Improving Rating Predictions Using Domain-sensitive Recommendation Algorithm with Collaborative Filtering Algorithms

Shilpa R. Bodkhe  
SGBAU, Amravati  
Maharashtra, India

Prof. Ratnashil N. Khobragade  
SGBAU, Amravati  
Maharashtra, India

Dr. V.M. Thakare  
SGBAU, Amravati  
Maharashtra, India

Abstract—Collaborative Filtering (CF) is one of the most successful recommendation approaches to cope with information overload in the real world. Typical CF methods equally treat every user and item, and cannot distinguish the variation of user’s interests across different domains. This violates the reality that user’s interests always center on some specific domains, and the users having similar tastes on one domain may have totally different tastes on another domain. This paper proposed an approach DsRec algorithm with the CF-based algorithms i.e. Memory-based and Model-based algorithms to find the rating predictions in less time.

Index Terms—Context-Aware Recommendation System, Collaborative Filtering, Domain-sensitive Recommendation Systems, Memory-based, Model-based algorithms.

I. INTRODUCTION

Although the area of context-aware recommender systems (CARS) has made a significant progress over the last several years, the problem of comparing various contextual pre-filtering, post-filtering and contextual modeling methods remained fairly unexplored. Most traditional Recommender Systems (RSs) provide recommendations of items to users and vice versa and do not take into consideration the circumstances and other contextual information when recommendations take place. Recently, some companies started taking into account the contextual information. Experimental research on customer modeling suggests that including context in a customer behavior model improves the ability to predict her behavior in some cases because it allows the identification of more homogeneous patterns in the data describing the purchasing history of a customer.

Collaborative Filtering (CF) is an effective and widely adopted recommendation approach. Different from content-based recommender systems which rely on the profiles of users and items for predictions, CF approaches make predictions by only utilizing the user-item interaction information such as transaction history or item satisfaction expressed in ratings, etc. As more attention is paid on personal privacy, CF systems become increasingly popular, since they do not require users to explicitly state their personal information. A novel Domain-sensitive Recommendation (DsRec) algorithm is proposed to make the rating prediction by exploring the user-item subgroup analysis simultaneously, in which a user-item subgroup is deemed as a domain consisting of a subset of items with similar attributes and a subset of users who have interests in these items. The proposed framework of DsRec includes three components: a matrix factorization model for the observed rating reconstruction, a bi-clustering model for the user-item subgroup analysis, and two regularization terms to connect the above two components into a unified formulation. Extensive experiments on Movielens-100K and two real-world product review datasets show that our method achieves the better performance in terms of prediction accuracy criterion over the state-of-the-art methods[1]. In this approach, different item-based recommendation generation algorithms are analysed. There are two fundamental challenges. First is to improve the scalability of the collaborative filtering algorithms. Second is to improve the quality of recommendations for the users. These two challenges are in conflict since the less time an algorithm spends searching for neighbours, the more scalable it will be, and the worst its quality. Here, these issues are addressed by applying an approach- item-based algorithm. The bottleneck in this algorithm is the search for neighbour among a large user population of potential neighbours. Item-based algorithms avoid this bottleneck by exploring the relationship between users [2]. Context-aware recommender systems have been proven to improve the performance of recommendations in a wide array of domains and applications. Despite individual improvements, little work has been done...
on comparing different approaches, in order to determine which of them outperform the others, and under what circumstances. This issue is addressed by conducting an empirical comparison of several pre-filtering, post-filtering and contextual modeling approaches on the movie recommendation domain. To acquire confident contextual information, a user study is performed where participants were asked to rate movies, stating the time and social companion with which they preferred to watch the rated movies [3]. In the field of recommender systems (RSs), several scholars have shown that adding contextual information helps to improve performance of RSs. This scheme focuses on comparing the pre-filtering and the post-filtering approaches and identifying which method dominates the other and under which circumstances. Since there are no clear winners in this comparison, an alternative more effective method of selecting the winners is proposed in the pre- vs. the post-filtering comparison. This strategy provides analysts and companies with a practical suggestion on how to pick a good pre- or post filtering approach in an effective manner to improve performance of a context-aware recommender system [4]. Distributional-Semantics Pre-filtering (DSPF) is analogous to Generalized Pre-filtering: it is a reduction-based approach, but instead of searching for the optimal segmentation of the ratings, it exploits similarities between situations to generate segments that aggregate the ratings tagged with situations similar to the target one. Hence, the key difference is that this approach leverages the knowledge of the situation-to-situation similarity of contextual conditions instead of relying on the usually limited condition-to-condition hierarchical relationships defined in context taxonomy [5].

In this paper, the Domain-sensitive Recommendation Algorithm is used with other efficient algorithms which are Memory-based and Model-based. This will reduce the required time to find the predictions and will increase the accuracy of the recommendation systems.

II. BACKGROUND

Methods for generating context-aware recommendations were classified into the pre-filtering, post-filtering and contextual modeling approaches. In this scheme, a novel type of contextual modeling is proposed, that is called contextual neighbors, based on the idea of using context to compute the neighborhood in a collaborative filtering approach, and introduces four variants of this method. In addition, the scheme presents the results of the comparison among these four approaches and among the contextual neighbors approach to the other contextual approaches and to the un-contextual one. A novel Domain-sensitive Recommendation (DsRec) algorithm is proposed to make the rating prediction by exploring the user-item subgroup analysis simultaneously, in which a user-item subgroup is deemed as a domain consisting of a subset of items with similar attributes and a subset of users who have interests in these items. The proposed framework of DsRec includes three components: a matrix factorization model for the observed rating reconstruction, a bi-clustering model for the user-item subgroup analysis, and two regularization terms to connect the above two components into a unified formulation. Extensive experiments on Movielens-100K and two real-world product review datasets show that our method achieves the better performance in terms of prediction accuracy criterion over the state-of-the-art methods[1].

Different item-based recommendation generation algorithms are analysed in this approach. There are two fundamental challenges. First is to improve the scalability of the collaborative filtering algorithms. Second is to improve the quality of recommendations for the users. These two challenges are in conflict since the less time an algorithm spends searching for neighbours, the more scalable it will be, and the worst its quality. Here, these issues are addressed by applying an approach- item-based algorithm. The bottleneck in this approach is the search for neighbor among a large user population of potential neighbours. Item-based algorithms avoid this bottleneck by exploring the relationship between users in [2].

Context-aware recommender systems have been proven to improve the performance of recommendations in a wide array of domains and applications. Despite individual improvements, little work has been done on comparing different approaches, in order to determine which of them outperform the others, and under what circumstances. In this scheme, this issue is addressed by conducting an empirical comparison of several pre-filtering, post-filtering and contextual modeling approaches on the movie recommendation domain. To acquire confident contextual information, a user study is performed where participants were asked to rate movies, stating the
time and social companion with which they preferred to watch the rated movies in [3].
In the field of recommender systems (RSs), several scholars have shown that adding contextual information helps to improve performance of RSs. This scheme focuses on comparing the pre-filtering and the post-filtering approaches and identifying which method dominates the other and under which circumstances. Since there are no clear winners in this comparison, an alternative more effective method of selecting the winners is proposed in the pre- vs. the post-filtering comparison. This strategy provides analysts and companies with a practical suggestion on how to pick a good pre- or post filtering approach in an effective manner to improve performance of a context-aware recommender system is described in [4].
Distributional-Semantics Pre-filtering (DSPF) is analogous to Generalized Pre-filtering: it is a reduction-based approach, but instead of searching for the optimal segmentation of the ratings, it exploits similarities between situations to generate segments that aggregate the ratings tagged with situations similar to the target one. Hence, the key difference is that this approach leverages the knowledge of the situation-to-situation similarity of contextual conditions instead of relying on the usually limited condition-to-condition hierarchical relationships defined in context taxonomy and it is presented in [5].
The rest of the paper is organized as follows. In this paper, Section III provides work which is done previously, Section IV gives idea about existing technology, in Section V analysis and discussion about techniques is carried out, proposed methodology is explained in Section VI, Possible outcomes and Result is described in Section VII. Finally, Section VIII concludes the paper.

III. PREVIOUS WORKDONE
Jing Liu et al.(2016)[1] proposed In this approach, a novel Domain-sensitive Recommendation (DsRec) algorithm is proposed to make the rating prediction by exploring the user-item subgroup analysis simultaneously, in which a user-item subgroup is deemed as a domain consisting of a subset of items with similar attributes and a subset of users who have interests in these items. The proposed framework of DsRec includes three components: a matrix factorization model for the observed rating reconstruction, a bi-clustering model for the user-item subgroup analysis, and two regularization terms to connect the above two components into a unified formulation. Extensive experiments on Movielens-100K and two real-world product review datasets show that our method achieves the better performance in terms of prediction accuracy criterion over the state-of-the-art methods.
Badrul Sarwar et al. (2001)[2] In this approach, different item-based recommendation generation algorithms are analysed. There are two fundamental challenges. First is to improve the scalability of the collaborative filtering algorithms. Second is to improve the quality of recommendations for the users. These two challenges are in conflict since the less time an algorithm spends searching for neighbours, the more scalable it will be, and the worst its quality. Here, these issues are addressed by applying an approach- item-based algorithm. The bottleneck in this algorithm is the search for neighbor among a large user population of potential neighbours. Item-based algorithms avoid this bottleneck by exploring the relationship between users
Pedro G. Campos et al. (2013)[3] has proposed a scheme which is addressed by conducting an empirical comparison of several pre-filtering, post-filtering and contextual modeling approaches on the movie recommendation domain. To acquire confident contextual information, a user study is performed where participants were asked to rate movies, stating the time and social companion with which they preferred to watch the rated movies.
Michele Gorgoglione et al. (2009)[4], proposed scheme which focuses on comparing the pre-filtering and the post-filtering approaches and identifying which method dominates the other and under which circumstances. Since there are no clear winners in this comparison, an alternative more effective method of selecting the winners is proposed in the pre- vs. the post-filtering comparison. This strategy provides analysts and companies with a practical suggestion on how to pick a good pre- or post filtering approach in an effective manner to improve performance of a context-aware recommender system.
Victor Codina et al. (2015)[5] proposed a Distributional-Semantics Pre-filtering (DSPF) which is analogous to Generalized Pre-filtering: it is a reduction-based approach, but instead of searching for the optimal segmentation of the ratings, it exploits similarities between situations to
generate segments that aggregate the ratings tagged with situations similar to the target one. Hence, the key difference is that this approach leverages the knowledge of the situation-to-situation similarity of contextual conditions instead of relying on the usually limited condition-to-condition hierarchical relationships defined in context taxonomy.

IV. EXISTING METHODOLOGY
This section gives a brief overview of the methods and various approaches used for improvement in Context-Aware Recommendation Systems. If readers wanted to get more information may refer the references given.

A. Domain-sensitive Recommendation Algorithm
In this approach, a novel Domain-sensitive Recommendation (DsRec) algorithm is proposed to make the rating prediction by exploring the user-item subgroup analysis simultaneously, in which a user-item subgroup is deemed as a domain consisting of a subset of items with similar attributes and a subset of users who have interests in these items. The proposed framework of DsRec includes three components: a matrix factorization model for the observed rating reconstruction, a bi-clustering model for the user-item subgroup analysis, and two regularization terms to connect the above two components into a unified formulation. Extensive experiments on MovieLens-100K and two real-world product review datasets show that our method achieves the better performance in terms of prediction accuracy criterion over the state-of-the-art methods [1].

B. Item-based Recommendation Generation Algorithm
In this approach, different item-based recommendation generation algorithms are analysed. There are two fundamental challenges. First is to improve the scalability of the collaborative filtering algorithms. Second is to improve the quality of recommendations for the users. These two challenges are in conflict since the less time an algorithm spends searching for neighbours, the more scalable it will be, and the worst its quality. Here, these issues are addressed by applying an approach- item-based algorithm. The bottleneck in this algorithm is the search for neighbors among a large user population of potential neighbors. Item-based algorithms avoid this bottleneck by exploring the relationship between users [2].

C. Method using Contextual Signals
Time context and social context are the two contextual signals (i.e., the user’s current companion). Exploiting time context has been proved to be an effective approach to improve recommendation performance, e.g., Netflix Prize. Additionally, social context has also been found as a source for improving CARS performance. Time can be represented both as continuum information (e.g., current date/time), and as periodic, discrete information. When timestamps are available, both continuous and categorical context information can be extracted and exploited. Social context is a key factor for the users’ actions. One way to obtain social context signals is to take advantage of online social networks e.g. Facebook and Twitter, which help in social network-based recommender systems. However, the context information obtained in this way is used to find general preferences of related users, and generally does not correspond to the item usage/consumption context of the target user [3].

D. Comparison of Pre-filtering and Post-filtering
In this method, EPF (exact pre-filtering), Weight and Filter post-filtering methods are used. These methods are compared across two datasets and various other experimental settings for finding which method dominates the other. This implies that the comparison of the pre- and the post-filtering approaches depends very significantly on the type of the post-filtering method used. Pre-filtering, post-filtering and the uncontextual filtering methods are compared to choose better one [4].

E. Distribuitional Semantic Pre-filtering
Distribuitional-Semantics Pre-Filtering (DSPF) is a novel contextual, pre-filtering approach that tackles the data-sparsity problem of Context-Aware Recommender Systems which improves state-of-the-art CARS techniques by exploiting semantic similarities between contextual situations during local context modeling. DSPF is a reduction-based approach. It builds local rating prediction models trained with ratings collected in a target contextual situation and similar situations, i.e., with a similarity larger than a data-dependent similarity threshold. The experiment is carried out on six datasets. That shows DSPF performs better than state-of-the-art CARS techniques [5]. The rating estimation for user u and item I is given as:

\[
\hat{r}_{ui} = \mu + b_u + b_i + q_i^T p_u
\]
V. ANALYSIS AND DISCUSSION
A novel Domain-sensitive Recommendation algorithm is developed, which makes rating prediction assisted with the user-item subgroup analysis. DsRec is a unified formulation integrating a matrix factorization model for rating prediction and a bi-clustering model for domain detection. Additionally, information between these two components is exchanged through two regression regularization items, so that the domain information guides the exploration of the latent space. Systematic experiments conducted on three real-world datasets demonstrate the effectiveness of our methods. It is worth noting that this method is totally based on the user-item rating matrix [1].

<table>
<thead>
<tr>
<th>Techniques</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain-sensitive Recommendation Algorithm</td>
<td>1) Domain-sensitive Recommendation Algorithm makes it easy to find quality rating predictions assisted with the user-item subgroup analysis.</td>
<td>1) Both user-item interaction information and some external information simultaneously for domain detection are not discussed.</td>
</tr>
<tr>
<td>Item-based Recommendation Generation Algorithm</td>
<td>1) Improvement in quality is consistent. 2) The item neighborhood is fairly static.</td>
<td>1) As recommendations systems are becoming more crucial so in some cases it is difficult to find quality recommendations.</td>
</tr>
<tr>
<td>Method using Contextual Signals</td>
<td>1) Collecting time information of user interactions with a system does not require additional user efforts and it has been used as a key input for achieving significant improvements on recommendation accuracy.</td>
<td>1) The used dataset have a limited number of ratings, and experiments with a much larger dataset (and additional datasets) should be conducted, in order to test whether results obtained in this work are general or not.</td>
</tr>
<tr>
<td>Comparison of Pre-filtering and Post-filtering</td>
<td>1) Pre-filtering and post filtering are less time consuming in comparison to exhaustive comparison of most of the post filtering methods.</td>
<td>1) Comparison of post filtering methods with pre-filtering methods is a laborious and time consuming strategy.</td>
</tr>
<tr>
<td>Distributional Semantic Pre-filtering</td>
<td>1) DSPF approach supports a more flexible and effective aggregation of ratings and thus yields a better accuracy, especially when the contextual situations considered in the application are very specific. 2) Distributional-Semantics Pre-Filtering (DSPF), a novel contextual, pre-filtering approach that tackles the data-sparsity problem of Context-Aware Recommender Systems (CARSs).</td>
<td>1) DSPF needs to learn a local rating prediction model for each target contextual situation that the system may face. This means that, depending on the number of possible situations and the size of each local prediction model, DSPF can be more memory-demanding than other context-aware techniques where a unique global prediction model is needed. 2) The memory consumption may be significantly larger than other methods. 3) Another potential limitation of DSPF is the method used for computing the influence of contextual conditions and building the semantic vectors, which, by design, works only with explicit rating data.</td>
</tr>
</tbody>
</table>

Table 1: Comparison between existing methodologies.

From the experiments performed using the item-item collaborative filtering it is observed that this scheme provides better quality of prediction than the user-user scheme. The improvement in quality is consistent over different neighborhood size and training/test ratio. However, this improvement is not significantly large. The second observation is that the item neighborhood is fairly static, which can be potentially pre-computed, which results in very high online performance. Due to model based approach, it is possible to retain only a small subset of items and produce reasonably good prediction quality. It is observed that the item-based techniques hold the promise of allowing CF-based algorithms to scale to large datasets and same time produce high-quality recommendation [2].

In Contextual signals approach, diverse CARS have been compared, including various pre-
filtering, post-filtering and contextual modeling approaches. The results obtained in our experiments show that there is not a CARS clearly superior to others. It is observed that an Item Splitting pre-filtering using Matrix Factorization, as well as a Random Forest-based contextual modeling had a general good performance on the collected dataset, independently of the contextual information used, and thus, may represent good choices for the movie domain when different contextual signals are available. It is concluded that using all available context information does not have to be the best solution. The Item Splitting-based approach was able to properly deal with the combination of context signals. This approach has some limitation, the used dataset have a limited number of ratings, and experiments with a much larger dataset should be conducted [3].

In comparison of pre-filtering and post-filtering approach, the pre- and the post-filtering approaches are compared to generating contextual recommendations. In particular, the exact pre-filtering (EPF) is used for the former and the Weight and the Filter post-filtering methods for the latter approaches. These methods across two datasets and various other experimental settings are compared and showed that the Filter method dominates EPF and that EPF dominates the Weight method on 2 datasets. This implies that the comparison of the pre- and the post-filtering approaches depends very significantly on the type of the post-filtering method used [4].

In DSPF approach, Distributional-Semantics Pre-Filtering (DSPF), have described an novel contextual, pre-filtering approach that tackles the data-sparse problem of Context-Aware Recommender Systems (CARSs). DSPF improves state-of-the-art CARS techniques by exploiting semantic similarities between contextual situations during local context modeling. DSPF is a reduction-based approach. An experiment have been carried out on six contextually-tagged data sets shows that DSPF outperforms state-of-the-art CARS techniques when used in combination with a bias-based MF rating prediction model. The results show that this approach obtains better results in data sets. This method uses specific solutions to improve the reliability of distributional similarities. Potential limitation of DSPF is the method used for computing the influence of contextual conditions and building the semantic vectors, which, by design, works only with explicit rating data. The main difficulty of this fine-tuning procedure is finding the optimal thresholds and meta-parameters without over fitting the training data, especially in small-medium rating data sets. This is a major limitation of DSPF as in other local model techniques [5].

VI. PROPOSED METHODOLOGY
In this paper, a new approach is proposed which is the combination of the Domain-sensitive recommendation algorithm with the other two algorithms i.e. Memory-based and Model-based algorithms. In DsRec, user-item subgroup analysis allows a user or an item to appear in multiple subgroups. The goal of DsRec is to perform domain sensitive recommendation by jointly discovering user-item subgroups and predicting domain-specific user-item correlation, where only the user-item ratings are explored. The rating prediction model and domain detection model are integrated into a unified framework, to best use of the available rating data and make the both stages boost each other. In addition, DsRec could handle the data sparsity problem better. The real-world user-item rating data is used to empirically validate the effectiveness of proposed model for rating prediction. DsRec interpret how to detect user-item subgroups (domains) with a bi-clustering model, which is also a two-sided clustering solution. DsRec has many advantages. The time required for the rating predictions can be reduced by using Memory-based and Model-based algorithms. Memory-based collaborative algorithms utilize the entire user-item database to generate a prediction. These systems employ statistical techniques to find a set of users, known as neighbors, that have same history as the target user. Once the neighborhood of users is formed, these systems use different algorithms to combine the preferences of neighbors to produce prediction or top-N recommendation for active user. It is more popular and widely used. Model-based Collaborative filtering algorithms provide item recommendation by first developing a model of user ratings. The basic idea of CF-based algorithms is to provide item recommendations or predictions based on the opinions of the other like-minded users. Clustering techniques work by identifying groups of users who appear to have similar preferences. Once the clusters are formed, predictions for an individual can be made by averaging the opinions of the other users in that cluster.
VII. POSSIBLE OUTCOMES AND RESULT

The proposed approach is the combination of the Domain-sensitive Recommendation Algorithm and the Memory-based and Model-based algorithms. In this approach, the user history which is stored in database is used to find the rating predictions. After using the DsRec if further algorithms are used then the time required to find the rating predictions gets reduced. Thus the users will get predictions very quickly. These predictions include the user history and the opinions of likeminded users.

VIII. CONCLUSION

In the proposed approach, the Domain-sensitive Recommendation Algorithms and the Memory-based and Model-based algorithms are used together to reduce the time required to find the rating predictions. Memory-based algorithms made the use of the user databases for finding the users of similar interest. This increases the accuracy of recommendation system.

IX. FUTURE SCOPE

It can intend to analyze deeper different aspects to find the rating predictions. In particular, the future plan is to analyze the impact of different algorithms on the process of finding the rating predictions. Many collaborative filtering based algorithms can be used in future to improve such recommendation models.

REFERENCES


