Expert Systems with Applications 42 (2015) 4733–4742

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Transforming collaborative filtering into supervised learning

Filipe Braida [a,b,*,1], Carlos E. Mello [b,1], Marden B. Pasinato [a,2], Geraldo Zimbrão [a,2]

[a] PESC/COPPE, Universidade Federal do Rio de Janeiro, Cidade Universitária, Rio de Janeiro, P.O. Box: 68511, Brazil
[b] DCC/IM, Universidade Federal Rural do Rio de Janeiro, Nova Iguaçu, Rio de Janeiro 26020-740, Brazil

ARTICLE INFO

Article history:
Available online 17 January 2015

Keywords:
Recommender system
Collaborative filtering
Dimensionality reduction
Supervised learning

ABSTRACT

Collaborative Filtering (CF) is a well-known approach for Recommender Systems (RS). This approach extrapolates rating predictions from ratings given by user on items, which are represented by a user-item matrix filled with a rating $r_{i,j}$ given by an user i on an item j. Therefore, CF has been confined to this data structure relying mostly on adaptations of supervised learning methods to deal with rating predictions and matrix decomposition schemes to complete unfilled positions of the rating matrix. Although there have been proposals to apply Machine Learning (ML) to RS, these works had to transform the rating matrix into the typical Supervised Learning (SL) data set, *i.e.*, a set of pairwise tuples (x, y), where y is the correspondent class (the rating) of the instance $x \in \mathbb{R}^k$. So far, the proposed transformations were thoroughly crafted using the domain information. However, in many applications this kind of information can be incomplete, uncertain or stated in ways that are not machine-readable. Even when it is available, its usage can be very complex requiring specialists to craft the transformation. In this context, this work proposes a domain-independent transformation from the rating matrix representation to a supervised learning dataset that enables SL methods to be fully explored in RS. In addition, our transformation is said to be straightforward, in the sense that, it is an automatic process that any lay person can perform requiring no domain specialist. Our experiments have proven that our transformation, combined with SL methods, have greatly outperformed classical CF methods.

1. Introduction

The growth of Internet has brought a myriad of new business and applications by reshaping many human activities. For instance, there has been a significant shift in trading towards more flexibility, availability, and mobility. A million of e-commerce systems around the globe have been deployed providing all these features and offering their users an unlimited catalog of goods and services. Due this overabundance of possibilities, users may become anxious and abandon the purchase process, once making choices with too many options may be bothersome and difficult (Schwartz, 2005).

In this scenario, Recommender Systems (RS) have emerged playing the important role of providing users with recommendations. These systems have been widely adopted by big players of the Web such as Amazon[3], NetFlix[4], YouTube[5] and many others. Besides their obvious virtues of leveraging sales, RS also improve customer loyalty and increase cross-sales. In fact, these systems try to efficiently meet needs and interests of users by avoiding the burden of finding a needle in a haystack (Schafer, Konstan, & Riedi, 1999).

The most prominent and successful RS approach is Collaborative Filtering (CF). It aims to recommend an item to a user based only on item evaluations (often represented by a numerical rating) provided by other users in the system. This scheme may be modeled through a $n_{users} \times m_{items}$ matrix R, where each position $r[i,j]$ of R is filled with the corresponding rating value $r_{i,j}$ given by an user i to an item j. Often, the rating matrix is sparse, in the sense that, very few ratings are known. The positions with unknown ratings are conventionally filled with a nonexistent rating symbol $\varnothing$.

Accordingly, CF algorithms predict ratings for the unfilled positions (unknown ratings) by extrapolating from the filled ones in the rating matrix. Thus, each unfilled matrix position can be viewed as an instance that needs to be classified into one of the

* Corresponding author at: PESC/COPPE, Universidade Federal do Rio de Janeiro, Cidade Universitária, Rio de Janeiro, P.O. Box: 68511, Brazil. Tel.: +55 21 2562 8672.

E-mail addresses: filipebraida@ufrj.br (F. Braida), carlos.mello@ufrrj.br (C.E. Mello), marden@ufrj.br (M.B. Pasinato), zimbrao@cos.ufrj.br (G. Zimbrão).

[1] Tel.: +55 21 2669 0105.
[2] Tel.: +55 21 3938 8672.

[3] http://www.amazon.com.
[4] http://www.netflix.com.
[5] http://www.youtube.com.

http://dx.doi.org/10.1016/j.eswa.2015.01.023
0957-4174/© 2015 Elsevier Ltd. All rights reserved.

possible ratings. Note that, no other input data than the rating matrix is required in this approach.

The CF data scheme creates a barrier for applying Supervised Learning (SL) as it is not possible to represent ratings as points in a vector space $\mathbb{R}^k$, like in classification and regression tasks. Indeed, there are no straightforward features to build an input space from which SL methods could learn rating prediction models.

There have been proposals that attempt to transform the rating matrix into a typical SL dataset suitable for ML methods. Nevertheless, the proposed transformations rely on the domain information, which can be hard to extract and even misleading (Cunningham & Smyth, 2010; Hsu, Wen, Lin, Lee, & Lee, 2007; O'Mahony & Smyth, 2009, 2010).

Moreover, a transformation process based on available domain information is still a complex task, once only specialists are able to execute. The transformations proposed so far are thoroughly crafted using the domain information and are also very specific. Thus one may claim that no straightforward domain-independent methodology has been proposed to transform the CF task into a SL task.

In this context, this work proposes a novel straightforward domain-independent methodology to transform the CF scheme into a SL scheme, in the sense that, this is an automatic process which only requires the rating matrix as input. The positions of the rating matrix R are mapped into a k-dimensional vector space corresponding to the k most important latent factors of R. Each point in this feature space is associated with its corresponding rating value forming a typical SL dataset. SL methods are trained with this dataset and applied to predict the users' unknown ratings. Their performance was given according to metrics such as MAE and RMSE, which indicate that this approach has greatly outperformed classical CF techniques.

This paper is organized in 5 sections of which this is the first one. Section 2 presents related work on how to apply supervised learning techniques in recommender systems. In Section 3, the general proposal is described in details and some theoretical insights are provided. Along Section 4, the experimental settings and results are presented. Finally, a discussion about the weakness and strengths of our proposal is carried out in Section 5 as well as some future works.

2. Related work

Collaborative Filtering (CF) approach has been largely used in Recommender Systems (RS). This approach aims to extrapolate rating predictions from a user-item matrix filled with the corresponding ratings given by users to items. The main related issue of this approach relies on the sparsity of this rating matrix. To handle this, many CF algorithms have been proposed based on adaptations of classic classification and regression methods (Ricci, Rokach, Shapira, & Kantor, 2011). For instance, the User-based CF algorithm is a k-Nearest Neighbors classifier with similarity measures between users adapted to consider only the co-rated items in the computation (Adomavicius & Tuzhilin, 2005).

Alternatively, there have been many works that apply dimensionality reduction techniques so as to transform the sparse matrix into a fixed space of features. Sarwar, Karypis, Konstan, and Riedl (2000) proposes the use of the Singular Value Decomposition (SVD) and the recommendations are generated by operations between the resulting matrices.

CF methods based on Matrix Factorization (MF) techniques have received great attention after NetFlix Prize. In this challenge, one of the top accurate methods was based on the factorization of the rating matrix through SVD (Koren, Bell, & Volinsky, 2009; Salakhutdinov & Mnih, 2008b). In addition, this sort of technique has shown to be scalable and accurate even under high sparsity. The idea behind these techniques is to complete the rating matrix through a low-rank approximation (Koren, 2008; Koren et al., 2009; Paterek, 2007). There have been also proposals that exploits, instead of matrix factorization, a probabilistic latent variable framework, wherein hidden random variables constitutes the underlying users' preferences on items (Salakhutdinov & Mnih, 2008a, 2008b).

Another approach to try to deal with the sparsity problem is to treat it as classification/regression problem. An approach of this kind is to derive features from the rating matrix based on the domain knowledge (Cunningham & Smyth, 2010; Hsu et al., 2007; O'Mahony & Smyth, 2009, 2010). For instance, the users' mean of ratings and the standard deviation may constitute features to train classifiers.

In Billsus (1998), authors propose a domain-independent method in order to predict ratings. For each user a model is induced based on a matrix of features derived from the original rating matrix. Although this proposal performs well, it does not work for multiclass problems and is not scalable, since the number of models to learn increases with new users. This method uses SVD to reduce dimensionality and build a new feature space in order to train SL models.

To the best of our knowledge there is no work that proposes a general straightforward domain-independent transformation of the rating matrix into a classic training set for applying SL methods.

3. Proposal

In this section we start discussing the main issue in Collaborative Filtering (CF) that prevents the direct application of Supervised Learning (SL) methods to rating predictions. We follow by describing the proposed methodology of this work.

In CF, the raw dataset consists of ratings $r_{i,j}$ associated with their corresponding user-item pairs, *i.e.* a numerical evaluation $r_{i,j} \in \mathbb{R}$ given by a user i on an item j. Hence, CF only uses information about how users like or dislike certain items. This is often represented by a rating matrix R, where each line i corresponds to a user and each column j corresponds to an item (Adomavicius & Tuzhilin, 2005).

CF aims to learn users' preferences from the matrix R in order to extrapolate rating predictions for items not yet rated by the users, *i.e.* to complete the unfilled positions in R. Usually, the rating matrix is subject to high sparsity, since only few users usually provide ratings by leaving many matrix positions unfilled. This makes the CF task even harder, once the more sparse the rating matrix, the harder the learning and more unfilled positions should be completed with predicted ratings.

The main issue that differentiates the CF problem from the classic SL problem lies in the fact that there is only the rating data from which one may learn a model for the users' preferences. A rating dataset is fundamentally unlike classic SL datasets, where instances are defined in an input feature space and are associated with outputs, in case, ratings. In fact, there is no such well-defined input feature space in a CF scheme so that one may learn a mapping from instances in a domain of features to output values (ratings). Instead, there are user-item pairs that determine the rating values, which are not represented by features.

For applying SL methods to handle CF rating predictions there should be a feature vector associated with each rating in the CF dataset. Each user-item pair in CF data must be transformed into an instance vector of features and labeled with its corresponding rating. In this form, one could deal with this set of labeled instances as a classic supervised training set from which SL methods may be applied.

Although many CF algorithms have been proposed in the literature of recommender systems (Ricci et al., 2011), to the best of our knowledge, there is no clear methodology or method that