the new user that would be able to pull out the user preferences. The RSsystems make accurate recommendations to the user and improve its predictions also. They proposed the following elicitation technique to acquire user preferences in the sign-up stage: *Entropy* - The items with the largest rating entropy are preferred. Random: The items are to be presented are chosen randomly with a uniform probability over the universe of items, *Popularity:* The items are sorted in descending order based on their number of ratings, *Pure entropy*: The users ask about items which give the most information for each rating, *Balanced strategy*: This technique combines popularity strategy and entropy strategy, and *Item–Item personalized*: The items are proposed randomly until one rating is acquired. The various strategies along two important dimensions of the new user problem (***** best, * worst performance) is shown in Table 1.

Recommendation Process

Step 1. Acquire the user rating based on any one of the elicitation methods during the initial interview process.

Step 2. Predict the ratings for items that the user is possible to see.

Table 1. Various strategies along two important dimensions of the new user problem (***** best, * worst performance[10])

Strategy	User Effort					Accuracy				
Random	★					★	★			
Entropy	★					★	★			
Popularity	★	★	★	★	★	★	★	★	★	
(log) Pop*Ent	★	★	★			★	★	★	★	★
Item-Item	★	★	★	★	★	★	★			

Step 3. Generate the item recommendation list to recommend the items to the new user.

5. Solution to Cold-Start Problem - Social Network Analysis Based Approach

Social network based recommender system: This recommender system integrates the social networking features to improve the recommendation accuracy. The Social Network Analysis (SNA) is the process of examining social structures through the use of network and graph theory concept. The social network represented in terms of nodes (individual actors or people within the network) and edges (relationships or interactions) that connect them.

The author focused the new user cold-start problem in recommender system and suggested a solution based on user social network interaction and analysis of their features[11]. The author proposed mobile app named Foursquare, which is a location-based social networking website. This app monitor and collects the user checked-in detail in a venue, which helps to overcome the new user cold-start issue in RS. In this approach, they used the eigenvector centrality measure to identify the most impor-

Table 2. Weights distribution for unknown user and friend user[11]

Unknown user	Friend user
Default weight is 1	Default weight is 3
If check-in interval <=1 hour, the weight is 3	If check-in interval <=1 hour, , the weight is 3
If 1 hour < check-in interval<= 2 hour, the weight is 2	If 1 hour < check-in interval<= 2 hour, the weight is 2
If 2 hour < check-in interval<=3 hour, the weight is 1	If 2 hour < check-in interval<=3 hour, the weight is 1

Table 3. User's Check-ins and Timestamps[11]

User's check-ins	Time stamps	Legend
Current user	1:00 pm	Friend user (1&4)
User 1	2:00 pm	Unknown user (3&2)
User 2	3:00 pm	1 hour interval
User 3	1:00 pm	2 hours interval
User 4	4:00 pm	3 hours interval

tant nodes in the network. The sample graph is built from a user check-in time. The weight distribution is based on the check-in time for unknown user and friend user are shown in Table 2. The user's check-in and timestamps are shown in Table 3. The graph shown in Figure 6 is represented in terms of an adjacencymatrix that is shown in Table 4. The eigenvalue is calculated from the collected user information. The highest eigenvalue nodes will be considered as the most important node in RS. This node used for making recommendations to the new user. In Table 4, A' is derived from matrix A (the default value for the unknown user and friend-user added), A" is derived from matrix A' (the relationship between the current user and other user (including friends) and the check-in time interval. In Figure 6, where there are 4 more foursquare users who had checked in the same venue.

Recommendation Process

Step 1. Built the graph from a user check-in detail.

Step 2. Find the adjacency matrix of the graph.

Step 3. Add the weight to the adjacency matrix based on the check-in time interval for the friend user and the unknown user.

Step 4. Calculate the eigenvalue from the resultant matrix.

Step 5. Recommend the highest eigenvalue items to the new user.

Table 4, A - Adjacency matrix for the graph given in Figure 6, A' – Derived from matrix A (the default value for the unknown-user and friend-user added), A" -Derived from matrix A' (the relationship between the current user and other user (including friends) and the check-in time interval.

The author proposed Enhanced Content-based Social Networking (ECSN) algorithm in academic social networks to solve a user cold-start problemin RS[12]. The faculty and friend's profile and their transactions are recorded and stored in the database based on the preference score. When a new user enters, ECSN algorithm uses the friend's and faculty mates profile and users' own preference. Then, it recommends the relevant academic items based on the preference score.

Table 4. Matrix value of the graph given in Figure 6[11]

0	1	1	1	1
1	0	1	1	1
1	1	0	1	1
1	1	1	0	1
1	1	1	1	0

A

0	3	1	1	3
3	0	1	1	3
1	1	0	1	1
1	1	1	0	1
3	3	1	1	0

A'

0	6	3	4	4
6	0	3	3	1
3	3	0	2	3
4	3	2	0	1
4	1	3	1	0

A"

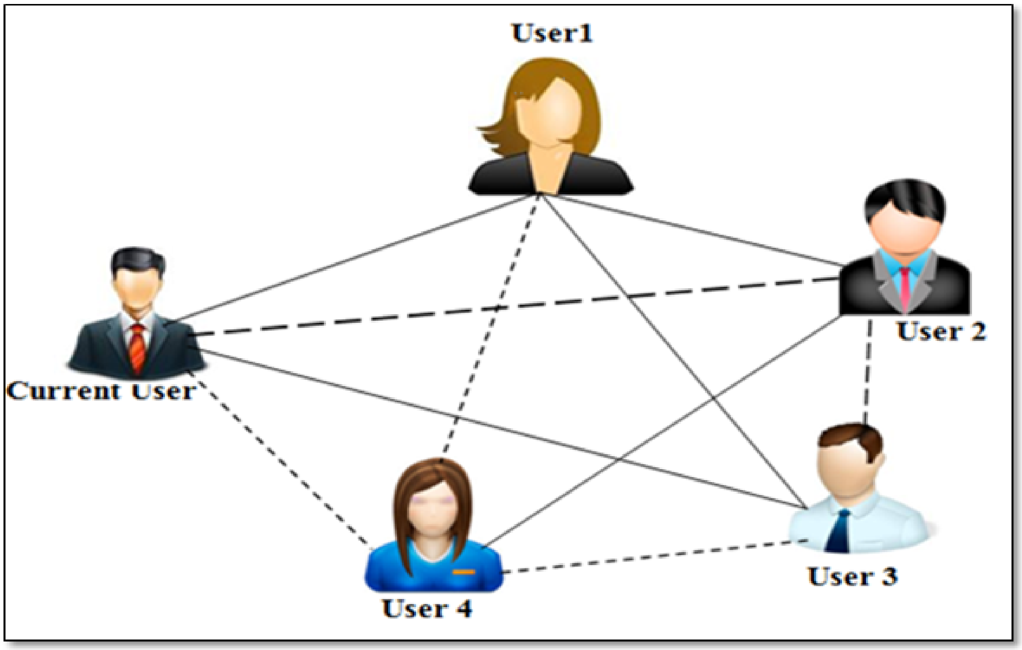

Figure 6. A sample graph built from a user Check-in[11].

The ECSN recommendation process is shown in Figure 7. The ECSN algorithm considers the following user'sdetails:

1. Interests and preferences of users' friends.
2. Interests and preferences of faculty mates.
3. Users own preferences.

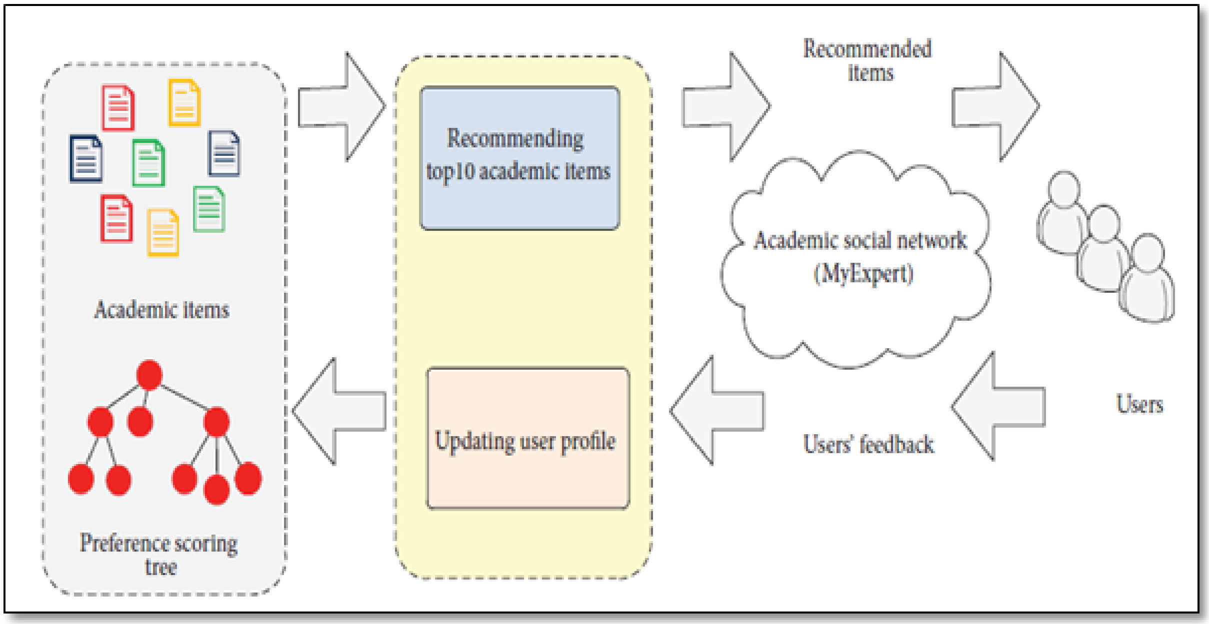

Figure 7. ECSN recommendation process[12].

Recommendation Process

Step 1. Store the interests and preferences of users in a hierarchy tree structure.
Step 2. Analyze the transaction records of a new user's friend using ECSN algorithm.
Step 3. Extract the most appropriate nodes from the preference tree structure of academic items.
Step 4. Update the nodes rating values in the preference tree of the new user.
Step 5. Repeat the same process for the faculty member of the new user.
Step 6. Calculate the preference score for the item stored in the preference tree.
Step 7. Recommend the top preference score items to the new user.

5.1 Community Based Solution to the Cold-Start Problem

The author proposed a solution to the cold-start problem in a collaborative recommender system based on Cold Start Community-Based Algorithm (CSCBA)[13]. The community detection algorithm detects the community based on the user relationship available in the known domain (eg: digital distribution systems). The external community profile available in the known domain can be used for the unknown domain to recommend the items to the new user because the new user community members (friends) already exist in that domain. Similarly, description about the new item was used as the seed to grow around the community. Automatically the new item would become known to the user in the community. RS provides a recommendation within smaller and highly similar community users, instead of the entire user in the database. This method provides high accuracy within a small amount of time.

Recommendation Process

Step 1. Apply the community detection algorithm in the user-movie rating database.
Step 2. List the top rated items in each community.
Step 3. Assign the new user to the appropriate community based on the known domain information.
Step 4. Recommend the top rated item to the new user.
Step 5. Create a short description of the new item. This description was used as the seed to grow around the community.
Step 6. Recommend the new item to all the communities, including the new user.

6. Conclusion

Recommender Systems are software applications that provide support to people by identifying interesting products and services in e-commerce sites. The new user cold-start problem and new item cold-start problem is a very big challenge in the collaborative filtering recommender system. In this paper, we have reviewed several methods for dealing with the cold-start problem via model based approach, demographic data based approach, ask-to-rate technique, and social network analysis methods. In future, a new method to be proposed to increase the recommendation accuracy in recommender systems.

7. References

1. Pereira ALV, Hruschka ER. Simultaneous co-clustering and learning to address the cold start problem in recommender systems, Knowledge-Based Systems. 2015; 82:11–19. Crossref.
2. Chekkai N, Chikhi S, Kheddouci H. A weighted-graph-based approach for solving the cold start problem in collaborative recommender systems. IEEE Symposium on Computers and Communication. Cappadocia, Turkey; 2012. p. 759–64. Crossref.
3. Latora V, Marchiori M. How the science of complex networks can help developing strategies against terrorism, Chaos, Solitons and Fractals. 2004; 20(1):69–75. Crossref.
4. Li C, Ma L, Dong K. Collaborative filtering cold-start recommendation based on dynamic browsing tree model in e-commerce. International Conference on Web Information Systems and Mining. Shanghai, China; 2009. p. 620–24. Crossref.

5. Safoury L, Salah A. Exploiting user demographic attributes for solving cold-start problem in recommender system, Lecture Notes on Software Engineering. 2013; 1(3):303–07. Crossref.
6. Seung-Taek P, Chu W. Pair-wise preference regression for cold-start recommendation. Proceedings of the third ACM conference on Recommender systems; 2009. p. 23–25.
7. Chen T, He L. Collaborative filtering based on demographic attribute vector. IEEE International conference on future Computer and Communication; 2009. p. 225–29. Crossref.
8. Bouras C, Tsogkas V. Clustering to deal with the new user problem. Proceedings of the 15thIEEE International Conference on Computational Science and Engineering, Nicosia, Cyprus; 2012. p. 1–8. Crossref.
9. Kohrs A, Merialdo B. Improving collaborative filtering for new users by smart object selection. CiteSeer. 2001. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.61.5243
10. Rashid AM, Albert I, Cosley D, Lam SK, Mc Nee SM, Konstan JA, Riedl J. Getting to know you: Learning new user preferences in recommender systems. Proceedings of the 7th International Conference on Intelligent user Interfaces; 2002, p. 127–34. Crossref.
11. Castillejo E, Almeida A, Lopez-de-Ipina D. Social network analysis applied to recommendation systems: Alleviating the Cold-User Problem. International Conference on Ubiquitous Computing and Ambient Intelligence; 2012. p. 306–13. Crossref.
12. Rohani VA, Kasirun ZM, Kumar S, Shamshirband S. An effective recommender algorithm for cold-start problem in academic social networks, Mathematical Problems in Engineering. 2014; 2014:1–11.
13. Vairachilai S, Devi MKK, Raja M. Alleviating the cold start problem in recommender systems based on modularity maximization community detection algorithm, Circuits and Systems. 2016; 7:1268–79. Crossref.