Neurocomputing 000 (2018) 1–10

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Sparse latent model with dual graph regularization for collaborative filtering

Xiaodong Feng[a], Sen Wu[b,*], Zhiwei Tang[a], Zhichao Li[a]

[a] School of Public Administration, University of Electronic Science and Technology of China, Chengdu, 611731, Sichuan, China
[b] Donlinks School of Economics and Management, University of Science and Technology Beijing, Beijing 100083, China

ARTICLE INFO

Article history:
Received 8 July 2017
Revised 25 November 2017
Accepted 8 January 2018
Available online xxx

Communicated by Dr Xin Luo

Keywords:
Collaborative flittering
Matrix factorization
Sparse representation
Graph Laplacian
Iterative optimization

ABSTRACT

Matrix factorization (MF) has been one of the powerful machine learning techniques for collaborative flittering, and it is also widely extended to improve the quality for various tasks. For recommendation tasks, it is noting that a single user or item is actually shown to be sparsely correlated with latent factors extracted by MF, which has not been developed in existing works. Thus, we are focusing on levering sparse representation, as a successful feature learning schema for high dimensional data, into latent factor model. We propose a Sparse LAtent Model (SLAM) based on the ideas of sparse representation and matrix factorization. In SLAM, the item and user representation vectors in the latent space are expected to be sparse, induced by the ℓ_1-regularization on those vectors. Besides, we extend a dual graph Lapalacian regularization term to simultaneously integrate both user network and item network knowledge. Also, an iterative optimization method is presented to solve the new learning problem. The experiments on real datasets show that SLAM can predict the user–item ratings better than the state-of-the-art matrix factorization based methods.

1. Introduction

Recommender Systems that automatically suggest particular items to users have had a rising presence with the rapid growth of Web applications on e-commerce and thus attracted a lot of attention from researchers in the past decade. Among the recommendation techniques, collaborative filtering [1] that predicts the user's preference about items by exploiting information from other similar users or items using user–item matrix (binary or real matrix), is very effective and has been widely used in practice.

Methods of collaborative filtering mainly include memory-based [2,3] and model-based approaches, where model-based methods have been illustrated to achieve better performance especially on large tasks. The low rank matrix factorization (MF) [4] that approximates the user–item-rating matrix with a low rank matrix is one of the popular model-based methods. Besides, SVD++ [5] is also a form of matrix factorization that takes into consideration user/item biases and the influence of rated items rather than user/item-specific vectors on rating prediction, while MF+ [6] focuses on incorporating rich features, such as visual features. Matrix factorization can also be interpreted as a latent factor model that each user or item lies in a latent feature space where the feature correspond to a factor, or probabilistic graphic model [7] that assumes each rating is determined by a conditional distribution of the corresponding user and item.

Besides the traditional collaborative filtering methods that only use the user–item rating data, social network-aware recommendation [8,9] is being developed to address the issues of cold start users that will severely degrade quality of recommendations. Matrix factorization has also been extended in ways to integrate the social network information, such as trust network between users [10], which outperforms methods without this information a lot. However, the improved performance owes a good deal to more information using, but the matrix factorization model does not help it, which can be obviously understood that utilizing more external information do can improve the performance. However, the performance is still bounded by the benefits of matrix factorization and what is exploited for collaborative filtering. Therefore, it comes to a question that whether we can continuously improve the performance of matrix factorization-based collaborative filtering by developing new models as well as including more information.

Noting that sparse coding [11] that simultaneously learns a dictionary and a sparse decomposition over it, has a successful application in high dimensional data [12] and also for recommendation [9,13] in recent works. However, sparsity constraint in these

* Corresponding author.
E-mail address: wusen@manage.ustb.edu.cn (S. Wu).

https://doi.org/10.1016/j.neucom.2018.01.011

works is to regulate the neighbor selection and weight learning to construct a sparse graph for Top-N recommendations but not for user–item high dimensional feature learning in rating prediction. Also, in [14], the sparsity is exerted on the hyperparameters of Bayesian model. Thus, if we apply sparse coding as a new latent model for rating prediction problem in recommendation, it can be seen as a sparsity regularized latent model to extract factors from rating matrix, where the dictionary and coefficient respectively corresponds to the item and user feature in the latent factor space. As each user or each item is actually shown to be sparsely correlated with latent factors in recommendation problem (a user or item contributes to part of the factors rather than all features), sparse coding will be a potentially effective tool to model user–item-rating matrix for collaborative filtering.

In this paper, we are focused on improving the performance of collaborative filtering by proposing a Sparse LAtent Model (SLAM) based on symmetrized ℓ_1-constrained graph regularized sparse coding for collaborative filtering. SLAM learns two sparse matrices for users and items at the same time using symmetrized sparse coding algorithm, and it is also capable to simultaneously integrate user network and item network by dual graph Lapalacian regularization to the objective function. Experiments on real datasets demonstrate that SLAM can predict the user–item ratings better than the state-of-the-art matrix factorization based methods.

The rest of this paper is organized as follows. Section 2 presents review on related works and preliminaries. In Section 3, the proposed method is introduced in detail along with optimization method which is described in Section 4. Section 5 illustrates the experiments and results, and we summarize the work and provide future directions in Section 6.

2. Preliminaries and related works

In this section, we mainly review and discuss some representative matrix factorization based collaborative filtering models and sparse coding theory. Given m users and n items in a recommender system forming a $m \times n$ rating matrix R, where part of the ratings are observed, the goal of collaborative filtering is to predict the unknown entries in R, where rating R_{ij} corresponding to user i and item j.

Generally, two types of methods are widely used for conventional recommendation: neighborhood-based ones and model-based ones. Neighborhood-based methods include user-based [2] which focus on user–user similarity, and item-based [3,15] recommendations which use item–item similarity. Among these, item-based approaches have proved better than user-based approaches in terms of accuracy and efficiency for rating prediction. Model-based approaches can further be segregated into latent factor based matrix factorization methods and supervised sparsity inducing methods. The first category views collaborative filtering as a matrix completion task and addresses it by factorizing the rating/purchase matrix into low rank user and item factor matrices. These less number (low rank) latent features are essentially the communities among user in accordance to their preferences/tastes and categories of item types among items. Given a user and item, the dot product between corresponding feature vectors reveals how interesting this item might be for given user. Examples include probabilistic matrix factorization (PMF) [16], collective matrix factorization [17], regression-based latent factor model [18], and tensor factorization based methods [19]. This category of methods have proven to be excellent for rating prediction-oriented recommendations when compared to the item based neighborhood methods.

The model-based category which builds upon the sparse learning, is a more recent one which does matrix factorization along with sparsity inducing techniques. In this paper we improve upon this category by integrating sparse learning and graph regularization into existing models. Therefore, in the following subsection we provide an overview of both of them.

2.1. Matrix factorization for collaborative filtering

Since collaborative filtering is viewed as a matrix complete problem, low rank matrix approximation [4] aims to seek a low rank matrix that approximates the user rating matrix $R \in \mathbb{R}^{m\times n}$ with m users on n items by a multiplication of two l-rank factor matrices,

$$R \approx U^T V \text{ or } R_{ij} = < u_i, v_j >, \tag{1}$$

where $U = [u_1, u_2, \ldots, u_m] \in \mathbb{R}^{l\times m}$ and $V = [v_1, v_2, \ldots, v_n] \in \mathbb{R}^{l\times n}$ ($l \ll min(m, n)$) denote the user and item representations in the low dimensional feature space. Obviously R is sparse as a single user only rates a few items and also a single item is rated by a few users in the real applications. It is noted that the rank l is much less than the total number of the users m and the items n. To recover the large number of missing ratings in R, low rank matrix factorization is usually to solve the following unconstrained minimization problem:

$$\min_{U,V} \frac{1}{2}\sum_{i,j} I_{ij}\left(R_{ij} - u_i^T v_j\right)^2 + \frac{\lambda}{2}\left(\|U\|_F^2 + \|V\|_F^2\right), \tag{2}$$

where $\| \quad \|_F$ denotes the Frobenius norm of a matrix as $\|A\|_F^2 = \sum A_{ij}^2$. I_{ij} is the indicator of ratings, where $I_{ij} = 1$ if user i rated item j and $I_{ij} = 0$ otherwise. $\|U\|_F^2$ and $\|V\|_F^2$ are two regularization terms to avoid overfitting, with the positive regularization coefficient $\lambda > 0$ to balance the weight between the data reconstruction error term and regularization term, which is set empirically. Eq. (2) is usually solved by the gradient descent method or the alternative projection algorithm to find a local minimum. This may be one of the most popular collaborative filtering methods and can be parallelized in distributed system for large datasets. Based on the basic matric factorization model, some works are focused on how to improve the performance of collaborative filtering tasks, such as regularized single-element-based NMF [20], NMF with Alternating Direction Method [21], second-order optimization-based latent factor model [22] and symmetric non-negative latent factor model [23].

In the traditional recommendation system, users are assumed to be independent and identically distributed, and thus the social relationship or connections between users are ignored. However, in the real world, the social relationship can determine what we select to some extent, as we always turn to friends we trust for movie, music, or restaurant recommendations, meaning our favors can easily be affected by our social activities. Based on this intuition, social network-based collaborative filtering [10] is attracting much attention, which is to integrate the social connection between users, such as trust network. An intuitive way is to replace the step of finding similar users with the use of a trust metric. When it comes to matrix factorization based collaborative filtering, the interaction between users can be considered as a graph regularized latent model problem as Social Regularized matrix factorization (SR) in [8]:

$$\min_{U,V} \frac{1}{2}\sum_{i,j} I_{ij}\left(R_{ij} - u_i^T v_j\right)^2 + \frac{\alpha}{2}\sum_{i=1}^{m}\sum_{f\in F(i)} w_{if}\|u_i - u_f\|^2 + \frac{\lambda}{2}\left(\|U\|_F^2 + \|V\|_F^2\right) \tag{3}$$

where w_{if} represents similarity between user i and f, which can be computed using cosine similarity or Pearson correlation coefficient, and $F(i)$ denotes the set of friends of user i. The second term