Subgroup Analysis Based On Domain Sensitive Recommendation

R.ESWARI[1],K.KANMANI[2],S.SHAHINA BEGUM[3],P.VENNILA[4], Ms.SUMITHRA.R[5]

[1,2,3,4] students member Computer Science and Engineering,
[5]Assistant professror,
Dept of CSE,M.R.K.Institute of Technology, Kattumannar Koil.

Abstract- *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. 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. In existing we had taken movie data and analysis subgroup analysis in our proposed system we had taken ,Multiple product items and analysis subgroup analysis.*

Keywords: *Matrix factorization, user-item subgroup, collaborative filtering.*

I. INTRODUCTION

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 pre- dictions, 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 [1].

Last decades have witnessed the overwhelming supply of online information with the evolution of the Internet. Thus, 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. Numerous efforts have been paid on this direction. Generally, these efforts can be divided into two types. The first type is to discover domains with the help of external information such as social trust network [2], product category information [3], etc.

In this paper we focus on the second type called clustering CF, which only exploits the user-item interaction information and detects the domains by clustering methods. Among algorithms of this type, some are one- side clustering in the sense that they only consider to cluster either items or users [4], [5], [6], [7], [8]. And others are two- side clustering, which make use of the duality between users and items to partition both dimensions simultaneously [9], [10], [11], [12], [13].

In most of clustering CF approaches, each user or item is assigned to a single cluster (domain). However, in reality, the user interests and item attributes are not always exclusive, e.g., a user likes roman- tic movies does not means the user does not like other genre movies, and a romantic movie could also be an war movie. Thus, it is more natural to assume that a user or an item can join multiple domains.

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. One advantage of this approach is to overcome the problem of scalability brought by many memory-based CF techniques where the heavy computational burden is brought by the similarity calculations. 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.

II. RELATED WORK

Y. Zhang, B. Cao, and D.-Y. Yeung proposed that 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. To solve the MCF problem, we 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

Zhang, J. Cheng, T. Yuan, B. Niu, and H. Lu have revealed 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 .we propose TopRec, which detects topical communities to construct interpretable domains for domain-specific collaborative filtering. Experimental results on real-world data from Epinions and Ciao demonstrate the effectiveness of the proposed framework.

Jiang, J. Liu, X. Zhang, Z. Li, and H. Lu reviewed to develop a novel product recommendation method called TCRec, which takes advantage of consumer rating historyrecord, 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 TCR

Han, S. Chee, J. Han, and K. Wang have suggested Many people rely on the recommendations of trusted friends to find restaurants or movies, which match their tastes..CF is a promising tool for dealing challenging to scale these methods to large databases. In this study, we develop an RecTree (which stands for RECommendation Tree) that addresses the scalability problem with a divide- and-conquer approach. 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. Based on our experiments and performance study, RecTree outperforms the well-known collaborative filter, CorrCF , in both execution time and accuracy

B. M. Sarwar, J. Konstan, and J. Riedl have suggested Recommender systems apply knowledge discovery techniques to the problem of making personalized product recommendations during a live customer interaction. These systems, especially the k-nearest neighbor collaborative filtering based ones, are achieving widespread success in E-commerce nowadays. These are producing high quality recommendations and performing many recommendations per second for millions of customers and products. We address the performance issues by scaling up the neighborhood formation process through the use of clustering techniques.

G.-R. Xue, C. Lin, Q. Yang, W. Xi, H.-J. Zeng, Yu, and Z.Chen have provided Memory-based approaches for collaborative filtering identify the similarity between two users by comparing their ratings on a set of items. In the past, the memory-based approaches have been shown to suffer from two fundamental problems: data sparsity and difficulty in scalability. In our approach, clusters generated from the training data provide the basis for data smoothing and neighborhood selection. As a result, we provide higher accuracy as well as increased efficiency in recommendations. Empirical studies on two datasets (EachMovie and MovieLens) show that our new proposed approach consistently outperforms other state-of-the-art collaborative filtering algorithms. Categories and Subject Descriptors.

III. METHODOLOGY

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. 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 item can be predicted directly.

The proposed system is divided into four Modules:

A) Data collection
B) Rating Prediction
C) Collaborative filtering
D) Bi-clustering

A.DATA COLLECTION

Product items dataset is collected through the help web site. the items have been divided into a few fixed categories. It consists of 263776 ratings (1-5) from 8351 users on 84652 product items and each user has rated at least 20 items. yelp are well-known consumer opinion websites where users can assign their familiar products integer ratings from 1 to 5. The two datasets used in this study are published by the authors of including data records until May 2011.Note that the original yelp dataset consists of 8351users who have rated on 84652different items, To build a compact and informative dataset for model learning, we expect to maintain those active users and popular items in original dataset. Specifically, we first remove the users who rate less than 10 items and then remove the items which has less than 10 ratings by the users. Thus we obtain a yelp subset whose detailed statistics. An alternate optimization scheme is developed to solve the unified objective function, and the experimental analysis on three real-world datasets demonstrates the effectiveness of our method. The real-world product items review datasets show that our method achieves the better performance in terms of prediction accuracy criterion over the state-of- the-art methods

B.RATING PREDICTION

Rating prediction in our work. Suppose we have a user item rating matrix describing N user's 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, 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. 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. that domain detection can improve the rating prediction accuracy.

C.COLLABORATIVE FILTERING

Collaborative filtering approach based on a weighted co-clustering algorithm. this method generates predictions based on the average ratings of the co-clusters (user-item neighborhoods) while taking into account the individual biases of the users and items. Collaborative Filtering (CF) is an effective and widely adopted recommendation approach. Different from content-based

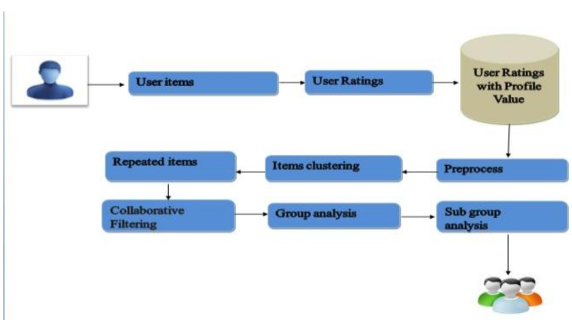
recommender systems which rely on the profiles of users and items for predictions, CF approaches make predictions by only utilizing the user-item interaction. A cluster is a collection of data objects that are similar to one another within the same cluster and are dissimilar to the objects in other clusters. Thus, In essence, the task of clustering approach in clustering CF is to discover domains. Recently, with the development of internet, various contextual information as well as the rating matrix are integrated to discover some meaningful domains where the typical contexts include item attributes, user trust.

D.BI-CLUSTERING

A bi-clustering model for the user-item subgroup analysis, and two regularization terms to connect the above two components into a unified formulation. 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.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, bi-clustering model for domain detection, 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. ARCHITECTURE



V. SCREENSHOTS

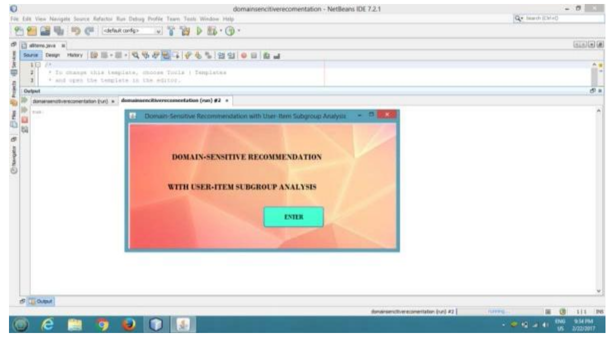


Fig1. login page

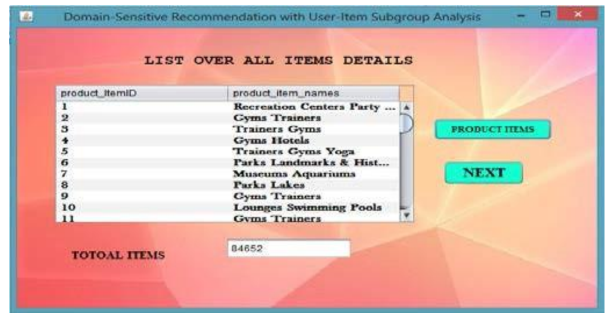


Fig2.Listing overall items

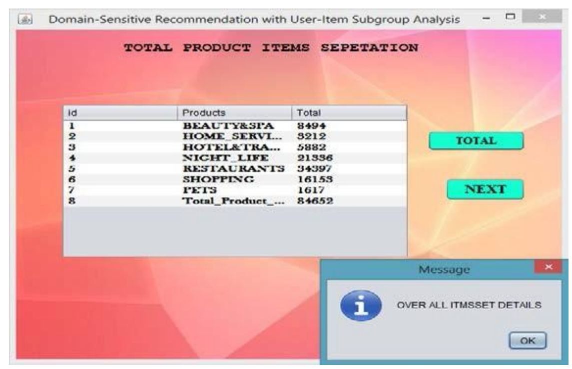


Fig3.Separation of product items

Fig4 .sub-items of product

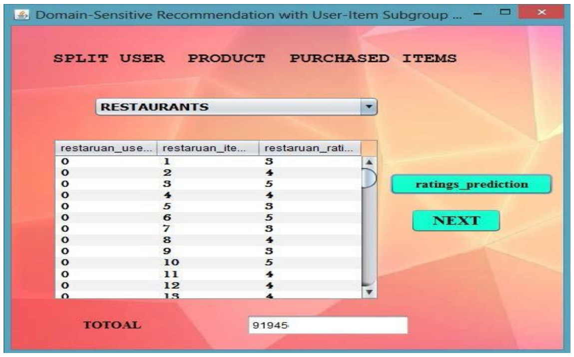


Fig5. Split user product purchased items

Fig6. User rating

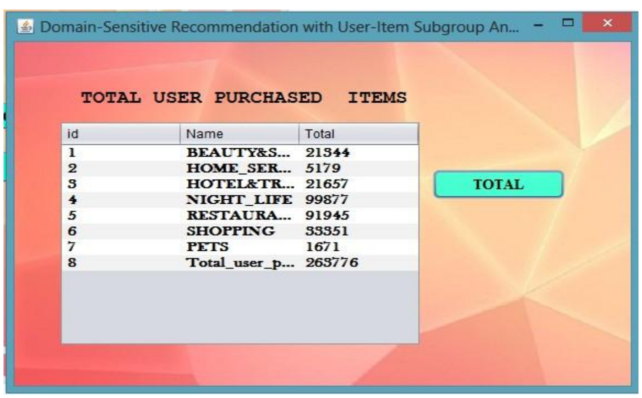


Fig7. Total user purchased items

VI. COMPARITIVE MEASURES

For each dataset, we use different observed data divisions (20, 50 and 80 percent) in our experiments. Training data 80 percent, for example, means we randomly select 80 percent of observed ratings from user-item rating matrix as the training data to predict the remaining 20 percent ratings. We also set different latent factor dimension (K) to test the matrix factorization methods. 10 random divisions of observed ratings are carried independently, and the average Results are reported.

VII. CONCLUSION

The user-item subgroup analysis in multiple product item dataset ,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 proposed three components : a matrix factorization model for the observed rating reconstruction, a bi-clustering model for the user-item subgroup analysis, by considering multiple collaborative filtering for Multiple product items and analysis subgroup analysis.

VIII. FUTURE ENHANCEMENT

In Existing technology is proceeded without the subgroups ,which can be done for movie based ratings. Now, we were proposed our work with subgroup analysis which is field of online products. This would be useful for vast technologies.

IX. REFERENCE

[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]. Zhang, J. Cheng, T. Yuan, B. Niu, and H.Lu,"TopRec:Domainspecific 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. G.-R. Xue, C. Lin, Q. Yang, W. Xi, H.-J.

[6]. Zeng, Y. Yu, and Z. Chen,"Scalable collaborative filtering using cluster-based smoothing,"in Proc. Res. Develop. Inf. Retrieval, 2005,pp. 114–21.

[7]. 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.
[12]. T. George and S. Merugu, "A scalable collaborative filtering framework based on co-clustering," in Proc. 5th IEEE Int. Conf. Data Mining, 2005, pp. 625–628.
[13]. B. Xu, J. Bu, C. Chen, and D. Cai, "An exploration of improving collaborat ive recommender systems via user-item subgroups," in Proc. 21st Int. Conf. World Wide Web, 2012, pp. 21–30.
[14]. J. S. Breese, D. Heckerman, and C. M. Kadie, "Empirical analysis of predictive algorithms for collaborative filtering," in Proc. 14th Conf. Uncertainty Artif. Intell., 1998, pp. 43–52.
[15]. J. L. Herlocker, J. A. Konstan, A. Borchers, and J. Riedl, "An algorithmic framework for performing collaborative filtering," in Proc.22nd Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 1999, pp. 230–237.