Engineering Applications of Artificial Intelligence 46 (2015) 58–66

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

An imputation-based matrix factorization method for improving accuracy of collaborative filtering systems

Manizheh Ranjbar [a], Parham Moradi [a,*], Mostafa Azami [a], Mahdi Jalili [b]

[a] Department of Computer Engineering, University of Kurdistan, Sanandaj, Iran
[b] School of Electrical and Computer Engineering, RMIT University, Melbourne, Australia

ARTICLE INFO

Article history:
Received 12 April 2015
Received in revised form
19 August 2015
Accepted 24 August 2015
Available online 25 September 2015

Keywords:
Matrix factorization
Collaborative filtering
Recommender system
Multiplicative update rules, Impute rates

ABSTRACT

Matrix-Factorization (MF) is an accurate and scalable approach for collaborative filtering (CF)-based recommender systems. The performance of matrix MF methods depends on how the system is modeled to mitigate the data sparsity and over-fitting problems. In this paper we aim at improving the performance of MF-based methods through employing imputed ratings of unknown entries. A novel algorithm is proposed based on the classic Multiplicative update rules (MULT), which utilizes imputed ratings to overcome the sparsity problem. Experimental results on three real-world datasets including *MovieLens*, *Jester*, and *EachMovie* reveal the effectiveness of the proposed strategy over state of the art methods. The proposed method is more tolerant against the sparsity of the datasets as compared to other methods including Alternating Least Squares (ALS), Stochastic Gradient Descent (SGD), Regularized Stochastic Gradient Descent (RSGD), Singular Value Decomposition Plus Plus (SVD++) and MULT methods.

1. Introduction

Recently, there has been much attention in the community of computer science to Recommender Systems (RS), which have evolved in the extremely interactive environment of the Web. They apply data analysis techniques to help e-commerce customers to better find the products they would like to purchase (Konstan et al., 1997; Resnick et al., 1994a; Resnick and Varian, 1997; Shardanand and Maes, 1995). The main sources of data used to design RSs are user profiles, item profiles and user-item interactions (i.e., user ratings to the available items). RS approaches can be potentially classified into three categories: content-based approaches (Adomavicius and Tuzhilin, 2005a; Blanco-Fernández et al., 2011; Lang, 1995; Mooney and Roy, 2000), collaborative filtering approaches (Goldberg et al., 1992; Miller et al., 2003; Resnick et al., 1994b), and hybrid approaches (Burke, 2002; Pazzani, 1999). In content-based approaches, the system learns to recommend items based on available contextual information on the users and items (Formoso et al., 2013; Pazzani and Billsus, 2007; Salter and Antonopoulos, 2006). Collaborative filtering (CF) constructs the recommendation list based on the preferences of a group of users that are similar to the active user. This approach is basically classified into memory-based and model-based approaches (Ahmadian et al., 2014; Breese et al., 1998; Candillier et al., 2008; Ramezani et al., 2014; Russell and Yoon, 2008). Memory-based approaches employ user-item matrix to find similar users and generate a prediction (Moradi and Ahmadian, 2015; Moradi et al., 2015; Sarwar et al., 2001). On the other hand, model-based approaches first build a user model in an offline learning phase, and then apply this model online to generate the recommendations (Al-Shamri and Bharadwaj, 2007; Bell et al., 2007; Bojnordi and Moradi, 2012; Javari and Jalili, 2014, 2015). Finally, hybrid RSs combine the contextual information with CF methods in order to provide more accurate recommendations.

Previous research showed that CF approaches are affective in designing industrial RSs (Adomavicius and Tuzhilin, 2005b; Bobadilla et al., 2012, 2010; Breese et al., 1998; du Boucher-Ryan and Bridge, 2006; Herlocker et al., 2004; Javari et al., 2014; Javari and Jalili, 2014; Li et al., 2007; Pu and Chen, 2007; Ramezani et al., 2014; Victor et al., 2008). In CF-based techniques, users' activity is recorded in a user-item rating matrix, which includes their interest in items. The objective of CF is to recommend the most relevant items to a target user based on the available rates in the user-item rating matrix. In many practical RSs, the users rate only a limited number of items; thus, the rating matrix is heavily sparse and one has to estimate the missing rates in order to find proper recommendation list for users. Therefore, the task is to estimate the unknown ratings based on known entries with minimum accumulative error (Bobadilla et al., 2013; Luo et al., 2012b). Matrix Factorization-based RSs are one of the most successful methods resulting satisfactory performance in various applications

* Correspondence to: University of Kurdistan, Pasdaran street, Sanandaj 66177-15177, Iran. Tel.: +98 8733668513.
E-mail addresses: manij.ranjbar@gmail.com (M. Ranjbar), p.moradi@uok.ac.ir (P. Moradi), m.azami@b-iust.ac.ir (M. Azami), mahdi.jalili@rmit.edu.au (M. Jalili).

http://dx.doi.org/10.1016/j.engappai.2015.08.010
0952-1976/

(Bojnordi and Moradi, 2012; Luo et al., 2012a, 2012c, 2013; Navgaran et al., 2013; Takács et al., 2009; Zhang et al., 2012). They transform both items and users into the same latent factor space. Moreover, each entry is characterized by a feature vector which is inferred from the existing ratings and the unknown rates are predicted using the inner products of the corresponding vector pairs.

There are a number of techniques to solve the Matrix factorization (MF) problem such as Alternating Least Squares (ALS) (Paatero and Tapper, 1994), Multiplicative update rules (MULT) (Lee and Seung, 2001), Stochastic Gradient Descent (SGD) (Lin, 2007), SVD++ (Koren, 2008) and Regularization Stochastic Gradient Descent (RSGD) (Takács et al., 2009). Compared to others, MULT converges faster (Lee and Seung, 2001), and here we further extend it to deal with the sparsity problem. The proposed method is different from state of the art methods in that it includes a pre-processing step to impute a subset of unknown ratings. The proposed method, denoted by Imputed MULT (IMULT), helps to guide the iterative solutions to the true ratings. Our method outperforms MULT, ALS, SGD, RSGD and SVD++ in all numerical experiments.

The rest of this paper is organized as follows. In Section 2, MF concept is first explained and then MULT approach is described. Section 3 describes the proposed method followed by proof of convergence in Section 4. Then experimental results are presented in Section 5 and finally Section 6 concludes the paper.

2. Background

This section gives a brief overview of MF approach for CF and describes MULT method, which has been widely used to implement MF approach.

2.1. Matrix factorization

MF-based approaches have proven to be effective techniques in designing efficient CF recommenders (Bojnordi and Moradi, 2012; Koren and Bell, 2011; Luo et al., 2012c; Navgaran et al., 2013). This technique was introduced by Paatero and Tapper employing ALS algorithm (Paatero and Tapper, 1994). Since then, MF has been widely used in machine learning (Lee and Seung, 1999, 2001), pattern recognition, multimedia, and text mining (Chen et al., 2009). In this section, we first briefly review the basic idea of MF, and then describe MULT approach. Consider a matrix $R \in \mathbb{R}^{n\times m}$ and a positive integer $k \ll \min(n,m)$. The main objective is to find matrix factors $W \in \mathbb{R}^{n\times k}$ and $H \in \mathbb{R}^{k\times m}$ in which $R \approx WH$.

In CF, W denotes the users matrix with k latent features such that each row represents information of a specific user. Moreover, H shows the items matrix including k latent features in which each column shows a specified item. Thus, the main task in MF is to estimate the values of W and H. AS these matrices are completed, the predicted rating of user u to item i ($\tilde{r}_{u,i}$) can be estimated by the inner product of the corresponding user-item feature vector pairs, given by $\tilde{r}_{u,i} = w_u h_i$ where w_u and h_i denote the u-th row and i-th column of W and H, respectively. Moreover, the accumulative error between the approximated matrix $\tilde{R}$ and the actual rating matrix R can be obtained as follows:

$$\text{Error} = \|R - \tilde{R}\| = \sum_{u,i} (r_{u,i} - W_u H_i)^2$$

where $\|\,.\,\|$ denotes the standard Euclidean norm. Therefore, the cost function can be defined as follows:

$$F(W,H) = \frac{1}{2}\|R - WH\|^2, \; s.t. \;\; W, H \geq 0$$

Generally, the main goal is to find the unknown parameters of W and H minimizing the cost function. In CF systems, MF model may deal with missing value matrices. Thus, the cost function can be reformulated as follows:

$$F(W,H) = \frac{1}{2}\|P_{\Omega}(R - WH)\|^2, \; s.t. \; W, H \geq 0 \tag{1}$$

where Ω is the set of known entries of R, and $P_{\Omega}(\,.\,): \mathbb{R}^{n\times m} \to \mathbb{R}^{n\times m}$ is defined as:

$$P_{\Omega}(x_{i,j}) = \begin{cases} x_{i,j} & (i,j) \in \Omega \\ 0 & \text{otherwise} \end{cases}$$

Eq. (1) can be rewritten as:

$$F(W,H) = \frac{1}{2}\sum_{(u,i)\in\Omega}\left(r_{u,i} - \sum_{l=1}^{k} w_{u,l} h_{l,i}\right)^2, \; s.t. \; W, H \geq 0 \tag{2}$$

where $w_{u,l}$ corresponds to entry (u,l) of W and likewise $h_{l,i}$, k indicates the number of latent features of W and H.

According to the above objective function, estimating the parameters of W and H can be achieved by searching around the global (or local) minimum of the quadratic cost function F as shown in Eq. (2).

2.2. Multiplicative update rules method

Multiplicative update rules method (MULT), is a fast and effective solver for computing the solution of MF problem (Lee and Seung, 2001). This method is a gradient descent-based approach with a special choice of learning step-sizes. Thus, it can be rewritten as a special case of the projected gradient method. The iterative MULT rules for updating the elements of W and H can be described as (Lee and Seung, 2001):

$$w_{u,l} = w_{u,l} - \beta_{u,l}\left(\frac{\partial F(W,H)}{\partial w_{u,l}}\right) \tag{3}$$

$$h_{u,l} = h_{u,l} - \alpha_{u,l}\left(\frac{\partial F(W,H)}{\partial h_{u,l}}\right) \tag{4}$$

where $\alpha_{l,i}$ and $\beta_{u,l}$ are learning step-sizes, which are defined as $\alpha_{l,i} = h_{l,i}/\left(W^T W H\right)_{l,i}$ and $\beta_{u,l} = w_{u,l}/\left(WHH^T\right)_{u,l}$. The multiplicative update rules can be rewritten as

$$w_{u,l} = w_{u,l}\frac{\left(RH^T\right)_{u,l}}{\left(WHH^T\right)_{u,l}} \tag{5}$$

$$h_{l,i} = h_{l,i}\frac{\left(W^T R\right)_{l,i}}{\left(W^T W H\right)_{l,i}} \tag{6}$$

It has been shown trough simulation that if there is not a sufficient number of known entries, the MULT update rules as shown by Eq. (3) does not converge to the optimum value. Furthermore, RSs often deal with sparse data, which further influence effectiveness of MULT. In this paper, a novel method is proposed to overcome this problem using further pre-processing in the estimated process.

3. Proposed method

The proposed method is based on a pre-processing step in the estimation process, which helps improve the results of the final estimation phase. This can guide the solutions to the optimum and mitigate the over-fitting problem. Moreover, unknown entries can be estimated using Item-wise, User-wise, Mean-wise and Hybrid-wise