



Collaborative filtering with social regularization for TV program recommendation



Ya Zhang*, Weiyuan Chen, Zibin Yin

Shanghai Key Lab. of Digital Media Processing & Transmissions, Shanghai Jiao Tong University, Shanghai 200240, China

ARTICLE INFO

Article history:

Received 5 February 2013

Received in revised form 23 September 2013

Accepted 23 September 2013

Available online 9 October 2013

Keywords:

TV program recommendation

Microblog

Recommender system

Collaborative filtering

Social regularization

Item similarity

ABSTRACT

In recent years, we have witnessed the explosive growth of microblogging services. As a popular platform for users to communicate and share information with friends, microblog has opened up new opportunities for recommendation. In this paper, we explore the possibility of recommending TV programs with microblogs. In particular, we leverage the following two important features of microblogs: (1) the rich user generated content reveals users' preferences on TV programs as well as the properties of TV programs and (2) the social interactions of the users suggest the mutual influences among the users. Taking into consideration of the above two properties, we proposed a hybrid recommendation model based on probabilistic matrix factorization, a popular collaborative filtering method. Two regularizers are added during matrix factorization: the social regularizer and the item similarity regularizer. We validate the proposed algorithm with Sina Weibo data set for TV program recommendation. The experimental results show that the proposed algorithm significantly outperforms the state-of-the-art collaborative filtering method, demonstrating the importance of incorporating social trust and item similarity in recommendation. In addition, we show that the proposed method is robust in recommending to new users, a typical cold-start scenario.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

In recent years, the dramatic increases in the number of TV programs have made the accessibility to the TV programs a challenging issue. Recommendation plays an important role in enhancing users' ability to access relevant information. Over the last several years, many recommendation systems have been proposed for recommending TV programs according to users' TV viewing behaviors [6,22,3]. The explosive growth of microblog services such as Twitter¹ and Weibo² has opened up new opportunities for improving the accuracy of TV program recommendation. As popular platforms for users to communicate with friends and share information, microblog services generates huge amount of content every day. The rich content helps form a better understanding of the preferences of the users as well as the characteristics of the items. Moreover, information about users' social connections and their mutual interactions are explicit with microblog services. All the above information may be leveraged by recommender systems to improve the accuracy of predictions. In this paper, we explore the possibility of leveraging user social interactions as well as user generated content at microblogging sites to improve the accuracy of TV programs recommendation.

* Corresponding author. Tel.: +86 21 34204468; fax: +86 21 34204155.

E-mail address: ya_zhang@sjtu.edu.cn (Y. Zhang).

¹ <http://www.twitter.com>.

² <http://weibo.com>.

Two widely used approaches in recommender systems are content-based filtering [13] and collaborative filtering [23]. Content-based filtering methods are to recommend items similar to the ones that a user liked in the past. Content-based filtering requires knowledge of the characteristics of the items, which is not always available. On the contrary, the collaborative filtering approaches do not rely on the characteristics of the items. User-based collaborative filtering methods assume that similar users tend to share similar preferences on items. By collecting and analyzing a large amount of information on users' behaviors, collaborative filtering methods predict what users would like based on similarity among the users or items in terms of the observed behaviors. Collaborative filtering approaches often suffer from three problems, cold start, data sparsity, and scalability. In fact, hybrid approaches combining collaborative filtering and content-based filtering have been demonstrated to be able to relief the above problems to some extent. See [1] for a complete review of recommender systems.

When leveraging the microblog services for recommendation, it is necessary to note that users' social network and their respective preference form a dynamic equilibrium. People of similar interests and preferences are more likely to be connected and interact with each other. On the other hand, one tends to be influenced by his/her social connections in many aspects including interests and preferences. For example, it is more likely for an individual to watch a new TV program if it has been watched by many of his/her friends. In fact, users' interests and friendship networks are

revealed to be highly correlated [24]. Hence, a user's social ties and interactions embed important information about the user's preferences. Considering the following–follower relationship is very noisy, we here measure the similarity in users' interests using their social interactions in terms of retweeting behaviors. Another advantage of microblog platforms is the large volume of content generated by users. For example, when individuals tweets about a TV program, the names of its characters, the names of the actors and actress, as well as the plot of the play are likely to be mentioned recurrently. Such content makes it possible for us to learn about the properties of TV programs, which is expected to contribute positively to TV program recommendation.

Given the availability of user interaction data and user generate content on the microblog platforms, we attempt to leverage them to improve recommendation accuracy on top of Probabilistic Matrix Factorization (PMF), a model-based collaborative filtering method. The similarity of user interests is measured based on the social interactions among users on a microblogging platform. Latent Dirichlet Allocation (LDA) algorithm is employed to derive topic models from the user generated content, based on which we measure the similarity of items. By incorporating the two types of similarity in probabilistic matrix factorization through regularization, we expect to partially relief the data sparsity problem and the cold start problem. Specifically, two regularizers are added to the objective function of probabilistic matrix factorization. The first regularizer requires friends of similar interests as measured by social interacts to have similar user profile. The second regularizer requires items associated with similar topic models to share similar item profiles. It is worth noting that the regularization may also be applicable to the objective function of other model-based collaborative filtering methods.

We validated the proposed algorithm with the task of recommending TV programs using data retrieved from Sina Weibo, the largest microblog site in China. The experimental results indicate that the proposed algorithm significantly outperforms the state-of-the-art-algorithms, demonstrating the effectiveness of the two regularizers. We also test the proposed algorithm under data sparsity and cold start scenarios and show that it is more robust than probabilistic matrix factorization in dealing with these problems. Finally, the methods are compared in terms of the coverage and popularity of the recommended items. The proposed method is shown to keep a good balance of coverage and popularity by incorporating the two regularizers.

The rest of the paper is organized as follows. In Section 2, we summarize the related work and background of major approaches for recommender systems. In Section 3, we introduce the definition and notations of the problem. Section 4 briefly reviewed the probabilistic matrix factorization method. The proposed algorithm and the learning method are shown in Section 5. Section 6 introduces the data set we used to validate the proposed algorithm. The experimental results and discussion are presented in Section 7. Finally, we conclude in Section 8.

2. Related work

Recommender systems have been widely studied and implemented in recent years, both in academia and in industry. As one of the most successful approaches to building recommender systems, collaborative filtering uses the preferences of a group of users to make recommendations or predictions of users' preferences on unrated items.

There are three main categories of collaborative filtering techniques: memory-based, model-based, and hybrid collaborative filtering algorithms [23]. Memory-based collaborative filtering algorithms are based on the entire collection of previously rated

items by the users. The prediction on the rating R_{ui} of the user u to the item i is obtained by identifying the N most similar neighbors of a user or an item. Memory-based collaborative filtering algorithms suffers significantly from the data sparsity problem. To overcome the drawback of memory-based collaborative filtering algorithms, model-based collaborative filtering algorithms develop a model of user ratings, thus enabling the system to learn complex patterns based on training data. In general, classification algorithms are used when the user ratings are categorical and regression models or Singular Value Decomposition (SVD) based methods are used for numerical ratings.

There are many studies on model-based collaborative filtering algorithms. Koren [7] merged the factor and neighborhood models and build a more accurate combined model to exploit both explicit and implicit feedback by the users. Paterek [15] tried several methods: addition of biases to the regularized SVD, postprocessing SVD with kernel ridge regression, using a separate linear model, and using methods similar to the regularized SVD. They propose a framework for combining them to obtain a good prediction. Gu et al. [5] proposed a unified model for collaborative filtering based on graph regularized weighted nonnegative matrix factorization. The proposed method not only inherits the advantages of model-based method, but also owns the merits of memory-based method which considers the neighborhood information. It has the ability to make use of content information and any additional information regarding user–user such as social trust network. Zhen et al. [26] proposed a novel framework, called tag informed collaborative filtering which seamlessly integrate tagging information into the collaborative filtering procedure.

By modeling the actual choice process in recommender systems (i.e. users are presented with a few number of items for rating at a time), Yang et al. [25] proposed the collaborative competitive filtering framework for learning user preferences. Collaborative competitive filtering employs a multiplicative latent factor model to characterize the dyadic utility function. Unlike collaborative filtering, collaborative competitive filtering models the user behavior of choices by encoding a local competition effect.

Latent Dirichlet Allocation (LDA) is a three-level hierarchical Bayesian model which has been used in recommender systems [16,18]. Purushotham et al. [17] proposed a hierarchical Bayesian model to integrate social network structure (using matrix factorization) and item content-information (using LDA model) for item recommendation. Liu et al. [9] proposed an interest expansion strategy via personalized ranking based on the topic model for building an interest-oriented collaborative filtering framework. Agarwal et al. [2] proposed a novel matrix factorization method to predict ratings in recommender system applications. To avoid overfitting, user and item factors in this model are regularized through Gaussian linear regression and LDA priors respectively. Krestel et al. [8] introduced an approach based on LDA for recommending tags of resources in order to improve search.

Furthermore, traditional recommender systems assume that all the users are independent and identically distributed; this assumption ignores the connections among users, which is not consistent with the real-world observations where we always turn to our trusted friends for recommendations. In recent years, several studies attempted to exploit trust in making recommendations. A hybrid trust-based multi-criteria recommendation approach was proposed to handle business partner matching e-services, which integrates trust-based filtering with the multi-criteria collaborative filtering [19,20]. Within the framework of collaborative filtering, Shambour and Lu [21] proposed a TrustSemantic Fusion (TSF)-based recommendation approach which leverages information from the users' social trust network and the items' semantic domain knowledge to alleviate the problem of data sparsity and cold-start in recommendation. The approach was validated with

Download English Version:

<https://daneshyari.com/en/article/6862768>

Download Persian Version:

<https://daneshyari.com/article/6862768>

[Daneshyari.com](https://daneshyari.com)