A Survey of Recent Articles in the Field of Recommendation Diversification

Zahra Eslamian Koupaee
Faculty of Graduate Studies
Safahan Institute of Higher Education Isfahan, Iran
z.eslamian@safahan.ac.ir

Mohammad Reza Khayyambashi
Department of Computer Architecture
Faculty of Computer Engineering, University of Isfahan Isfahan, Iran
M.R.Khayyambashi@eng.ui.ac.ir

Abstract—Today, with the explosive growth of the internet applications and in the current age of information overload, recommender systems are steadily becoming more important in filtering relevant information and items for users and also in keeping customers and gaining more benefits. Based on the assumption that users with similar preferences in history would also have similar interests in the future, collaborative filtering algorithms have shown significant successes and become one of the most pervasive branches in the study of personalized recommendation. However, while most research focused on improving the accuracy of recommender systems, other important aspects of recommendation quality, such as the diversity of recommendations, have often been overlooked. The ability of recommending a diverse set of items is very important for user satisfaction, because it gives users a richer set of items to choose from and increases the chance of discovering new items. With the development of electronic markets and arrival of diverse and new goods, addressing the diversity factor has become very important and received a lot of attentions in recommender systems literature. But there is a big vacancy in the classification of the works done in this area which is necessary in order to clarify and streamline the directions for future research. Therefore, in this paper, the existing approaches and some of the authentic papers published in recent two or three years are introduced.

Keywords: Recommender Systems, Collaborative Filtering, Diversity, Accuracy.

1. INTRODUCTION

In the currentage of information overload, finding relevant content is becoming increasingly harder and also a big concern. However, over the last 10-15 years, recommender systems technologies have been introduced to help people deal with this problem and they have been widely used in research as well as e-commerce applications [1].

Although, different algorithms has been proposed in recommender systems literature, but collaborative filtering (CF) and content – based (CB) are two of the most practical approaches.

A collaborative filtering approach can be designed based on either user similarities or object similarities, which is called user – based or item – based respectively. A user-based CF approach assumes that users with similar preferences in history would also have similar interests in the future. So, the performance of such an approach depends largely on the calculation of similarity scores between the users. An item- based CF approach is the same as user- based but just the roles of users and items should be changed. In contrast, a CB recommender system uses contents and features of items to discover their similarities and then recommends to the active user new items that are similar to the ones which already got high rating by the user. In order to increase the advantages of these two methods, some hybrid approaches has also been proposed [2].

Collaborative filtering approach, unlike content-based methods, does not require any content information and can recommend any items, even the ones that are dissimilar to those seen in the past which is more appropriate in the field of recommendation diversification [3]. So, CF can be applied to any domain, independent of items’ features and it is also relatively easier to implement, compared to other algorithms and provides the most satisfactory effects [4].

In order to satisfy the users and encourage them to continue using the proposed recommender system, different aspects such as accuracy, diversity, novelty, serendipity, and trust must be considered [5], [6]. Among these factors, accuracy and diversity are of major importance [5], [7], [8].

Many studies have been made on the development of new algorithms that can be used to increase the accuracy of recommendations and different solutions have been proposed to decrease cold-start and sparsity problems [9], especially with the advent of social networks [10]. However, in recent years, several studies have demonstrated that high accuracy does not always lead to user satisfaction. They suggest that having accuracy as the sole criteria in measuring recommendation quality is not sufficient, and considering other important aspects, such as diversity is necessary to generate recommendations that are helpful to users [5]. In fact, accuracy-oriented algorithms, sometimes even decrease user satisfaction because the recommendation results often have similar content and reflect only some aspects of the user interests [11].

Therefore, in today's economic crisis when having a good sale and profitability of all products that a business has invested on and also not limiting a market to a few specific and famous items, are of major importance, addressing the diversity factor is essential. However, many studies [1], [2], [6], [12] have demonstrated that the gains of the recommendation accuracy are often accompanied by the losses of the diversity, the phenomenon which is called accuracy-diversity dilemma; Because to achieve diversity in the recommendations, the recommender system should remove its focus on just recommending famous items and take into consideration a wider range of items including niche and long tail items [13], [14]. Also it is sometimes necessary for the recommender system to accept the risk of losing the reliance of ratings predicted by the conventional collaborative filtering methods and interference new weights and factors in computing ratings.
The main contribution of this paper, which is the result of months of effort and research in the field of recommender systems and especially collaborative filtering, is to introduce some works done in the last two or three years to improve the recommendation diversity and it attempts to provide an overview of existing approaches and pave the way for those interested in this domain.

The remainder of this paper is organized as follows. Section 2 provides a background of CF algorithm and the procedure of it. Section 3 introduces the existing approaches and papers in the field of recommendation diversity. In section 4, famous datasets, which are applied in the most articles in order to do the experiments, are introduced. In section 5, some of the most commonly used metric for evaluating recommender systems from the perspective of accuracy and diversity are explained. Finally, section 6 concludes the paper and summarizes the future directions.

2. BACK GROUND

2.1 Collaborative Filtering

The CF algorithm procedure generally has three steps. The first step is to create a matrix from users’ ratings on different items. Then a neighborhood for the active user or item is formed based on the similarity value, which is computed by a similarity measure such as cosine, Pearson correlations and so on. In the second step, the unknown rating of the active user on an item, which is not consumed by him yet, is calculated using a weighted combination of selected neighbors’ ratings. Finally, the system puts the top N items, according to their ratings, in descending order and makes the top-N recommendation list for the active user [9].

2.2 Similarity Measures

Assume that I is the set of n items and U is the set of m users and R is the rating matrix of size m×n. As mentioned earlier, CF makes recommendations for an active user by first finding a set of similar users and exploiting their ratings. Several similarity measures have been proposed for computing the similarity between two users.

Next, the most commonly used similarity measures will be described.

2.2.1 Pearson correlation coefficient

\[
\text{sim}(u, v) = \frac{\sum_{i \in I(u) \cap I(v)} (r_{u,i} - \bar{r}_u)(r_{v,i} - \bar{r}_v)}{\sqrt{\sum_{i \in I(u) \cap I(v)} (r_{u,i} - \bar{r}_u)^2 \sum_{i \in I(u) \cap I(v)} (r_{v,i} - \bar{r}_v)^2}}
\]

(1)

where, I(u) and I(v) are the sets of items rated by users u and v respectively, and \(\bar{r}_u\) is the average rating of user u [14].

2.2.2 Cosine similarity

In this approach, the similarity between two users is measured by computing the cosine of the angle between them [14].

\[
\text{sim}(u, v) = \frac{\sum_{i \in I(u) \cap I(v)} r_{u,i} r_{v,i}}{\sqrt{\sum_{i \in I(u) \cap I(v)} r_{u,i}^2 \sum_{i \in I(u) \cap I(v)} r_{v,i}^2}}
\]

(2)

2.2.3 Jaccard index

If each user is considered as a set of items preferred by the user, Jaccard index calculates the similarity score between two users as follows [2]:

\[
\text{sim}(u, v) = \frac{|u \cap v|}{|u \cup v|}
\]

(3)

2.3 Rating Prediction

Resnick’s prediction formula is the most common formula used to predict unknown ratings, which calculates the rating of active user a on an item i as follows [14]:

\[
P_i(a, i) = \hat{r}_a + \frac{\sum_{u \in N(a)} \text{Sim}(a, u) (r_{u,i} - \bar{r}_a)}{\sum_{u \in N(a)} |\text{Sim}(a, u)|}
\]

(4)

where \(\bar{r}_a\) denotes the average rating of user a and \(N(a)\) represents the neighbors of user a.

3. INTRODUCTION AND CLASSIFICATION OF EXISTING APPROACHES

Existing approaches in the field of recommendation diversity can be divided into several categories (see Figure 1):

1) Popularity-bias problem.
2) Diversifying top-N recommendation.
3) Probabilistic Approaches.
4) Clustering-Based Approaches.

Next, some of the articles published in the years 2012 to 2014 in the prestigious international journals and conferences, are introduced.

3.1 Popularity-bias problem

- Gan and Jiang [2], while keeping a reasonable tradeoff between accuracy and diversity, introduced a power function to adjust user similarity scores, with the aim of reducing adverse effects of popular items.

- Zhao et al. [15] propose an opinion-based CF and introduce weighting functions used in measuring users’ similarities to adjust the influence of popular items.

- Again, Gan and Jiang [16], in order to overcome the adverse influence of popular items, propose a network-based CF approach. This method first constructs a user similarity network from historical data by using a nearest neighbor approach, then calculates discriminate scores for candidate items.
Diversifying top-N recommendation

Diversity is divided into two categories: Individual diversity and aggregate diversity. The individual diversity refers to increasing item’s similarity to the user and its dissimilarity to the items that are already recommended to the user. The aggregate diversity is considered as the total number of items recommended to all users. High individual diversity does not necessarily mean high aggregate diversity, e.g. if each user receives recommendations for the same 10 movies from 10 different genres, the individual diversity is high, but the aggregate diversity is still low [6].

There are two main ways to improve the aggregate diversity. The first way make predictions for the unknown ratings using existing filtering approaches and then re-rank the items to bring long tail items to the recommendation lists. The second way tries to improve the prediction formula especially for items that are rarely offered [6].

The first research direction to improve the diversity, namely ranking top-N recommendations, can be divided into two categories: Hybrid Ranking and Re-ranking.

3.2.1 Ranking top-N recommendations

- Premchaiswadi et al. [12] proposed a new recommendation ranking method, namely “Total Diversity Effect Ranking”, and combined this method with Standard Ranking (SRank) in order to make a compromise between accuracy and diversity.
- Su et al. [17] introduced a set-oriented personalized ranking model and integrated the concept of diversity into traditional matrix factorization model in order to achieve personalized top-N recommendation results which are both relevant and diversified.
- Another work in this regard, which combines different ranking approaches, is the work done by Adomavicius and Kwon [1].

3.2.2 Improving prediction process

- Niemann and Wolpers[6] propose a new CF approach that is based on the items’ usage contexts. This means that an item is described by the items it is usually used with rather than by its users or features. The approach improves predictions for niche items which usually have less information available.

  • In another work, Zhang et al. [18] propose an adaptive CF method based on multiple features. First, user vector and item vector are built based on their features; second, the rating is modified based on the principle of locality of time and user-item relation; then user-user similarity and item-item similarity are recalculated using the vectors created in the first step and also the initial similarity; finally, a new prediction formula is proposed.

3.3 Probabilistic Approaches

- Wu et al. [7] focus on how to help users find more interests in the recommendation list and propose a sampling-based algorithm “Probabilistic Top-N Selection” to recommend potential interests to users. They said that traditional methods in selecting top-N items are very probable to miss the truly interesting items to user. They suggest that losing the reliance of predicted ratings properly and enlarging the range of potential candidates can give the users a better chance to find their true interests.

- Kabutoya et al. [11] propose a method which create a recommendation list so as to maximize the probability that a user buys at most one item from the list. They also use a greedy heuristic in order to approximately optimizing the objective function.

3.4 Clustering-Based Approaches

- AytekinandKarikaya [8] propose a method, called “ClusDiv”, which its idea is to cluster items into groups and then create the recommendation list by selecting items from different groups,
such that, recommendation diversity is maximized while maintaining an acceptable level of accuracy.

- Liand Murata [19] present a multidimensional clustering approach which facilitates obtaining user groups in order to improve the diversity of recommendations. In the first step, user groups are collected by applying multidimensional clustering algorithm. Then, the clusters of the first step are pruned in order to choose the appropriate clusters. Finally, the recommendations are generated for the active user.

- Another [20] work which is done in the field of recommendation diversity, is the work done by Wang and Yin. They, with doing some experiments on Movielens dataset, showed that:” 1) When N is small, user-based CF (UBCF) and item-based CF (IBCF) tend to recommend popular items. 2) As N increases, UBCF tends to recommend popular items, while IBCF tends to recommend long tail items. 3) As N increases, compared to UBCF, IBCF can recommend more distinct items”. So in order to make a balance between accuracy (recommending popular items) and diversity (recommending long tail items), this study proposed a synthetically CF model which combined the user-based and item-based CF techniques.

4. DATASETS

Two of the datasets which are used in most articles in order to do the experiments are Movielens and Netflix. In the following, a brief description of each of these datasets will be listed.

4.1 Movielens Dataset

Movielens Dataset is provided by the GroupLens Research Group in the University of Minnesota. This dataset consists of 10,000 ratings from 943 users on 1682 movies. Each user has rated at least 20 movies. Collected ratings are in a 1-to-5 star scale. Movielens dataset also contains some simple demographic information including age, gender, occupation, zip. In this dataset, 19 different movie genres are considered.

4.2 Netflix Dataset

Netflix dataset consists of 100,480,507 ratings from 480,189 users on 17,770 movies. Ratings are in a 1-to-5 star scale. To protect customer privacy, each customer id has been replaced with a randomly-assigned id. The date of each rating and the title and year of release for each movie id are also provided.

5. EVALUATION METRICS

Since this paper focuses on two aspects of recommendation quality, namely accuracy and diversity, in the following, the most important and common metrics in this regard will be introduced.

5.1 Accuracy Metrics

Various metrics have been applied for measuring the recommendation accuracy, including statistical accuracy metrics and decision-support measures. Statistical accuracy metrics in fact compare the predicted ratings to the actual ratings in the user–item matrix [1], [21]. Mean absolute error (MAE) and root mean squared error (RMSE) are the examples of statistical accuracy metrics [22].

\[
\text{MAE} = \frac{1}{K} \sum_{u,i} |p_{u,i} - r_{u,i}| \quad (5)
\]

\[
\text{RMSE} = \sqrt{\frac{1}{K} \sum_{u,i} (p_{u,i} - r_{u,i})^2} \quad (6)
\]

In the above equations, \(p_{u,i}\) represents the predicted rating of user \(u\) on item \(i\) while \(r_{u,i}\) symbolizes the actual rating and \(K\) refers to the number of items under evaluation.

Decision support measures specify how well a recommender system can generate predictions of high-relevance items. Precision and recall are two examples of such measures [21].

\[
\frac{|\text{relevant_items} \cap \text{recommended_items}|}{|\text{recommended_items}|} \quad \text{precision} = (7)
\]

\[
\frac{|\text{relevant_items} \cap \text{recommended_items}|}{|\text{relevant_items}|} \quad \text{recall} = (8)
\]

These measures are often, conflicting in nature. For example, increasing the number of recommended items \(N\), leads to increase recall but decreases precision [21]. Therefore, in most papers, F1-measure, which is a weighted average of precision and recall, is employed [9].

\[
\frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad \text{F1-measure} = (9)
\]

5.2 Diversity Metrics

Coverage is one of the crucial metrics in evaluating recommendation diversity. The coverage of a recommender system is actually the domain of items over which the system can generate recommendations. If it is assumed that \(I\) is the set of available items and \(I_r\) is the set of items for which a prediction can be made, prediction coverage can be computed as follows [23]:

\[
\text{Prediction Coverage} = \frac{|I_r|}{|I|} \quad (10)
\]

As mentioned earlier, diversity can be divided into two categories, individual diversity and aggregate diversity. Aggregate diversity can be measured by the Hamming distance or coverage. Since, the aim usually is to measure the performance of the recommender system based on the top-\(N\) recommendation lists, most papers use coverage as the metric of aggregate diversity [20]:

\[
\text{AD} = \frac{1}{|I|} \sum_{u \in U} |I_u\| - |I_u| \quad (11)
\]

where \(I\) is the item set and \(|I|\) is the number of items in the set. \(Nu\) represents the recommendation list which is recommended to the user \(u\) and \(|J/|\) refers to the total number of items recommended across all users.

If recommended items for user \(u\) is represented as the set \(\{i_1, i_2, ..., i_N\}\), the individual diversity is calculated as the averaged dissimilarity of all pairs of items in therecommendation list [20]:

\[
ID_u = \frac{1}{N(N-1)/2} \sum_{i=0}^{N} \sum_{j \neq i} \text{d}(i,j) \quad (12)
\]
where $\text{sim}(i,j)$ is the similarity between items $i$ and $j$ and $d(i,j)$ refers to the dissimilarity, which is defined as $1$ minus the $\text{sim}(i,j)$.

6. CONCLUSIONS AND FUTURE WORK

Recommender systems have achieved remarkable progress in recent years and various techniques have been proposed to improve recommendation quality. However, in most cases, new approaches have been designed to improve recommendation accuracy, while the diversity of recommendation has often been overlooked. As mentioned throughout this paper, because of the importance that recommendation diversity has gained in recent years with the arrival of diverse merchandise in electronic markets, and the crucial role that this factor has in helping users find more interests, some works have been done in the field of recommendation diversification.

However, the lack of a classification of the existing works, which is necessary in order to clarify and streamline the directions for future research, was the main motivation for writing this paper. This paper introduced the existing approaches and some of the authentic papers published in the years 2012 to 2014 and classified them according to the applied approach.

As mentioned before, there is a trade-off between accuracy and diversity. Since improving the diversity of recommendations requires that the recommender system removes its focus on just recommending famous items and takes into consideration a wider range of items which may have less available information, increasing the diversity usually causes the loss of accuracy. This question that how diversity of recommendations can be increased more and how all of the items, that an electronic market has invested on, can be entered into recommendation process, is still an open issue and needs lots of work. Even in some studies [1] it is mentioned that another interesting subject for future research is to discover possible methods in order to simultaneously improving both accuracy and diversity.

It seems that some solutions can be applied in this regard and some of them which is now just an idea and their feasibility will be tested in future studies, includes:

- **Considering the problems such as cold-start and sparsity and incorporating them with the approach proposed to improve recommendation diversity.**

- **Integrating different approaches in the field of recommendation diversification; For instance, combining ranking methods with the approaches which present a better prediction formula.**

- **Applying more information about users and items in the recommendation process; For example, considering some demographic information and recommending items that are relevant to user's working field and interest to enhance personalization.**

ACKNOWLEDGMENTS

This article is part of the M.S.C. thesis of the first author in Safahan Institute of Higher Education. The authors give their sincere thanks to institute officials.
Filtering to Solve Popularity Bias in Recommender Systems”,


