Neurocomputing 204 (2016) 17–25

Contents lists available at ScienceDirect

Neurocomputing

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

Pairwise probabilistic matrix factorization for implicit feedback collaborative filtering

Gai Li [a,*], Weihua Ou [b]

[a] Department of Electronics and Information Engineering, Shunde Polytechnic, Foshan 528300, China
[b] School of Mathematics and Computer Science, Guizhou Normal University, Guiyang 550001, China

ARTICLE INFO

Article history:
Received 15 March 2015
Received in revised form
11 July 2015
Accepted 20 August 2015
Available online 8 April 2016

Keywords:
Recommended systems
Collaborative filtering
Collaborative ranking
Implicit feedback
Probabilistic matrix factorization

ABSTRACT

Implicit feedback collaborative filtering has attracted a lot of attention in collaborative filtering, which is called one-class collaborative filtering (OCCF). However, the low recommendation accuracy and the high cost of previous methods impede its generalization in real scenarios. In this paper, we develop a new model named pairwise probabilistic matrix factorization (PPMF) by using the advantages of RankRLS. PPMF model takes RankRLS integrated with PMF (probabilistic matrix factorization) to learn the relative preference for items. Different from previous works, PPMF minimizes the average number of inversions in ranking rather than maximize the gaps of the binary predicted values for OCCF problem. Meanwhile, we propose to optimize the PPMF model by the pointwise stochastic gradient descent algorithm based on bootstrap sampling, which is more effective for parameter learning than the original optimization method used in previous works. Experiments on two datasets show that PPMF model achieves satisfactory performance and outperforms the state-of-the-art implicit feedback collaborative ranking models by using different evaluation metrics.

1. Introduction

Recommender systems are a subclass of information filtering systems that seek to predict the 'rating' or 'preference' that user would give to an item [1]. The preference relation is learned from the user's past interactions with the system, such as purchase history or click-through log. The aim of the system is to create a user-specific ranking for a set of items by learning the user–item relationships.

The core technology used in recommender systems is collaborative filtering (CF). Traditional CF can be put into use by collecting users' explicit feedback. Explicit feedback is more widely used in the research fields of recommender system and often in the form of numeric ratings input by users to express their preferences regarding items. However, such explicit feedback is hard to collect in practice because of the intensive user involvement. As a matter of fact, most of the feedback is not explicit but implicit. The research on implicit feedback about CF is also called one-class collaborative filtering (OCCF) [2], in which only positive implicit feedback or only positive examples can be observed, and it is a new emerging research field in CF. The explicit and implicit feedback data can be expressed in matrix form as shown in Fig. 1. In the explicit feedback matrix, an element can be any real number, but often ratings are integers in the range (1–5), such as the ratings on Netflix, where a missing element represents a missing example. In the implicit feedback matrix, the positive-only users' preferences data can be represented as a single-valued matrix, where a 1-valued element represents a positive example and a missing element represents a missing example. Compared with the traditional CF setting where the data have ratings, OCCF is more realistic in many scenarios when no ratings are available, such as scientific paper suggestions, product recommendations, and personalized news recommendations. Therefore, in this paper we focus on item recommendation from implicit, positive-only feedback [3].

In the early research, OCCF was treated as a one-class classification problem and was solved by an algorithm such as one-class SVMs [4]. Following one-class SVMs, some matrix factorization models have been proposed to tackle the OCCF problem. Pan et al. [2] proposed to weight and sample unobserved user-item elements and to learn a matrix factorization model by minimizing the weighted element-wise squared loss. The essential idea is to treat all missing user-item examples as negative and to assign proper weights to these entries. Hu et al. [5] proposed to treat the data as indicating positive and negative preferences associated with vastly varying confidence levels. Sindhwani et al. [6] proposed a non-negative matrix factorization (NMF) model as the optimized model in group matrix factorization for OCCF problems. Gabor et al. [7]

* Corresponding author. Tel.: +86 13420672172.
E-mail address: ligai999@126.com (G. Li).

http://dx.doi.org/10.1016/j.neucom.2015.08.129
0925-2312/

investigated some applications of the Conjugate Gradient method for new and existing implicit feedback CF models. Recent studies [2,5–7] can improve the performance of OCCF by solving the sparsity problem and imbalance of rating data to a certain extent. The disadvantage of those methods is that it obtains the ranked item list through intermediate rating prediction and creates the list indirectly, which influences the performance improvement of those methods.

Solving OCCF problem by the idea of ranking is a new emerging research field in CF, skipping the intermediate rating prediction step, and creating the ranked list directly. The foremost of these is Bayesian personalized ranking (BPR) [3], which converts the OCCF problem into a ranking one. Recently, Pan et al. [8] have proposed group Bayesian personalized ranking (GBPR), via introducing richer interactions among users. In GBPR, it introduces group preference, to relax the individual and independence assumptions of BPR. Shi et al. [9] proposed collaborative less-is-more filtering (CLiMF), in which the model parameters are learned by directly maximizing the well-known information retrieval metric: Mean Reciprocal Rank (MRR), but CLiMF is not suitable for other evaluation metrics (e.g., MAP, AUC). Solving OCCF problem by the idea of ranking can further improve the performance of OCCF [3,8,9]. But all of them focus on maximizing the gaps of the binary predicted values, which also influences the further performance improvement of those methods.

The above researches about OCCF all focus on maximizing the gaps of the binary predicted values so as to improve the performance of those methods. From the references [10,11], we know that minimizing the average number of inversions in ranking is better than maximizing the gaps of the binary predicted values for the binary ranking problem (OCCF). The objective function of RankRLS [10] (a pairwise learning-to-rank (LTR) approach) is to minimize the average number of inversions in ranking. In this paper, in order to further improve the recommendation accuracy and reduce the time complexity of OCCF method, we develop a new model named pairwise probabilistic matrix factorization (PPMF) by using the advantages of RankRLS. PPMF model takes RankRLS integrated with PMF (probabilistic matrix factorization) to learn the relative preference for items. Therefore, the objective function of our model is also to minimize the average number of inversions in ranking for OCCF problem. Experiments on two datasets show that our proposed PPMF model achieves satisfactory performance and outperforms the state-of-the-art implicit feedback collaborative ranking models (which are all to maximize the gaps of the binary predicted values for OCCF problem) by using different evaluation metrics.

The main contributions of this paper can be summarized as follows:

(1) We develop a novel PPMF model, which takes RankRLS integrated with PMF. PPMF model can further improve the recommendation accuracy of OCCF by minimizing the average number of inversions in ranking.
(2) We propose to optimize the PPMF model by the pointwise stochastic gradient descent algorithm based on bootstrap sampling, which can reduce the time cost of parameter learning compared with the original optimization method used in previous methods.
(3) Experiments empirically show that our PPMF model achieves satisfactory performance and outperforms other state-of-the-art implicit feedback collaborative ranking methods.

The rest of paper is organized as follows. We provide overview of the related works in Section 2 and introduce some preliminaries in Section 3. We present our proposed model and discuss how parameters are learned and how inference is carried out in Section 4. The experimental results and discussion are presented in Section 5, followed by the conclusions and future work in Section 6.

2. Related work

2.1. Rating prediction

In conventional CF tasks, the most frequently used evaluation metrics are the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE). Therefore, rating prediction (such as the Netflix Prize) has been the most popular method of solving the CF problem. Rating prediction methods are always regression based: they minimize the error of predicted ratings and true ratings. The simplest algorithm for rating prediction is k-nearest-neighbor (KNN) [12], which predicts the missing ratings from the neighborhood of users or items. KNN is a memory-based algorithm, and one needs to compute all the similarities between different users or items. More efficient algorithms are model based: they build a model from the visible ratings and compute all the missing ratings from the model. Widely used model-based rating prediction methods include PLSA [13], the restricted Boltzmann machine (RBM) [14], and a series of matrix factorization techniques [15,16].

	Item					
User	1			3		
		3			2	
		4			4	
	4		3			1
		5			5	
			2			5

	Item					
User		1			1	
			1			
	1			1		
			1			1
		1			1	
			1			1

Fig. 1. Examples of an explicit feedback matrix (left) and an implicit feedback matrix (right) for a recommendation system.