Neurocomputing 177 (2016) 447–453

Contents lists available at ScienceDirect

Neurocomputing

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

A survey of transfer learning for collaborative recommendation with auxiliary data

Weike Pan

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

ARTICLE INFO

Article history:
Received 2 March 2015
Received in revised form
9 November 2015
Accepted 23 November 2015
Communicated by Peng Cui
Available online 2 December 2015

Keywords:
Collaborative recommendation
Auxiliary data
Transfer learning

ABSTRACT

Intelligent recommendation technology has been playing an increasingly important role in various industry applications such as e-commerce product promotion and Internet advertisement display. Besides user feedbacks (e.g., numerical ratings) on items as usually exploited by some typical recommendation algorithms, there are often some additional data such as users' social circles and other behaviors. Such auxiliary data are usually related to user preferences on items behind numerical ratings. *Collaborative recommendation with auxiliary data* (CRAD) aims to leverage such additional information so as to improve personalized services. It has received much attention from both researchers and practitioners.

Transfer learning (TL) is proposed to extract and transfer knowledge from some auxiliary data in order to assist the learning task on the target data. In this survey, we consider the CRAD problem from a transfer learning view, especially on how to enable knowledge transfer from some auxiliary data, and discuss the representative transfer learning techniques. Firstly, we give a formal definition of transfer learning for CRAD (TL-CRAD). Secondly, we extend the existing categorization of TL techniques with three knowledge transfer strategies. Thirdly, we propose a novel and generic knowledge transfer framework for TL-CRAD. Fourthly, we describe some representative works of each specific knowledge transfer strategy in detail, which are expected to inspire further works. Finally, we conclude the survey with some summarized discussions and several future directions.

1. Introduction

Intelligent recommendation technology [1,4,18,31,45,48] has been a standard component embedded in many Internet systems such as e-commerce and advertisement systems to provide personalized services. There are two main approaches widely used in personalized recommendation for an active user, i.e., content-based recommendation [3] and collaborative recommendation [14]. Content-based methods promote an item based on the relevance between a candidate item and the active user's consumed items, while collaborative recommendation techniques focus on collective intelligence and exploit the community's data so as to recommend preferred items from users with similar tastes. However, both methods are limited to users' feedbacks of explicit scores or implicit examinations, which may result in a challenging problem, data sparsity, due to the lack of users' behaviors.

Fortunately, there are often some additionally available data besides the users' feedbacks (e.g., numerical ratings) in a recommender system. There are at least four types of auxiliary data as shown in Table 1, such as content information [52,56], time contextual information [23,36], social or information networks [21,49,54] and additional feedbacks [19,29,39]. These auxiliary data have the potential to help relieve the aforementioned sparsity problem and thus improve the recommendation performance. In this survey, we study on how to exploit different types of auxiliary data in collaborative recommendation, which is coined as *collaborative recommendation with auxiliary data* (CRAD).

Specifically, we study the CRAD problem from an *inductive transfer learning* [37] view (instead of unsupervised or transductive transfer learning views [2]), in which we consider the users' feedback data as our *target data* or supervised information, and all the other additional information as our *auxiliary data*. In particular, we focus on how to enable knowledge transfer from some auxiliary data to the target data in order to address the aforementioned sparsity challenge. We discuss some representative transfer learning techniques, aiming to answer the fundamental question of transfer learning [37], i.e., "how to transfer". With this focus in our survey, we extend previous categorization of transfer learning techniques in collaborative filtering [38,43], and answer the above question from two dimensions, including *knowledge transfer algorithm styles* (i.e., adaptive, collective and integrative knowledge transfer) and *knowledge transfer strategies* (i.e.,

E-mail address: panweike@szu.edu.cn

http://dx.doi.org/10.1016/j.neucom.2015.11.059
0925-2312/

prediction rule, regularization and constraint). Then, we propose a novel and generic knowledge transfer framework and describe some representative works in each category to answer the "how to transfer" question in detail, in particular the main idea that may be generalized to other applications. Finally, we conclude the survey with some summarized discussions and several exciting future directions.

2. Transfer learning for collaborative recommendation with auxiliary data

2.1. Problem definition

We have a target data set and an auxiliary data set. In the target data set, we have some feedbacks from n users and m items, which is usually represented as a rating matrix $\mathbf{R} = [r_{ui}]^{n \times m}$ and an indicator matrix $\mathbf{Y} \in \{0,1\}^{n \times m}$, where $y_{ui} = 1$ means that the feedback r_{ui} is observed. In the auxiliary data set, we have some additional data such as content, context, network and feedback information as shown in Table 1. Our goal is to predict the unobserved feedbacks in $\mathbf{R}$ by transferring knowledge from the available auxiliary data. We illustrate the studied problem in Fig. 1, where the left part is the target data of user feedbacks and the right part denotes different types of auxiliary data.

2.2. Categorization of transfer learning techniques

Following the fundamental question of "how to transfer" in transfer learning [37,43], we first categorize various transfer learning algorithms into (i) adaptive knowledge transfer, (ii) collective knowledge transfer and (iii) integrative knowledge transfer w.r.t. *knowledge transfer algorithm styles*. For each type of algorithm styles, we then study the related works in three specific *knowledge transfer strategies*, including (i) transfer via prediction rule, (ii) transfer via regularization and (iii) transfer via constraint, which are closely related to the three parts of a typical optimization problem [5], i.e., loss function, regularization and constraint.

Note that the binary categorization of adaptive knowledge transfer and collective knowledge transfer was first briefly described in [38], and was later expanded with one more category of integrative knowledge transfer in [43]. And in this survey, we further expand it with three specific knowledge transfer strategies in each algorithm style.

Table 1
List of auxiliary data.

Content
User's static profile of demographics, affiliations, etc.
Item's static description of price, brand, location, etc.
User–item pair's user generated content (UGC), etc.
Context
User's dynamic context of mood, health, etc.
Item's dynamic context of remaining quantities, etc.
User–item pair's dynamic context of time, etc.
Network
User–user social network of friendship, etc.
Item–item relevance network of taxonomy, etc.
User–item–user network of sharing items with friends, etc.
Feedback
User's feedback of rating on other items, etc.
Item's feedback of browsing by other users, etc.
User–item pair's feedback of collection, etc.

2.3. A generic knowledge transfer framework

We mainly survey some recent works of low-rank transfer learning methods for collaborative recommendation with auxiliary data (CRAD), in particular of matrix factorization based methods. The prosperity of matrix factorization based methods is mostly due to many successful stories in various public competitions and reported industry applications. Matrix factorization based methods are also the state-of-the-art in TL-CRAD because they are able to digest the sparse rating data well via learning latent variables and are also flexible to incorporate different types of auxiliary data.

Mathematically, matrix factorization based methods can be formulated with a loss function and a regularization term, i.e., $\min_{\Theta} \mathcal{E}(\Theta \mid \mathbf{R}) + \mathcal{R}(\Theta)$, where Θ is the model parameter. We extend such basic formulation and propose a novel and generic framework for TL-CRAD,

$$\min_{\Theta, \mathbb{K}} \quad \mathcal{E}(\Theta, \mathbb{K} \mid \mathbf{R}, \mathbb{A}) + \mathcal{R}(\Theta \mid \mathbb{K}, \mathbb{A}) + \mathcal{R}(\mathbb{K}),$$
$$\text{s.t.} \quad \Theta \in \mathcal{C}(\mathbb{K}, \mathbb{A}), \tag{1}$$

which contains a loss function $\mathcal{E}(\Theta, \mathbb{K} \mid \mathbf{R}, \mathbb{A})$, two regularization terms $\mathcal{R}(\Theta \mid \mathbb{K}, \mathbb{A})$ and $\mathcal{R}(\mathbb{K})$, and a constraint $\Theta \in \mathcal{C}(\mathbb{K}, \mathbb{A})$. Specifically, $\mathbf{R}$ is the target user–item rating matrix, $\mathbb{A}$ is the auxiliary data, $\mathbb{K}$ is the extracted knowledge from $\mathbb{A}$, and Θ is the model parameter. Note that the prediction rule is not explicitly shown but embedded in the loss function $\mathcal{E}(\Theta, \mathbb{K} \mid \mathbf{R}, \mathbb{A})$. In the following sections, we will describe some representative works of TL-CRAD, which are instantiations of the generic framework in Eq. (1).

3. Adaptive knowledge transfer

Adaptive knowledge transfer aims to *adapt* the knowledge extracted from an auxiliary data domain to a target data domain. This is a *directed* knowledge transfer approach similar to traditional domain adaptation methods. In this section, we describe two adaptive knowledge transfer strategies as instantiated from Eq. (1), including (i) transfer via regularization, $\min_{\Theta} \mathcal{E}(\Theta \mid \mathbf{R}) + \mathcal{R}(\Theta \mid \mathbb{K})$, and (ii) transfer via constraint, $\min_{\Theta} \mathcal{E}(\Theta \mid \mathbf{R})$, s.t. $\Theta \in \mathcal{C}(\mathbb{K})$.

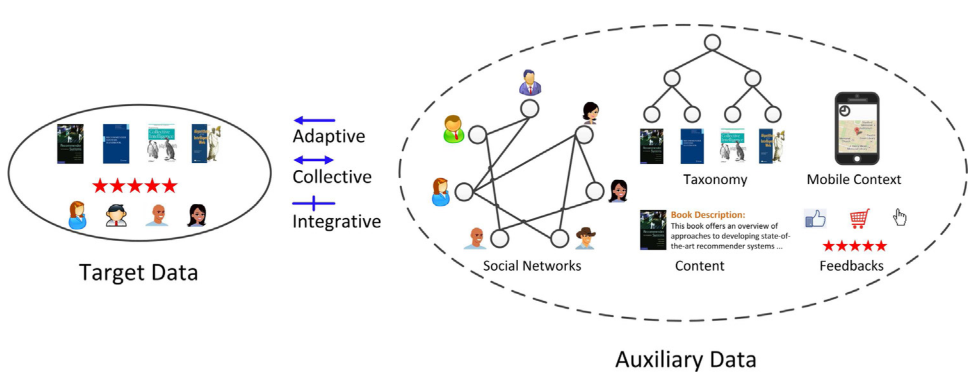

Fig. 1. Illustration of transfer learning for collaborative recommendation with auxiliary data (TL-CRAD).