Community Detection-Based Recommendation Framework

Karen K. Mkhitaryan

Institute for Informatics and Automation Problems of NAS RA
e-mail: karenmkhitaryan@gmail.com

Abstract

Recommender system is a type of information filtering system predicting users preferences about items, aiming to generate personalized recommendations. Various recommendation approaches exist in the literature that differ in terms of methodology and types of systems they can be utilized on. In recent years some attempts have been made to incorporate community detection methods into recommender systems to make the process of recommendation generation more accurate in terms of rating or preference prediction and efficient in terms of computational complexity. In this paper we propose a community detection-based approach for recommender system, which is more reasonable in certain applications.

Keywords: Community Detection, Recommender Systems.

1. Introduction

In this era of Big Data it is hard to imagine an area, which does not deal with the collection and analysis of data. Recommender systems being part of information filtering system is such area, aiming to suggest relevant information to users based on available data. Applying recommendation techniques help companies to increase revenues and customer satisfaction, make more personalized user profiles, etc. Several real world examples of RS include movie recommendation (MovieLens), song recommendation (Last.fm) and product recommendation (Amazon).

One of the most popular recommendation techniques is **collaborative filtering (CF)**, which predicts users interests i.e., how the user will rate the item, by analyzing the data about the users and their preferences. Two popular types of CF are **memory-based** and **model-based** approaches [1]. In memory-based or **neighborhood-based** approach, ratings are predicted by applying similarity measure on user-item rating data, while in **model-based** approaches machine learning and data mining techniques (e.g., clustering, singular value decomposition, Bayesian networks, etc.) are used to predict users preferences. There are also **hybrid CF** algorithms that use both memory-based and model-based approaches, which can improve prediction performance. Moreover, various approaches from other disciplines such as machine learning and network science were incorporated into recommender systems for the purpose to increase efficiency of recommendation generation.

Community detection is a research area from network science, providing tools to partition

the relational data (network) composed of entities (nodes) and interactions among them (edges) into subgroups, called communities or clusters that have dense connections inside and sparse connections with other subgroups in the network.

Some efforts have been made to combine community detection and recommender system techniques. In [2] authors showed the principal stages in community-based recommender system and how recommender systems can benefit from detection of communities. In [3] Deng et. al incorporated community detection into SVD++ model which has improved accuracy compared with SVD model. Abdrabbah et. al [4] proposed dynamic community-based collaborative filtering approach combining both collaborative filtering and dynamic community detection techniques. Proposed approach outperforms both item-based collaborative filtering and collaborative filtering based on static community detection.

In this work we propose community detection based recommendation framework which is more reasonable in certain applications compared with traditional recommendation techniques.

The paper is organized as follows: In Section 2 we overview the popular recommendation approaches and present our proposed framework in Section 3.

2.   Recommender Systems

Recommender systems play an important role in suggesting relevant information to users, which are utilized in many areas such as recommending books, research articles, songs and products. Modern literature provides many recommendation techniques and algorithms designed to predict users preferences. Currently recommender systems are mostly developed using the collaborative filtering, content-based and hybrid model approaches that we discuss in the following subsections. (Fig. 1)
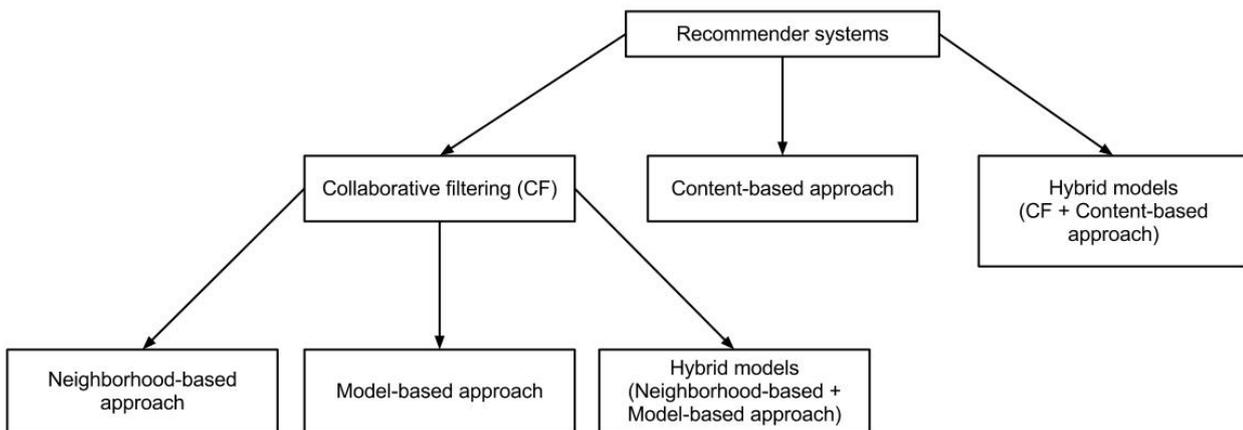
Fig. 2. Different implementations of recommender systems

2.1 Collaborative Filtering

Collaborative filtering (CF) is used to make predictions about X user's interest by analyzing the tastes or preferences of other users similar to user X. It has three main approaches

which are memory-based or neighborhood-based CF, model-based CF and hybrid CF.

Memory-based approach
Memory-based CF considers the similarities of users and items and is divided into two subsections: user-user filtering and item-item filtering. In user-user filtering, the idea is to find similar users to the given user X and recommend items to X those similar users liked. In contrast, item-item filtering finds users that liked an item and recommends other items that those users liked. Below we demonstrate a simple example how user-user CF is working. Consider a user-item rating matrix R where each R_{ij} element shows the rating user i gave to book j (Table 1).

Table 1: User-rating matrix.

R	Book 1	Book 2	Book 3	Book 4
User 1	5	3	1	2
User 2	4	1	?	1
User 3	5	5	1	3
User 4	4	?	3	1
User 5	1	?	?	2

In the matrix we see unknown ratings denoted by "?". The goal of both user-user and item-item filtering is to predict the unknown ratings using the R matrix. In the first step, similarities of users are calculated using similarity measure (e.g., cosine similarity, Pearson correlation, Manhattan distance, Euclidean distance, etc.). In our example we use cosine similarity defined as:

$$\text{Cosine}(x, y) = \sum_i \frac{r_{xi} r_{yi}}{\sqrt{\sum_i r_{xi}^2} \sqrt{\sum_i r_{yi}^2}}, \tag{1}$$

where r_{xi} and r_{yi} are the ratings users x and y gave to item i respectively. Symmetric matrix U of similarities between users is shown in Table 2.

Table 2: Cosine similarity values between users.

U	User 1	User 2	User 3	User 4	User 5
User 1	1	0.943	0.971	0.785	0.644
User 2	0.943	1	0.852	0.785	0.632
User 3	0.971	0.852	1	0.658	0.635
User 4	0.785	0.785	0.658	1	0.526
User 5	0.644	0.632	0.635	0.526	1

Next, suppose we want to predict how User 5 will rate the Books 2 and 3, which are missing in Table 1. To do this we calculate the weighted average of item ratings and user similarities.

$$R(\text{User 5}, \text{Book2}) = \frac{3 * 0.644 + 1 * 0.632 + 5 * 0.635}{0.644 + 0.632 + 0.635} = 3.003. \tag{2}$$

The same way $R(\text{User 5}, \text{Book 3}) = 1.582$.
Finally as $R(\text{User 5}, \text{Book 2}) > R(\text{User 5}, \text{Book 3})$, Book 2 will be recommended to User 5.

Item-item filtering approach works in similar way considering item-item similarities instead of user-user similarities.

Model-based and hybrid approaches
In model-based approaches machine learning and data mining algorithms are used to develop models predicting the rating user will give to item. General approaches include clustering algorithms (k-nearest neighbors), matrix factorization (singular value decomposition) and deep learning (multi-layered neural nets). Hybrid approaches combine both memory-based and model-based approaches and can improve prediction performance and data sparsity issues. However, hybrid approaches may have higher computational complexity.

2.2 Content-Based Filtering

Unlike collaborative filtering approach, which is based on large amount of "collaboration" data between users and items where recommendation engine does not rely on the "understanding" of the product itself, content-based approaches are based on the item descriptions and the users taste. Recommendation engines developed on the principle of content-based approach, learn about users preferences and recommend items that have similar features to items the user likes. Bayesian classifiers, decision trees, neural networks can be used to develop content-based filtering algorithms.

3.   Proposed Framework

As we saw in previous section, different CF approaches take "collaboration" data between users and items as input and predict how the user will rate the items, which are not previously rated by him/her. On the other hand, content-based approaches recommend similar items to what user liked based on the description of an item itself. Both collaborative filtering and content-based filtering approaches rely on similarities of users and items, which is not necessarily typical to some recommendation tasks. In proposed framework, recommendation to target user is done by considering preferences of users lying in the same community in the network constructed using the similarities of users based on predefined features. Recommendations are generated by the implementation of the following steps: (Fig. 3.):
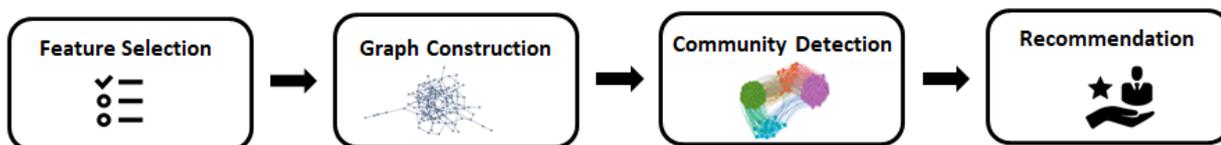
Fig. 2. Steps to make recommendation.

Step 1 - Feature Selection
Recommendation process starts with identifying features, which are a set of attributes that describe the user (e.g., age, favorite music genre). Features help to discriminate users based on their preferences, however their selection highly depends on the nature of recommender system it will be implemented on.

Step 2 - User Similarity Calculation and Graph Construction

In this step similarity between users is calculated. Some popular measures of similarity are cosine similarity, Euclidean distance, Manhattan distance, Pearson correlation coefficient, etc., which can be used to compare users based on predefined features. After the measure is applied and the similarity scores between users are obtained, the next step is a construction of a weighted network or graph where nodes and edges represent the users and similarity scores between them, respectively.

Step 3 - Community Detection

After a weighted network is obtained, community detection algorithms designed for weighted networks such as Louvain [5] or fast greedy modularity optimization [6] can be applied to partition the network into community structure.

Step 4 - Recommendation

In the final step, recommendation is done using the preferences of users about the items in the community of a target user, i.e., items that majority of users preferred in the community are recommended to the user.

This framework also enables to tackle some drawbacks that exist in collaborative filtering and content-based filtering techniques. Running algorithms on systems composed of millions or billions of users and items cause scalability issues, however in our framework community detection algorithms do not have too large complexity ($O(n \log n)$ for Louvain and $O(n \log^2 n)$ for fast greedy). Another issue typical to content-based filtering is over-specialization i.e., recommendation of items that are very close to what user likes and is aware of. This problem was also tackled in this framework as item similarities and description of an item are not taken into consideration.

4. Conclusion

In this paper we discussed several popular recommendation algorithms that are utilized in recommender systems. We propose community detection-based recommendation framework, the use of which is justified in certain applications. The framework also enables to eliminate several drawbacks that are typical to traditional recommendation techniques such as scalability and over specialization.

Acknowledgement

References

[1] F. Ricci, L. Rokach and B. Shapira, "Introduction to Recommender Systems Handbook", *Springer*, pp. 1-35, 2011.

[2] F. Gasparetti, A. Micarelli and G. Sansonetti, "Community Detection and Recommender Systems", *Encyclopedia of Social Network Analysis and Mining*, pp. 1-14, 2017.

[3] W. Deng, R. Patil, L. Najjar, Y. Shi and Z. Chen, "Incorporating Community Detection and Clustering Techniques into Collaborative Filtering Model", *Procedia Computer Science*, vol. 31, pp. 66-74, 2014.

[4] S. B. Abdrabbah, R. Ayachi, N. B. Amor, "Collaborative Filtering based on Dynamic Community Detection", *DyNaK*, 2014.

[5] V. D. Blondel, J. Guillaume, R. Lambiotte and E. Lefebvre, "Fast unfolding of communities in large networks", *Journal of Statistical Mechanics: Theory and Experiment,* vol. 2008, 2008.

[6] A. Clauset, M. E. J. Newman, and C. Moore, "Finding community structure in very large networks", *Phys. Rev. E*, vol. 70, no. 066111, 2004.

Համայնքների հայտնաբերման վրա հիմնված խորհրդատվական միջավայր

Կ. Մխիթարյան

Ամփոփում

Խորհրդատվական համակարգերը ինֆորմացիայի մշակման համակարգերի տեսակ են, որոնք կանխագուշակում են օգտատերերի նախընտրությունները և ստեղծում են առաջարկություններ: Գրականության մեջ գոյություն ունեն բազմապիսի ալգորիթմներ, որոնք տարբերվում են իրենց մոտեցումներով: Այս հոդվածում դիտարկված են որոշ հայտնի խորհրդատվություն իրականցնող ալգորիթմներ և առաջարկված է համայնքների հայտնաբերման մեթոդների վրա հիմնված խորհրդատվական համակարգի մոդել, որն ավելի հիմնավորված է որոշակի կիրառություններում: Միջավայրը հնարավորություն է տալիս նաև վերացնել մի շարք թերություններ, որոնք բնորոշ են ավանդական խորհրդատվական մեթոդներին, ինչպիսիք են, օրինակ՝ հաշվողական բարդությունը մեծածավալ տվյալների դեպքում և խորհրդատվությունների նեղ մասնագիտացումը:

Рекомендательная среда на основе обнаружения сообществ

К. Мхитарян

Аннотация

Рекомендательные системы - это вид систем обработки информации, которые предсказывают предпочтения пользователей и создают предложения. В литературе существует множество алгоритмов, которые отличаются по своим подходам. В этой статье рассматриваются некоторые из наиболее популярных рекомендательных алгоритмов и предлагается модель системы на основе методов обнаружения сообществ, которая более обоснована в определенных приложениях. Среда также позволяет устранить ряд недостатков, типичных для традиционных методов рекомендации, такие как, например, вычислительная сложность в случае больших данных и узкая специализация рекомендаций.