Information Sciences 332 (2016) 84–93

Contents lists available at ScienceDirect

Information Sciences

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

Mixed factorization for collaborative recommendation with heterogeneous explicit feedbacks

Weike Pan, Shanchuan Xia, Zhuode Liu, Xiaogang Peng, Zhong Ming*

College of Computer Science and Software Engineering, Shenzhen University, China

ARTICLE INFO

Article history:
Received 28 November 2014
Revised 23 September 2015
Accepted 27 October 2015
Available online 10 November 2015

Keywords:
Preference learning
Collaborative recommendation
Heterogeneous explicit feedbacks
Mixed factorization
Transfer learning

ABSTRACT

Collaborative recommendation (CR) is a fundamental enabling technology for providing high-quality personalization services in various online and offline applications. Collaborative recommendation with heterogeneous explicit feedbacks (CR-HEF) such as 5-star grade scores and like/dislike binary ratings is a new and important problem, because it provides a rich and accurate source for learning users' preferences. However, most previous works on collaborative recommendation only focus on exploiting *homogeneous* explicit feedbacks such as grade scores or *homogeneous* implicit feedbacks such as clicks or purchases. In this paper, we study the CR-HEF problem, and design a novel and generic mixed factorization based transfer learning framework to fully exploit those two different types of explicit feedbacks. Experimental results on two CR-HEF tasks with real-world data sets show that our TMF is able to perform significantly better than the state-of-the-art methods.

1. Introduction

Collaborative recommendation [1,2] serves as an enabling technology for providing personalization services in various systems and applications, such as academic resource recommendation [3,4], entertainment video recommendation [5,6], telecom/mobile service recommendation [7,8], and people/community recommendation [9,10], etc. The main idea of collaborative recommendation is to learn users' hidden preferences via exploiting users' feedbacks in a collective way instead of studying each user separately [11,12]. However, we may not learn users' preferences well when users' feedbacks are few [13,14], in particular of the 5-star grade scores that most recommendation algorithms [15,16] rely on. This kind of data scarcity challenge will usually cause the overfitting problem when building a recommendation model.

Heterogeneous explicit feedbacks (HEF) such as grade scores from best to worst and binary ratings of like and dislike provide a rich and accurate source for learning users' preferences and constructing users' online profiles, which gives us an opportunity to address the data scarcity challenge of the grade scores [17]. However, most mathematical models in collaborative recommendation (CR) are designed for learning users' preferences from *homogeneous* explicit feedbacks such as 5-star grade scores [15,16], or from *homogeneous* implicit feedbacks such as clicks or purchases [18,19]. In this paper, we study this new and important problem, i.e., collaborative recommendation with heterogeneous explicit feedbacks (CR-HEF), which contains two different types of explicit feedbacks, i.e., 5-star grade scores and like/dislike binary ratings. We then focus on designing a novel and generic algorithm to fully exploit such heterogeneous explicit feedbacks in a principled way.

* Corresponding author. Tel.: +8675526534480.
E-mail addresses: panweike@szu.edu.cn (W. Pan), shaun.xia@outlook.com (S. Xia), zero.lzd@gmail.com (Z. Liu), patrickpeng@126.com (X. Peng), mingz@szu.edu.cn (Z. Ming).

http://dx.doi.org/10.1016/j.ins.2015.10.044
0020-0255/

Previous works on exploiting heterogeneous explicit feedbacks are very few, among which the major approach is matrix *collective* factorization [17,20,21]. Collective factorization based methods are usually designed to simultaneously factorize two preference matrices, i.e., a grade score matrix and a binary rating matrix in CR-HEF, where some latent features from the same users or the same items are shared so as to enable joint preference modeling and sharing. However, such two jointly conducted factorization tasks are still loosely coupled, which may not fully leverage the binary ratings to the grade scores.

Besides the collective factorization methods, there is a different technique called matrix *integrative* factorization for a related problem to CR-HEF, i.e., collaborative recommendation with grade scores and implicit feedbacks [15]. Integrative factorization based methods take the implicit feedbacks as instances (or feedback instances), and incorporate them seamlessly into the factorization task of grade scores as an additional term in the prediction rule. However, leveraging the implicit feedbacks to the grade scores in such an integrative manner may not well capture the implicit-feedback-dependent effect.

In this paper, we aim to overcome the aforementioned limitations of the two state-of-the-art factorization models, i.e., collective factorization and integrative factorization, and adapt them to our studied CR-HEF problem. Specifically, we first take the CR-HEF problem from a transfer learning view [22], in which grade scores are taken as target data and binary ratings are taken as auxiliary data. We then propose a novel and generic mixed factorization based transfer learning framework, i.e., *transfer by mixed factorization* (TMF), which consists of collective factorization and integrative factorization as different assembly components. The novelty of our TMF is that it unifies collective factorization and integrative factorization in one single transfer learning framework, which enables both feature-based and instance-based preference learning and transfer in a principled way. Furthermore, we can also derive some new algorithm variants from our TMF for leveraging different parts or combinations of auxiliary binary ratings.

TMF is expected to transfer more knowledge from binary ratings to grade scores than collective factorization, and to model binary-rating-dependent and -independent effect more accurately than integrative factorization. Experimental results on two real-world data sets show that our TMF performs significantly better than collective factorization or integrative factorization alone.

We organize the rest of the paper as follows. We first discuss some closely related works in Section 2. We then describe the proposed solution in detail in Section 3, and conduct extensive empirical studies of our TMF and the state-of-the-art methods in Section 4. Finally, we give some concluding remarks and future directions in Section 5.

2. Related work

Our TMF is designed from a transfer learning perspective with factorization techniques, which aims to learn users' preferences from heterogeneous explicit feedbacks (HEF) in collaborative recommendation. Hence, in this section, we discuss some related work of TMF in the context of transfer learning for heterogeneous feedbacks, and factorization for collaborative recommendation.

2.1. Transfer learning for heterogeneous feedbacks

Transfer learning [22] aims to conduct parameter learning and knowledge transfer for more than one domains, tasks or data, which has been a state-of-the-art solution in many applications, including text and image classification [23,24], personalization and social behavior learning [17,25], etc. Transfer learning algorithms have also been designed to learn users' preferences from heterogeneous feedbacks in collaborative recommendation, for which we will discuss about their problem settings and techniques.

The problem settings of heterogeneous feedbacks mainly include two categories, i.e., (i) different recommendation tasks with the same feedback type, and (ii) the same recommendation task with different feedback types. The first category includes recommendation tasks for books and movies with grade scores [13,26], recommendation tasks for books, movies and music with grade scores [27], and recommendation tasks for different Amazon products with grade scores [28,29], etc. This category is usually called cross-domain collaborative recommendation [30,31] because the items or product domains are different. The second category includes recommendation for movies with different types of feedbacks such as grade scores and implicit, binary and/or uncertain feedbacks [14,17,32]. Our studied CR-HEF problem in this paper also belongs to the second category. Transfer learning techniques for heterogeneous feedbacks in collaborative recommendation or other applications mainly contain model-based, feature-based and instance-based transfer [22,32], and the corresponding algorithms include adaptive, collective and integrative styles. We follow and expand the categorizations of transfer learning techniques in collaborative recommendation in [32], in particular of "how to transfer" in transfer learning [22], and show the relationship between our TMF and some typical works in Table 1. From Table 1, we can see that our TMF is different from all existing works w.r.t. either of the two dimensions, i.e., transfer learning approaches or transfer learning algorithm styles. Specifically, TMF is a *mixed* transfer learning approach of feature-based transfer and instance-based transfer, and a *mixed* transfer learning algorithm of collective factorization and integrative factorization.

2.2. Factorization for collaborative recommendation

Factorization techniques such as second order matrix factorization, higher order tensor factorization and their extensions have been well studied and successfully applied to many machine learning and data mining problems [33], among which collaborative recommendation is an important application [15,16,19,34–37].