Knowledge-Based Systems 73 (2015) 161–172

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

DualDS: A dual discriminative rating elicitation framework for cold start recommendation

Xi Zhang, Jian Cheng *, Shuang Qiu, Guibo Zhu, Hanqing Lu

National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

ARTICLE INFO

Article history:
Received 10 April 2014
Received in revised form 8 July 2014
Accepted 28 September 2014
Available online 13 October 2014

Keywords:
Cold start recommendation
Rating elicitation
Discriminative model
Dual regularization
Sparse regularization

ABSTRACT

Cold start problem is challenging because no prior knowledge can be used in recommendation. To address this cold start scenario, rating elicitation is usually employed, which profiles cold user or item by acquiring ratings during an initial interview. However, how to elicit the most valuable ratings is still an open problem. Intuitively, category labels which indicate user preferences and item attributes are quite useful. For example, category information can be served as a guidance to generate a set of queries which can largely capture the interests of cold users, and thus appealing recommendation lists are more likely to be returned. Therefore, we exploit category labels as supervised information to select discriminative queries. Furthermore, by exploring the correlation between users and items, a dual regularization is developed to jointly select optimal representatives. As a consequent, a novel *Dual Discriminative Selection* (DualDS) framework for rating elicitation is proposed in this paper, by integrating discriminative selection with dual regularization. Experiments on two real-world datasets demonstrate the effectiveness of DualDS for cold start recommendation.

1. Introduction

Recommender systems, as a sort of attractive tool on many famous online websites such as Netflix, YouTube, and Amazon, have become a popular platform to offer potential favorite items for users. Due to the promising performance, recommendation methods based on Collaborative Filtering (CF) are extensively studied in numerous literatures [1,2]. Nevertheless, these methods are not capable to give meaningful recommendation when it comes to cold entities[1] (users or items) with few collaborative information (i.e. ratings, clicks, purchases, etc.), which is named as cold start problem.

Several approaches have been proposed to alleviate this problem by making use of side information. Some of them extract features from auxiliary relationships such as social network for users [3,4], and others utilize additional attributes contents for items [5,6]. One shortcoming in above methods is that, extra data for cold entities are not always available on the web, even if there are generally plenty of meta data for warm entities. As an alternative, many recommender systems resort to *Rating Elicitation*, that is getting to know new users or items by ratings through an initial interview process. Specifically, the interview solicits tastes of cold users by querying them with carefully selected items, while judging characteristics of cold items by asking for opinions from elaborately chosen users. Therefore, the problem of rating elicitation converts to discovering the most informative users and items from warm entities, which can acquaint system with essential features of cold entities and ultimately improve accuracy of cold start recommendation. This paper aims to provide useful insights along this direction.

Faced with large amount of warm data resources in systems, most traditional methods select qualified items merely based on rating records while ignore other available knowledge [7–10]. However, category labels, as an indicator of users' interested topics, are quite helpful for the system to profile users. Considering a real-world scenario that if we plan to recommend movies to someone, the most natural question we might ask is "*what type of movie would you prefer*". The reason to ask this question is that understanding people preferences on topics can largely reflect their tastes on multiple facets and assist us to make a more effective recommendation. Consequently, interview conducted by recommender systems should benefit from the prior knowledge of categories. For this purpose, we exploit the category information as a guidance to select the most discriminative items such that the

* Corresponding author.
E-mail addresses: xi.zhang@nlpr.ia.ac.cn (X. Zhang), jcheng@nlpr.ia.ac.cn (J. Cheng), shuang.qiu@nlpr.ia.ac.cn (S. Qiu), gbzhu@nlpr.ia.ac.cn (G. Zhu), luhq@nlpr.ia.ac.cn (H. Lu).

[1] Our framework can tackle not only cold entities with few ratings but also entirely new entities without any rating.

http://dx.doi.org/10.1016/j.knosys.2014.09.015
0950-7051/

overall preferences are captured for users. Similarly, the most discriminative users can also be selected by introducing category labels to embody items' intrinsic topics. The architecture of our proposed rating elicitation process is shown in Fig. 1.

Additionally, existing rating elicitation algorithms such as [8] model user and item selection separately. In fact, the correlation between the two selection tasks should be taken into account. As we know, there exists some common subspace for both users and items to represent their low-dimensional features, which have been widely demonstrated by latent factor models in CF [1,2]. Furthermore, inspired by multi-task learning, knowledge of the two tasks can be transferred via the common feature subspace and mutually enhance each other. The crucial problem is how to establish such a shared subspace in rating elicitation context. It is worth to notice that the set of categories is same for users and items, such that the correlation between users and items can be explored via a common low-dimensional category space. For example, if a user gives a high rating score to the movie "Toy Story", she is probably interested in the categories "Comedy" and "Animation", which are close to "Toy Story" in category space. We summarize the above example into a hypothesis: if a user rates an item with a high score, it could indicate that the user and the item also have similar category labels. Based on the hypothesis, a dual regularizer is designed to bridge user selection task and item selection task into an integrated framework. Thereby, only one unified selection model needs to be trained in the two selection tasks, so as to informative users and items can be jointly mined.

With all above concerns, in this paper, we develop a novel Dual Discriminative Selection framework to elicit ratings for cold start recommendation, which is called as DualDS for short. For discriminative selection, a least square loss function with category as supervised label is exploited. To reduce redundant users or items in the selected set, $\ell_{2,1}$-norm constraint is incorporated into the objective function to learn a sparse set of representatives. Combining the $\ell_{2,1}$-norm regularized discriminative selection with the dual regularization, a unified framework DualDS is obtained. The major contributions of this work are as follows:

- Proposing a novel rating elicitation framework, DualDS, that integrates the $\ell_{2,1}$-norm regularized discriminative selection with dual regularization.
- Exploring the most discriminative users and items by encoding category labels in the selection process, so that the essential features of users or items are comprehensively exhibited.
- Introducing a dual selection strategy by modeling relation between users and items in category space, so that the users and items selection tasks can be jointly achieved.
- Evaluating DualDS extensively on two real-world datasets to verify that DualDS can improve the performance for cold start problem.

The reminder of the paper is organized as follows. In Section 2, some related works are summarized. Section 3 details a novel framework DualDS for representatives selection. Section 4 presents the method for cold start recommendation. We report the experimental results in Section 5 and conclude the paper in Section 6.

2. Related work

In the section, we review the related work on general cold start recommendation, especially for various rating elicitation approaches.

2.1. Cold start recommendation

From a general viewpoint, cold start scenarios not only refer to entirely new entities without ratings but also those with few ratings. Existing studies dealing with cold start problem mainly focus on three different strategies. The first one is incorporating additional attributes or contents from the profiles of entities (i.e. age, gender of users, or genre, director of movies) into a latent feature space, such that the lacked rating records can be compensated. For instance, the model [5] learns attribute-to-feature mappings from contents of entities to predict unknown latent factors and use them in matrix factorization. Collaborative Topic Regression [6] employs probabilistic topic modeling to analyze the contents of items (articles) including both warm and cold items. By combining the topic distribution with latent factor in traditional CF, the model can provide an interpretable latent structure meanwhile solving cold item problem.

The second strategy is extracting latent features from auxiliary relationship between entities (i.e. social networks). Relational learning method [11] describes diverse affiliations of users as latent social dimensions, which are extracted from social network. Then they regard social dimensions as features and cold start recommendation as a multi-label classification problem that can be solved by constructing a discriminative classifier. Also, in [3], the authors transfer information from auxiliary social relations using multi-relational factorization techniques [12]. Many other works using social network [13–15] aims to alleviate cold start problem as well as improve recommendation accuracy.

Based on above two strategies, there are models combining additional contents with auxiliary relationship. One example is the method in [16] which integrates Collaborative Topic Regression [6] with Social Recommendation [13]. In [17], the authors propose a unified model introducing kinds of external information via graph regularized nonnegative matrix factorization and show greatly improvement when collaborative information is sparse. In reality, however, for cold entity whose rating is usually missing, their extra meta information is probability absent. Therefore, the

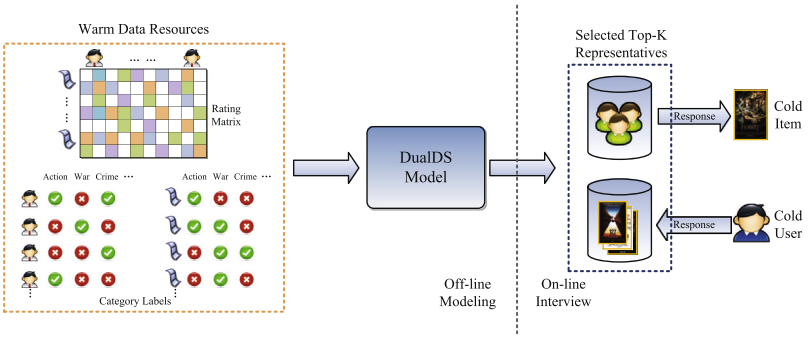

Fig. 1. The architecture of DualDS-based rating elicitation process for cold entities.