Domain-Sensitive Recommendation With User-Item Subgroup Analysis

Rubini.B[1], Nagarajan[2]
[1]Dept of MCA
[2]Professor, Dept of MCA
[1, 2] Sri Muthukumaran Institute of Technology, Mangadu

Abstract- *Collaborative Filtering (CF) is one of the most successful recommendation approaches to cope with information overload in the real world. However, 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. Motivated by the observation, in this paper, we propose a novel Domain-sensitive Recommendation (DsRec) algorithm, 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.*

I. INTRODUCTION

Generally, data mining (sometimes called data or knowledge discovery) is the process of analysing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases. While large-scale information technology has been evolving separate transaction and analytical systems, data mining provides the link between the two. Data mining software analyzes relationships and patterns in stored transaction data based on open-ended user queries. Several types of analytical software are available: statistical, machine learning, and neural networks. Generally, any of four types of relationships are sought:

Classes: Stored data is used to locate data in predetermined groups. For example, a restaurant chain could mine customer purchase data to determine when customers visit and what they typically order. This information could be used to increase traffic by having daily specials.

Clusters: Data items are grouped according to logical relationships or consumer preferences. For example, data can be mined to identify market segments or consumer affinities.

Associations: Data can be mined to identify associations. The beer-diaper example is an example of associative mining.

Sequential patterns: Data is mined to anticipate behaviour patterns and trends. For example, an outdoor equipment retailer could predict the likelihood of a backpack being purchased based on a consumer's purchase of sleeping bags and hiking shoes.

The Existing recommender systems have been indispensable nowadays, which support users with possibly different judgments and opinions in their quest for information, through taking into account the diversity of preferences and the relativity of information value. 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. Besides, most of these clustering CF approaches are performed in a two-stage sequential process: domain detection by clustering and rating prediction by typical CF within the clusters.

II. LITERATURE SURVEY

1) Multi-domain collaborative filtering
AUTHORS: Y. Zhang, B. Cao, and D.-Y. Yeung

Collaborative filtering is an effective recommendation approach in which the preference of a user on an item is predicted based on the preferences of other users with similar interests. A big challenge in using collaborative filtering methods is the data sparsity problem which often arises because each user typically only rates very few items and hence the rating matrix is extremely sparse. In this paper, we address this problem by considering multiple collaborative filtering tasks in different domains simultaneously and exploiting the relationships between domains. We refer to it as a multi-domain collaborative filtering (MCF) problem. To solve the MCF problem, we propose a probabilistic framework which uses probabilistic matrix factorization to model the rating problem in each domain and allows the knowledge to be adaptively transferred across different domains by automatically learning the correlation between domains. We also introduce the link function for different domains to correct their biases. Experiments conducted on several real-world applications demonstrate the effectiveness of our methods when compared with some representative methods.

2) TopRec: Domain specific recommendation through community topic mining in social network
AUTHORS: X. Zhang, J. Cheng, T. Yuan, B. Niu, and H. Lu

Traditionally, Collaborative Filtering assumes that similar users have similar responses to similar items. However, human activities exhibit heterogenous features across multiple domains such that users own similar tastes in one domain may behave quite differently in other domains. Moreover, highly sparse data presents crucial challenge in preference prediction. Intuitively, if users' interested domains are captured first, the recommender system is more likely to provide the enjoyed items while filter out those uninterested ones. Therefore, it is necessary to learn preference profiles from the correlated domains instead of the entire user-item matrix. In this paper, we propose a unified framework, TopRec, which detects topical communities to construct interpretable domains for domain-specific collaborative filtering. In order to mine communities as well as the corresponding topics, a semi-supervised probabilistic topic model is utilized by integrating user guidance with social network. Experimental results on real-world data from Opinions and Ciao demonstrate the effectiveness of the proposed framework.

3) TCRec: Product recommendation via exploiting social-trust network and product category information
AUTHORS: Y. Jiang, J. Liu, X. Zhang, Z. Li, and H. Lu

In this paper, we develop a novel product recommendation method called TCRec, which takes advantage of consumer rating history record, social-trust network and product category information simultaneously. Compared experiments are conducted on two real-world datasets and outstanding performance is achieved, which demonstrates the effectiveness of TCRec.

4) RecTree: An efficient collaborative filtering method
AUTHORS: S. Han, S. Chee, J. Han, and K. Wang

Many people rely on the recommendations of trusted friends to find restaurants or movies, which match their tastes. But what if your friends have not sampled the item of interest? Collaborative filtering (CF) seeks to increase the effectiveness of this process by automating the derivation of a recommendation, often from a clique of advisors that we have no prior personal relationship with. CF is a promising tool for dealing with the information overload that we face in the networked world.

Prior works in CF have dealt with improving the accuracy of the predictions. However, it is still challenging to scale these methods to large databases. In this study, we develop an efficient collaborative filtering method, called *RecTree* (which stands for Recommendation Tree) that addresses the scalability problem with a divide-and-conquer approach. The method first performs an efficient k-means-like clustering to group data and creates neighborhood of similar users, and then performs subsequent clustering based on smaller, partitioned databases. Since the progressive partitioning reduces the search space dramatically, the search for an advisory clique will be faster than scanning the entire database of users. In addition, the partitions contain users that are more similar to each other than those in other partitions. This characteristic allows RecTree to avoid the dilution of opinions from good advisors by a multitude of poor advisors and thus yielding a higher overall accuracy.

5)Recommender systems for large-scale e-commerce: scalable neighborhood formation using clustering
AUTHORS: B. M. Sarwar, J. Konstan, and J. Riedl

Recommender systems apply knowledge discovery techniques to the problem of making personalized prod-uct recommendations during a live customer interaction. These systems, especially the k-nearest neigh-bor collaborative filtering-based ones, are achieving widespread success in E-commerce nowadays. The tremendous growth of customers and products in re-cent years poses some key challenges for recommender systems. These are: producing high quality recommen-dations and performing many recommendations per second for millions of customers and products. New recommender system technologies are needed that can quickly produce high quality recommendations, even for very large-scale problems.

III.PROPOSE IMPLEMENTATION OF SENSITIVE RECOMMENDATION

We propose a novel Domain-sensitive Recommendation (DsRec) algorithm, 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. Theproposed 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. There are three components in the unified framework. First, we apply a matrix factorization model to best reconstruct the observed rating data with the learned latent factor representations of both users and items, with which those unobserved ratings to users and items can be predicted directly. Second, a bi-clustering model is used to learn the confidence distribution of each user and item belonging to different domains. Actually, a specific domain is a user-item subgroup, which consists of a subset of items with similar attributes and a subset of users interesting in the subset of items. In the bi-clustering formulation, we assume that a high rating score rated by a user to an item encourages the user and the item to be assigned to the same subgroups together.

IV. MODULES

SYSTEM CONSTRUCTION MODULE:

In the first module, we develop the system with the entities needed to evaluate the effectiveness of our proposed system. We examine how DsRec behaves on the classic movie rating dataset Movie lens. 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. It is motivated by the assumption that the collaborative effect among users varies across different domains. To achieve such a goal, we design a unified framework with three components: the factorization model for rating prediction, the bi-clustering model for domain detection, and the regression regularization items as the bridge between the above two models.

MATRIX FACTORIZATION MODEL:

First, the typical matrix factorization model is adopted to find user-specific and item-specific latent factors to reconstruct the observable user-item ratings, and we can utilize the learned factors to predict the rating of any user item pair. Second, a bi-clustering model is formulated to make full use of the duality between users and items to cluster them into subgroups. The underlying assumption is that the labels of a user and an item for their subgroup identification should be the same if they are strongly associated, i.e., a high rated user-item pair should be grouped together. Third, the regression regularization attempts to learn the mappings from the latent factor representations of users (and items) to their confidence distribution belonging to different subgroups, where the former is learned from the factorization model, and the latter is explored in the biclustering model. Simply, the rating prediction model and the domain detection model are both estimated based on the observable user-item ratings. The regression terms are considered as a bridge between the both above models, in order to learn more discriminative latent spaces of users and items for recommendation and domain identification. From this view, the unified model is tightly integrated with the three models, and they enhance each other.

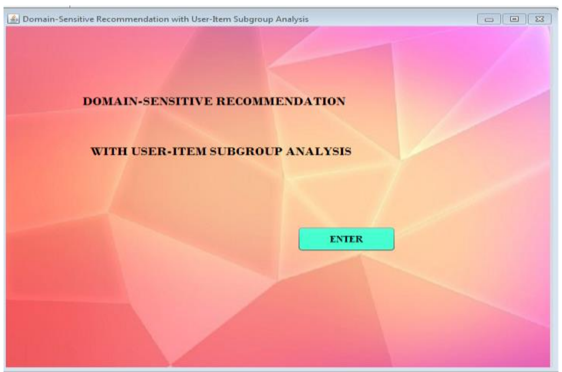
RATING PREDICTION MODEL:

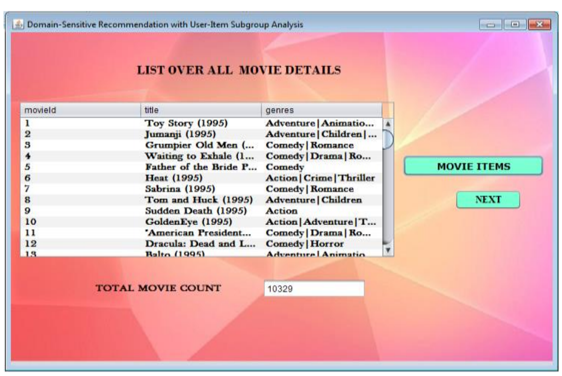
In this module, we develop the rating prediction model. As a typical solution, matrix factorization is adopted for rating prediction in our work. Suppose we have a user item rating matrix R 2 RN_M describing N users' numerical ratings on M items. Since in the real-world, each user usually rates a very small portion of items, the matrix R is extremely sparse. A matrix factorization approach seeks to approximate the rating matrix R by a multiplication of K-rank factors. However, due to the reason that R contains a large number of missing values, we only need to factorize the observed ratings in R.

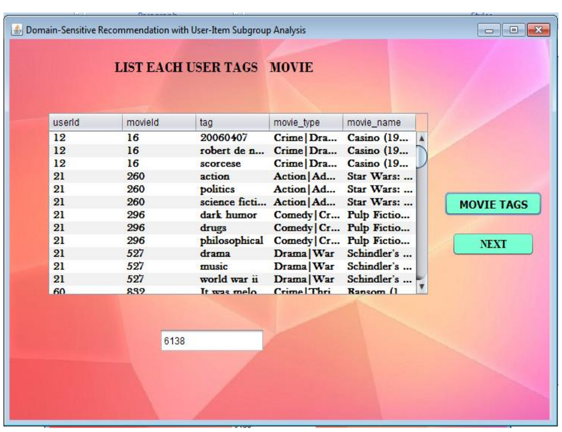
DOMAIN DETECTION MODEL:

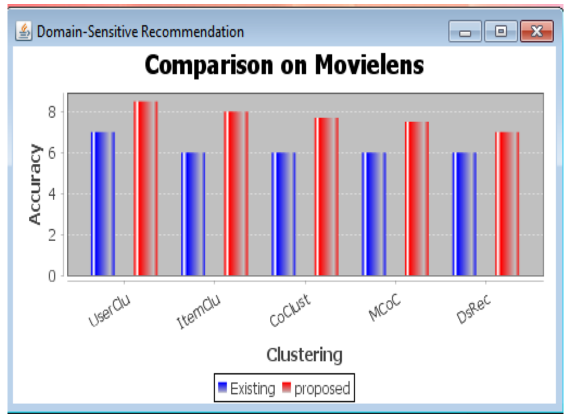
In this module, we will systematically interpret how to detect user-item subgroups (domains) with a bi-clustering model, which is also a two-sided clustering solution. It has been shown that the two-sided clustering often yields impressive performance over traditional one-sided clustering algorithms. More importantly, the resulting co-clustered subgroups may reveal valuable insights from the item attributes. In our framework, the domain detection model works on the assumption that a high rating score rated by a user to an item encourages the user and the item to be assigned to the same subgroups together. As a transitive result, the items highly rated by the same user are probably grouped together, and the same to the case that the users who prefers to the same item. Accordingly, we can obtain some user item subgroups each consistingof a subset of items and a group of users who interested in those items.

V. SCREENSHOTS









VI. ADVANTAGES AND DISADVANTAGES

ADVANTAGES

- Develop a novel Domain-sensitive Recommendation algorithm, 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.

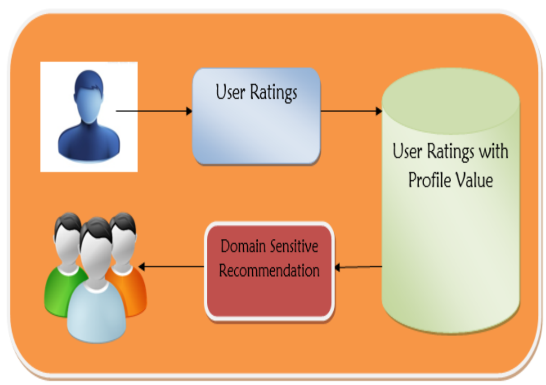


Fig.1.System Architecture of Proposed System

DISADVANTAGES

- The existing system has some problems which might limit the performance of typical CF methods.
- However, it is observed that this assumption is not always so tenable. Usually, the collaborative effect among users varies across different domains.
- However, such divide-and-conquer style brings a new problem, i.e., the algorithm cannot take full advantage of the observed rating data which is limited and precious

VII.CONCLUSION

In this project, we develop a novel Domain-sensitive Recommendation algorithm, which makes ratingprediction 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 are 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 our method is totally based on the user-item rating matrix. In the future, we will attempt to explore both user-item interaction information and some external information simultaneously for domain detection.

REFERENCES

[1] Y. Zhang, B. Cao, and D.-Y. Yeung, "Multi-domain collaborative filtering," in Proc. 26th Conf. Annu. Conf. Uncertainty Artif. Intell., 2010, pp. 725–732.

[2] X. Zhang, J. Cheng, T. Yuan, B. Niu, and H. Lu, "TopRec: Domain specific recommendation through community topic mining in social network," in Proc. 22nd Int. Conf. World Wide Web, 2013, pp. 1501–1510.

[3] Y. Jiang, J. Liu, X. Zhang, Z. Li, and H. Lu, "TCRec: Product recommendation via exploiting social-trust network and product category information," in Proc. 22nd Int. Conf. World Wide Web, 2013, pp. 233–234.

[4] S. Han, S. Chee, J. Han, and K. Wang, "RecTree: An efficient collaborative filtering method," in Proc. 3rd Int. Conf. Data Warehousing Knowl. Discovery, 2001, pp. 141–151.

[5] B. M. Sarwar, J. Konstan, and J. Riedl, "Recommender systems for large-scale e-commerce: scalable neighborhood formation using clustering," in Proc. 5th Int. Conf. Comput. Inf. Technol., 2002, pp. 1–6.

[6] G.-R. Xue, C. Lin, Q. Yang, W. Xi, H.-J. Zeng, Y. Yu, and Z. Chen, "Scalable collaborative filtering using cluster-based smoothing," in Proc. Res. Develop. Inf. Retrieval, 2005, pp. 114–121.

[7] A. Bellogin and J. Parapar, "Using graph partitioning techniques for neighbour selection in user-based collaborative filtering," in Proc. 6th ACM Conf. Recommender Syst., 2012, pp. 213–216.

[8] M. O'Connor and J. Herlocker, "Clustering items for collaborative filtering," in Proc. ACM SIGIR Workshop Recommender Syst., 1999, pp. 1–4.

[9] L. Ungar, D. Foster, E. Andre, S. Wars, F. S. Wars, D. S. Wars, and J. H. Whispers, "Clustering methods for collaborative filtering," in Proc. Workshop Recommendation Syst., 1998, pp. 114–129.

[10] L. Si and R. Jin, "Flexible mixture model for collaborative filtering," in Proc. 20th Int. Conf. Mach. Learn., 2003, pp. 704–711.

[11] A. Banerjee, I. Dhillon, J. Ghosh, S. Merugu, and D. S. Modha, "A generalized maximum entropy approach to Bregman co-clustering and matrix approximation," in Proc. 10th ACM KDD, 2004, pp. 509–514.