Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
July 2014, 21(Suppl. 1): 20–25
www.sciencedirect.com/science/journal/10058885
The Journal of China Universities of Posts and Telecommunications
http://jcupt.xsw.bupt.cn

Collaborative recommendation based on social community detection

LIU Xin[1] (✉), E Hai-hong[1], TONG Jun-jie[2], SONG Mei-na[1]

1. School of Computer, Beijing University of Posts and Telecommunications, Beijing 100876, China
2. Network Technology Research Institute, China Unicom, Beijing 100048, China

Abstract

Collaborative filtering algorithms have become one of the most used approaches to provide personalized services for users to deal with abundance of information. The traditional algorithms just use the explicit user-item rating matrix to find similar users or items. To improve the accuracy of the ratings predicted by the collaborative filtering algorithms, social information is widely incorporated into the traditional ones. Different with the existed works focus on directly connected neighbors, we consider the community between the users. We design the algorithms in two aspects: one is that the members in the same community have similar tastes and preferences, the other is that the member's taste is affected by the other members. We simplify these two factors as community similarity and community affection. Community similarity is incorporated into modifying the model-based collaborative filtering algorithm as the social community-based regularization (SCR), which improves 6.2% in mean absolute error (MAE) and 6.1% in root mean square error (RMSE) compared to the existed social recommendation algorithm. Community affection is incorporated into modifying the neighborhood-based collaborative filtering algorithm as the neighbor-based collaborative filtering based on community detection (NCFC) which improve 14.8% in MAE and 8.1% in RMSE compared to user-based collaborative filtering (UCF).

Keywords collaborative filtering, social community, global neighbor, matrix recommendation, social regularization

1 Introduction

The explosive growth and development of web and information technologies have led to the problem of information overload–the overwhelming plethora of choices and options available to a user, often varying in quality. To deal with abundance of information, personalized recommendation is a powerful tool which enables users to be presented information suiting his interests. And many recommendation systems (RSs) have been built. Generally, there are two variants of recommendation approaches: content-based and collaborative-filtering (CF) based approaches [1]. CF approaches can be further grouped into model-based CF (MCF) and neighborhood-based CF (NCF) [2]. Model-based approaches use user-item ratings to learn a predictive model. The general idea is to model the user-item interactions with factors representing latent features of users and items in the system, such as the preference class of users and the category class of items. In contrast, NCF approaches use user-item ratings stored in the system to directly predict ratings for new items.

Online social network (OSN) present new opportunities as to further improve the accuracy of RSs. In real life, people often resort friends in their social networks for advice before purchasing a product or consuming a service. Findings in the fields of sociology and psychology indicate that humans tend to associate and bond with similar others, also known as homophily [3]. Due to stable and long-lasting social bindings, people are more willing to share their personal opinions with their friends, and typically trust recommendations from their friend more than those from strangers and vendors.

How to utilize social network information has been widely studied in many research tasks. And numerous

Received date: 10-05-2014
Corresponding author: LIU Xin, E-mail: liuxinjackie@bupt.edu.cn
DOI: 10.1016/S1005-8885(14)60517-3

social matrix-factorization based RSs have recently been proposed as to improve recommendation accuracy. In Ref. [4], in order to predict review quality, Lu et al. proposed a generic framework for incorporating social context information by adding regularization constraints to the text-based predictor. In Ref. [5], Zhou et al. designed a matrix factorization objective function with social regularization. And it had two intuitions, one is that every user's taste is close to the average taste of this user's friends and the other one is the tastes may propagate between friends. However, due to data sparsity, the number of commonly rated items between friends could be very small or even zero. To address this problem, authors [6] improved the prediction accuracy by employing adaptive social similarities in the social regularization part. In Ref. [7], the authors improved the objective function in the social matrix factorization in three aspects: learning user similarity, directly learning of user-to-user information diffusion and learning restrict interests which means considering the users' similar preference in specific areas. In Ref. [8], the authors established a joint friendship-interest propagation (FIP) model that leverages both interest and friendship information to one unified framework. And FIP model bridges collaborative filtering in recommendation systems and random walk in social network analysis with a coupled latent factor model.

Besides incorporating the social information with matrix-factorization methods, there are many other methods. In Ref. [9], the authors proposed an architecture for social application recommendation with unifying three factors, which includes the popularity and reputation of the applications, the users' preference and the social similarity and social interaction based on human behavior analysis. Three prediction results are combined at last by a back-propagation neural network model. In Ref. [10], the authors proposed a collaborative ranking model for recommending useful tweets to twitter users with carefully designing features from the users' preference information.

The above methods improve on the accuracy of the traditional CF algorithms by taking social interest and social trust directly connected users in an OSN as additional inputs without considering the community or group formed in OSN. OSNs have the clustering or community property and there are two intuitions. One is that nodes of the same community are highly similar while on the contrary, nodes across communities present low similarity. And the other one is the social influence can change the users' preference.

In this paper, we aim to improve accuracy and performance of CF algorithms by community detection of OSN. And we use the similarity between the members of the same community to modify MCF and use the influence between the nodes to modify NCF. And we also conduct the experiments on the real world dataset to invalidate these two algorithms.

2 Overview of the two algorithms

2.1 Problem definition

We assume that $U = \{u_1, u_2, ..., u_m\}$ and $I = \{1, 2, ...n\}$ are the set of users and items respectively. And m is the number of users while n is the number of items. Here, we also have the social information from OSN. And we use $G = (U, E)$ to denote the network which is formed by the users. And the vertices in G are all the users and if user i have followed user j there will be an directed edge $(u_i, u_j) \in E$. We also have the explicit history ratings and $\boldsymbol{R}$ is the rating matrix. We use r_{ij} to denote the rating of user u_i on item j. And $\boldsymbol{R}$ is often sparse in practical scenarios.

Our goal is to, based on G and $\boldsymbol{R}$, to predict the value of missing ratings in $\boldsymbol{R}$ as accurately as possible while maintaining other characteristics such as high successful rate and coverage.

2.2 The main processes of the two algorithms

We modify the traditional CF algorithms by considering similarity and influence in communities in OSN. And form the NCFC and SCR. And the main processes of the two methods are shown in Fig. 1. In NCFC, we consider the influence between the members of the same and different communities and form the global similarity between any two users to enhance the accuracy of the traditional methods based on the local similarity based on the ratings. In SCR, we consider the similarity between the members of the same community and set the regularization to minimize the difference between them.

Both of the methods are based on the community detection, to simplify the processes, we use the community detection algorithms proposed in Ref. [11] to find the communities in the directed social network G.