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

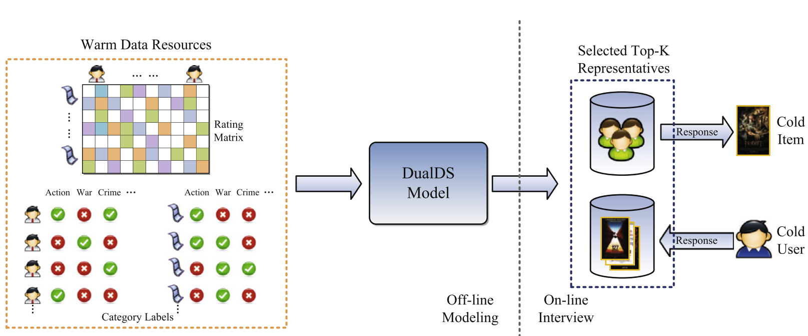


Fig. 1. The architecture of DualDS-based rating elicitation process for cold entities.

third strategy based on rating elicitation is widely studied in academia and commonly used in industry.

2.2. Rating elicitation

Rating elicitation is motivated by two crucial factors. On the one hand, it enlarges the set of available collaborative information for cold entities, because the more ratings are elicited, the more effective the recommendations are. On the other hand, it admits that ratings are not equally useful and different elicitation criteria often lead to significantly different performance. Hence, most of the rating elicitation methods place emphasis on developing specific selection criterion. Some early works [18,19] show the comparison of several basic criteria such as random, popularity, entropy and demonstrate the necessity of adopting a good elicitation method.

As an important research area, active learning can be connected with rating elicitation task in recommender systems. Active CF [20] casts the query selection problem by Expected Value Of Information and derive an interactive CF. The work [21] takes model uncertainty into account by averaging the expected loss function over the posterior distribution of models. Personalized active learning method [7] extends Bayesian active learning techniques for Aspect Model to generate personalized queries for users. But these methods are not efficient enough because they need to optimize selection criteria during the interview process. Another kind of methods is from the perspective of constructing a decision tree of queries. Golbandi et al. devise a greedy algorithm to build a query set [9], and then extent it to a tree-based bootstrapping process [22]. In [23], a decision tree with each node labeled by an item query is fitted on the basis of strategies given in [18]. Following this direction, several decision tree-based methods are developed [10,24]. By answering questions on the query tree, new users are classified into different leaf nodes that indicate their tastes.

Different with above methods, some methods define the selection problem as finding an optimal submatrix which can best represent the original rating matrix, and the users or items corresponding to the submatrix are selected as representatives. Even though the problem is NP-hard, some tractable approximations can be used. Liu et al. [8] resort to a mathematical method of searching for the maximal-volume submatrix in the latent space [25]. Transductive Experimental Design [26], which serves as general active learning problems, provides an efficient scheme by sequential greedy optimization and outputs questionnaire with selected representatives. More recently, [27] constructs an entity network by rating co-occurrence and determine the most influential seeds by PageRank and its variants.

Nevertheless, all above rating elicitation approaches only apply ratings of warm data while overlooks other useful information stored in real-world systems. In this study, we take full advantage of available warm data resources by leveraging rating matrix with category labels to select the most discriminative representatives. Besides, a variety of cold start recommendation methods based on the first two strategies [5,16,17] alleviate both cold user and item problems singly, whereas few works in rating elicitation can solve the two problems simultaneously. In contrast, our proposed framework DualDS is able to combine the two problems and solve them together.

3. Dual discriminative selection framework

In this section, we focus on: (1) how to select discriminative items and users; and (2) how to integrate the two selection tasks into one unified model. First of all, we define the problem of representative selection.

3.1. Problem formulation

Recommender system has a set of n warm users $\mathcal{U} = \{u_1, \ldots, u_n\}$ and a set of m warm items $\mathcal{V} = \{v_1, \ldots, v_m\}$. Given an observed rating matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$ for the users and items, each element X_{ij} describes the rating of user u_j to item v_i. Then row vector $\mathbf{x}_{i\cdot} \in \mathbb{R}^n$ and column vector $\mathbf{x}_{\cdot j} \in \mathbb{R}^m$ of $\mathbf{X}$ can be deemed as feature vectors for item v_i and user u_j.

Suppose there are k categories, $\mathcal{C} = \{c_1, \ldots, c_k\}$ is the label set which is the predefined item genre information such as romance, comedy for movies. For all the warm users, we denote their user-category relationship into a label matrix $\mathbf{Y}_1 \in \{0,1\}^{n \times k}$, such that $\mathbf{Y}_1(i,j) = 1$ if user u_i loves category c_j, and 0 otherwise. In addition, there is an indicator matrix $\mathbf{Y}_2$ for war items, where $\mathbf{Y}_2(i,j) = 1$ if item v_i belongs to category c_j. Note that the labels of users and items are defined by the same set $\mathcal{C}$.

With the defined notations, the task of *rating elicitation* can be formally stated as: *Given warm data including rating matrix* $\mathbf{X}$*, label matrix* $\mathbf{Y}_1$ *for users, and label matrix* $\mathbf{Y}_2$ *for items, the task is to respectively select informative user set* $\mathcal{U}_s$ *from* $\mathcal{U}$ *and item set* $\mathcal{V}_s$ *from* $\mathcal{V}$ *by building an unified selection model S as*

$$S : \{\mathcal{U}, \mathcal{V}; \mathbf{X}, \mathbf{Y}_1, \mathbf{Y}_2\} \to \{\mathcal{U}_s, \mathcal{V}_s\}$$

3.2. Discriminative item selection

In order to profile the tastes of a new user, a recommender system needs to query her/him with some informative items. Since ratings on an appealing queries are expected to capture user preferences on multiple facets as much as possible, the most discriminative items should be selected. To the aim, the category labels are utilized as supervised information to guide the selection process. Given the category labels $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n\} \in \mathbb{R}^k$ for n users, the problem of predicting the multi-faceted preference of users can be formulated as

$$\min_{\mathbf{P}} \sum_{i=1}^{n} \|\mathbf{P}^T \mathbf{x}_{\cdot i} - \mathbf{y}_i\|_2^2 \tag{1}$$

where $\mathbf{P} \in \mathbb{R}^{m \times k}$ is a projection matrix to map users into k-dimensional category space, such that the disagreement between the predicted preference distribution over categories $\mathbf{P}^T \mathbf{x}_{\cdot i}$ and true label $\mathbf{y}_i$ is minimized. Furthermore, to select the most discriminative items, a $\ell_{2,0}$-norm is added on the above least square regression problem as following

$$\min_{\mathbf{P}} \sum_{i=1}^{n} \|\mathbf{P}^T \mathbf{x}_{\cdot i} - \mathbf{y}_i\|_2^2 + \alpha \|\mathbf{P}\|_{2,0} \tag{2}$$

where $\ell_{2,0}$-norm restricts the number of nonzero rows of $\mathbf{P}$, which makes the model consistent with the intuitive explanation of item selection. However, as solving this problem with $\ell_{2,0}$-norm is NP-hard, we consider minimizing the tightest convex relation [28] of the $\ell_{2,0}$-norm

$$\min_{\mathbf{P}} \sum_{i=1}^{n} \|\mathbf{P}^T \mathbf{x}_{\cdot i} - \mathbf{y}_i\|_2^2 + \alpha \|\mathbf{P}\|_{2,1} \tag{3}$$

where the regularization term $\|\mathbf{P}\|_{2,1}$ jointly considers regression on k categories and ensures $\mathbf{P}$ is sparse in item dimension,[2] and the weight α controls the sparseness. With the predefined matrices $\mathbf{X}$ and $\mathbf{Y}_1$, Eq. (3) can be written into a matrix format as

$$\min_{\mathbf{P}} \|\mathbf{X}^T \mathbf{P} - \mathbf{Y}_1\|_F^2 + \alpha \|\mathbf{P}\|_{2,1} \tag{4}$$

[2] The $\ell_{2,1}$-norm is defined as $\|\mathbf{P}\|_{2,1} = \sum_{i=1}^{m} \sqrt{\sum_{j=1}^{k} \mathbf{P}^2(i,j)} = \sum_{i=1}^{m} \|\mathbf{P}(i,:)\|_2$.

Through sorting items according to $\|\mathbf{P}(i,:)\|_2$ in descending order, the items corresponding to the top-K ranked rows of $\mathbf{P}$ are selected as discriminative queries. Normally, category labels for items $\mathbf{Y}_2$ are easy to obtain from on-line websites, while category labels for users $\mathbf{Y}_1$ are often absent in most cases. For simplicity, we assign category labels by computing the rating distribution over categories for each user. That is, if a user has a quantity of ratings in a certain category, then we suggest that she is interested in the category.

3.3. Discriminative user selection

Similarly, to recommend a new item without any information, the system needs some user-offered ratings about this item. How to depict items well using as few ratings as possible is the central problem in rating elicitation. For this purpose, user selection aims to find an optimal set of users whose ratings are representative and can approximate the inherent attributes of items. As category information can directly describe the attribute of item, the label matrix $\mathbf{Y}_2$ is considered as a guidance in our model to select discriminative users. We formulate the procedure by the following problem

$$\min_{\mathbf{Q}} \|\mathbf{XQ} - \mathbf{Y}_2\|_F^2 + \alpha\|\mathbf{Q}\|_{2,1} \tag{5}$$

where $\mathbf{Q} \in \mathbb{R}^{n\times k}$ is a projection matrix to map items into k-dimensional category space, so that the disagreement between the learned attribute distribution over categories $\mathbf{XQ}$ and true label $\mathbf{Y}_2$ is minimized. After $\mathbf{Q}$ is learnt, we keep the users corresponding to the top-K ranked rows of $\mathbf{Q}$ to form an optimal user set. Then when there is a new item, judgements from chosen users are collected and recommendation can be made.

3.4. The proposed framework: DualDS

Most of the existing works address the rating elicitation for cold users and items separately while fail to consider the correlation between the two tasks. Clearly, the relations of user-item pairs are explicitly indicated by rating matrix $\mathbf{X}$, which can be viewed as a sort of mutual information connecting user and item selection process. Thanks to the symmetrical formulation introduced in Eqs. (4) and (5), users and items can be projected into a shared subspace with its dimensions related with categories, in which selection tasks may provide some useful knowledge to each other. With above reasons, we consider to develop a unified selection model which aims to: (a) extract users and items simultaneously; and (b) boost performance by learning projection matrices jointly. Consequently, a dual discriminative selection framework is proposed.

In our work, both of the one-sided selection models project users or items into one common low dimensional category space $\mathcal{C}$, hence there is a new constraint between projection matrices $\mathbf{P}$ and $\mathbf{Q}$. Concretely, a natural assumption of the constraint is that: *if item v_i and user u_j are strongly associated with a high rating score X_{ij}, they are more likely to have similar category labels*. Mathematically, the dual regularizer is formulated by minimizing the following loss function

$$\epsilon(\mathbf{P},\mathbf{Q}) = \sum_{i=1}^{m}\sum_{j=1}^{n} X_{ij}\left\|\frac{\mathbf{Q}^T\mathbf{x}_{i.}^T}{\sqrt{S_{ii}^{row}}} - \frac{\mathbf{P}^T\mathbf{x}_{.j}}{\sqrt{S_{jj}^{col}}}\right\|_2^2 \tag{6}$$

where $\mathbf{S}^{row} \in \mathbb{R}^{m\times m}$ is a diagonal degree matrix of items with its diagonal element $S_{ii}^{row} = \sum_{j=1}^{n} X_{ij}$, and $\mathbf{S}^{col} \in \mathbb{R}^{n\times n}$ is a diagonal degree matrix of users with $S_{jj}^{col} = \sum_{i=1}^{m} X_{ij}$.

By leveraging the two discriminative selection components with dual regularization component, we have the DualDS model with the parameter matrices $\mathbf{P}$ and $\mathbf{Q}$, which is formulated to solve the unified optimization problem

$$\min_{\mathbf{P},\mathbf{Q}} \lambda_1(\|\mathbf{X}^T\mathbf{P} - \mathbf{Y}_1\|_F^2 + \alpha\|\mathbf{P}\|_{2,1}) + \lambda_2(\|\mathbf{XQ} - \mathbf{Y}_2\|_F^2 + \alpha\|\mathbf{Q}\|_{2,1}) + \beta\sum_{i=1}^{m}\sum_{j=1}^{n} X_{ij}\left\|\frac{\mathbf{Q}^T\mathbf{x}_{i.}^T}{\sqrt{S_{ii}^{row}}} - \frac{\mathbf{P}^T\mathbf{x}_{.j}}{\sqrt{S_{jj}^{col}}}\right\|_2^2 \tag{7}$$

where the parameter λ_1 and λ_2 control weights of each selection task and the parameter β controls the contribution of dual regularizer. We use $\mathcal{L}(\mathbf{P},\mathbf{Q})$ to denote the objective function in Eq. (7). By finding optimal solution for $\mathbf{P}$ and $\mathbf{Q}$, the sets of informative user and item are selected simultaneously from the dyadic dimensions of rating matrix $\mathbf{X}$. The unified DualDS framework is illustrated in Fig. 2.

3.5. Optimization algorithm

While the minimization problem with $\ell_{2,1}$-norm has been studied in some feature selection works [28,29,11], it remains unclear how to optimize an objective function related with the dual regularizer. To solve the problem of Eq. (7), we present an efficient two-step algorithm, where the parameter matrices $\mathbf{P}$ and $\mathbf{Q}$ update iteratively and enhance each other mutually. For ease of optimization, we first derive an equivalent loss function for dual regularization term as follows:

Theorem 1. *The dual regularization in Eq. (7) is equivalent to the following loss function:*

$$\epsilon(\mathbf{P},\mathbf{Q}) = tr(\mathbf{P}^T\mathbf{X}\mathbf{X}^T\mathbf{P}) + tr(\mathbf{Q}^T\mathbf{X}^T\mathbf{X}\mathbf{Q}) - 2tr(\mathbf{Q}^T\mathbf{X}^T\mathbf{M}\mathbf{X}^T\mathbf{P}) \tag{8}$$

where $\mathbf{M}$ is defined as $\mathbf{M} = (\mathbf{S}^{row})^{-\frac{1}{2}}\mathbf{X}(\mathbf{S}^{col})^{-\frac{1}{2}}$.

Proof. To convert Eq. (7) into matrix form, two auxiliary matrices $\overline{\mathbf{P}} \in \mathbb{R}^{n\times k}$ and $\overline{\mathbf{Q}} \in \mathbb{R}^{m\times k}$ are introduced with their j-th row vector $\overline{\mathbf{p}}_j$ and i-th row vector $\overline{\mathbf{q}}_i$ satisfying $\overline{\mathbf{p}}_j = \mathbf{x}_{.j}^T\mathbf{P}$ and $\overline{\mathbf{q}}_i = \mathbf{x}_{i.}\mathbf{Q}$ respectively.

Then Eq. (7) becomes

$$\sum_{i=1}^{m}\sum_{j=1}^{n} X_{ij}\left\|\frac{\overline{\mathbf{q}}_i}{\sqrt{S_{ii}^{row}}} - \frac{\overline{\mathbf{p}}_j}{\sqrt{S_{jj}^{col}}}\right\|_2^2 = \sum_{i=1}^{m}\|\overline{\mathbf{q}}_i\|^2 + \sum_{j=1}^{n}\|\overline{\mathbf{p}}_j\|^2 - 2\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{X_{ij}\overline{\mathbf{q}}_i^T\overline{\mathbf{p}}_j}{\sqrt{S_{ii}^{row}S_{jj}^{col}}}$$
$$= tr\left(\overline{\mathbf{Q}}^T\overline{\mathbf{Q}}\right) + tr\left(\overline{\mathbf{P}}^T\overline{\mathbf{P}}\right) - 2tr\left(\overline{\mathbf{Q}}^T\mathbf{M}\overline{\mathbf{P}}\right)$$

where $\mathbf{M} = (\mathbf{S}^{row})^{-\frac{1}{2}}\mathbf{X}(\mathbf{S}^{col})^{-\frac{1}{2}}$, $\overline{\mathbf{P}} = \mathbf{X}^T\mathbf{P}$, and $\overline{\mathbf{Q}} = \mathbf{XQ}$. Eq. (8) can be obtained by substituting $\mathbf{P}$ and $\mathbf{Q}$ into above function, which completes the proof. □

With the equivalent form, we divide the objective function of Eq. (7) into two separate subproblems, then it can be minimized by updating $\mathbf{P}$ and $\mathbf{Q}$ in an alternating manner.

3.5.1. Optimize P given Q

When $\mathbf{Q}$ is fixed, the terms only involving $\mathbf{Q}$ in Eq. (7) can be omitted, which is reduced to the first subproblem with respect to $\mathbf{P}$,

$$\min_{\mathbf{P}} \lambda_1(\|\mathbf{X}^T\mathbf{P} - \mathbf{Y}_1\|_F^2 + \alpha\|\mathbf{P}\|_{2,1}) + \beta(tr(\mathbf{P}^T\mathbf{X}\mathbf{X}^T\mathbf{P}) - 2tr(\mathbf{Q}^T\mathbf{X}^T\mathbf{M}\mathbf{X}^T\mathbf{P})) \tag{9}$$

Theorem 2. *The minimization problems in Eq. (9) is equivalent to minimize function:*

$$\mathcal{L}_1(\mathbf{P}) = tr(\mathbf{P}^T\mathbf{G}_1\mathbf{P} - 2\mathbf{H}_1\mathbf{P}) + \lambda_1.\alpha\|\mathbf{P}\|_{2,1} \tag{10}$$

where

$$\begin{aligned}\mathbf{G}_1 &= (\lambda_1+\beta)\mathbf{X}\mathbf{X}^T\\ \mathbf{H}_1 &= \lambda_1\mathbf{Y}_1^T\mathbf{X}^T + \beta\mathbf{Q}^T\mathbf{X}^T\mathbf{M}\mathbf{X}^T\end{aligned} \tag{11}$$

Proof. By converting Frobenius norm into $tr(\cdot)$, the first term of Eq. (9) can be written as

$$\|\mathbf{X}^T\mathbf{P}-\mathbf{Y}_1\|_F^2 = \lambda_1 tr(\mathbf{P}^T\mathbf{X}\mathbf{X}^T\mathbf{P} - 2\mathbf{Y}_1^T\mathbf{X}^T\mathbf{P} + \mathbf{Y}_1^T\mathbf{Y}_1)$$

where term $tr(\mathbf{Y}_1^T\mathbf{Y}_1)$ is a constant and can be discarded. Substituting above equation into Eq. (9), $\mathcal{L}_1(\mathbf{P})$ is obtained. □

Inspired by [28], each subproblem can be solved by an alternating optimization strategy. Setting the derivation of $\mathcal{L}_1(\mathbf{P})$ as 0, and $\mathbf{P}$ can be computed by

$$\frac{\partial\mathcal{L}_1(\mathbf{P})}{\partial\mathbf{P}} = 2\mathbf{G}_1\mathbf{P} - 2\mathbf{H}_1^T + 2\alpha\mathbf{D}_P\mathbf{P} = 0 \Rightarrow \mathbf{P} = (\mathbf{G}_1+\alpha\mathbf{D}_P)^{-1}\mathbf{H}_1^T \tag{12}$$

where $\mathbf{D}_P$ is a diagonal matrix with $\mathbf{D}_P(i,i) = \frac{\lambda_1}{2\|\mathbf{P}(i,:)\|_2}$.

3.5.2. Optimize Q given P

Once given $\mathbf{P}$, we can calculate $\mathbf{Q}$ by discarding the irrelevant terms, Eq. (7) can be reduced to the second subproblem with respect to $\mathbf{Q}$,

$$\min_{\mathbf{Q}} \lambda_2(\|\mathbf{X}\mathbf{Q}-\mathbf{Y}_2\|_F^2 + \alpha\|\mathbf{Q}\|_{2,1}) + \beta(tr(\mathbf{Q}^T\mathbf{X}^T\mathbf{X}\mathbf{Q}) - 2tr(\mathbf{P}^T\mathbf{X}\mathbf{M}^T\mathbf{X}\mathbf{Q})) \tag{13}$$

Algorithm 1. DualDS.

Input: $\{\mathbf{X},\mathbf{Y}_1,\mathbf{Y}_2,\alpha,\beta,\lambda_1,\lambda_2,K\}$
Output: K most discriminative users and items
1: Compute $\mathbf{G}_1,\mathbf{G}_2$ and $\mathbf{M}$;
2: Initialize $\mathbf{D}_P^t$ as an identity matrix;
3: Initialize $\mathbf{Q}^t$ as a random matrix and compute initial $\mathbf{D}_Q^t$;
4: **while** not convergent **do**
5: Compute $\mathbf{H}_1$ in Eq. (11) by $\mathbf{Q}^t$;
6: Update $\mathbf{P}$ by $\mathbf{P}^{t+1} = (\mathbf{G}_1+\alpha\mathbf{D}_P^t)^{-1}\mathbf{H}_1^T$;
7: Compute $\mathbf{H}_2$ in Eq. (15) by $\mathbf{P}^{t+1}$;
8: Update $\mathbf{Q}$ by $\mathbf{Q}^{t+1} = (\mathbf{G}_2+\alpha\mathbf{D}_Q^t)^{-1}\mathbf{H}_2^T$;
9: Update $\mathbf{D}_P$ by $\mathbf{D}_P^{t+1}(i,i) = \frac{\lambda_1}{2\|\mathbf{P}^{t+1}(i,:)\|_2}$;
10: Update $\mathbf{D}_Q$ by $\mathbf{D}_Q^{t+1}(i,i) = \frac{\lambda_2}{2\|\mathbf{Q}^{t+1}(i,:)\|_2}$;
11: **end while**
12: Sort items and users respectively according to $\|\mathbf{P}(i,:)\|_2$ and $\|\mathbf{Q}(i,:)\|_2$ in descending order, and then select the top-K ranked ones;

Furthermore, it can be verified that Eq. (13) is equivalent to minimize following function

$$\mathcal{L}_2(\mathbf{Q}) = tr(\mathbf{Q}^T\mathbf{G}_2\mathbf{Q} - 2\mathbf{H}_2\mathbf{Q}) + \lambda_2\cdot\alpha\|\mathbf{Q}\|_{2,1} \tag{14}$$

where

$$\begin{aligned}\mathbf{G}_2 &= (\lambda_2+\beta)\mathbf{X}^T\mathbf{X}\\ \mathbf{H}_2 &= \lambda_2\mathbf{Y}_2^T\mathbf{X} + \beta\mathbf{P}^T\mathbf{X}\mathbf{M}^T\mathbf{X}\end{aligned} \tag{15}$$

Thus the objective functions $\mathcal{L}_1(\mathbf{P})$ and $\mathcal{L}_2(\mathbf{Q})$ have been presented by the similar formula. And $\mathbf{Q}$ can also be computed by setting $\frac{\partial\mathcal{L}_2(\mathbf{Q})}{\partial\mathbf{Q}} = 0$

$$\mathbf{Q} = (\mathbf{G}_2+\alpha\mathbf{D}_Q)^{-1}\mathbf{H}_2^T \tag{16}$$

where $\mathbf{D}_Q$ is a diagonal matrix with $\mathbf{D}_Q(i,i) = \frac{\lambda_2}{2\|\mathbf{Q}(i,:)\|_2}$. The inverse of Eqs. (12) and (16) exist, for the reason that $\mathbf{G}_1+\alpha\mathbf{D}_P$ and $\mathbf{G}_2+\alpha\mathbf{D}_Q$ are positive definite matrices. Besides, the value of $\mathbf{D}_P$ relies on $\mathbf{P}$ while the value of $\mathbf{D}_Q$ relies on $\mathbf{Q}$, thus the four matrices can be updated alternately until convergence. Based on above analysis, the detailed optimization procedure for DualDS is summarized in Algorithm 1. The convergence of our algorithm is proved by the following theorem.

Theorem 3. *The updating rules of Algorithm 1 monotonically decreases the objective function $\mathcal{L}(\mathbf{P},\mathbf{Q})$ given by Eq. (7) in each iteration.*

Proof. With the similar proof process presented in [28,30], an inequality chain is obtained

$$\mathcal{L}(\mathbf{P}^0,\mathbf{Q}^0) \geqslant \mathcal{L}(\mathbf{P}^1,\mathbf{Q}^0) \geqslant \mathcal{L}(\mathbf{P}^1,\mathbf{Q}^1)\ldots \tag{17}$$

Since $\mathcal{L}(\mathbf{P},\mathbf{Q}) \geqslant 0$, Algorithm 1 converges, which completes the proof. The detail is shown in Appendix A. □

To empirically demonstrate above analysis, the convergence curves over two datasets used in this paper are shown in Fig. 3. As we can see, the values of objective function decline quickly and approximate minima within 10 iterations on the two datasets.

4. Cold start recommendation

With the outputted discriminative users and items from DualDS, the cold start recommendation can be conducted for new users and new items. Note that the recommendation method for cold items is similar to the one for cold users, thus we place emphasis on cold user recommendation here.

Fig. 4 illustrates the flowchart of recommendation for cold users. Without loss of generality, a new user is denoted as u_{n+1}. In on-line interview process, the new user need to sequentially answer the orderly queries formed by selected items which are outputted by the model introduced in Section 3. Accordingly a column vector $\widetilde{\mathbf{x}}_{n+1}$ with ratings on those queries is generated. Note that new users may do not know about(have no rating on) the chosen items during interviews, then we do not use the rating of these items in prediction. Following the previous works [8,26], we do not retrain the selection model with the updated ratings of cold entities and only predict ratings by the obtained $\widetilde{\mathbf{x}}_{n+1}$ and a precalculated coefficient matrix $\mathbf{W}$.

To estimate ranking scores, the coefficient matrix $\mathbf{W}$ is computed through warm data to reconstruct the original rating matrix $\mathbf{X}$. For warm users, a rating matrix whose rows corresponding to the selected items is obtained from $\widetilde{\mathbf{X}} = \mathbf{X}(\mathcal{V}_s,:) \in \mathbb{R}^{K\times n}$. Here we use ridge regression to learn the coefficient matrix $\mathbf{W}^* \in \mathbb{R}^{K\times m}$ by

$$\mathbf{W}^* = \arg\min_{\mathbf{W}} \frac{1}{2}\|\mathbf{X}-\mathbf{W}^T\widetilde{\mathbf{X}}\|_F^2 + \frac{\gamma}{2}\|\mathbf{W}\|_F^2 \tag{18}$$

The optimum can be obtained by $\mathbf{W}^* = \left(\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^T + \gamma\mathbf{I}\right)^{-1}\widetilde{\mathbf{X}}\mathbf{X}^T$, where $\mathbf{I}\in\mathbb{R}^{K\times K}$ is an identity matrix. Then personalized ranking scores for new user u_{n+1} is generated by $\mathbf{W}^{*T}\widetilde{\mathbf{x}}_{n+1}$. Since $\mathbf{W}^*$ can be precomputed off-line. That is, when there is a new user with interview results, our method can predict rankings without retraining the entire model, which is coherent with online-updating principle of a real-world system.

5. Experiments

In this section, we investigate the performance of DualDS in top-N recommendation task for cold user and cold item scenarios,

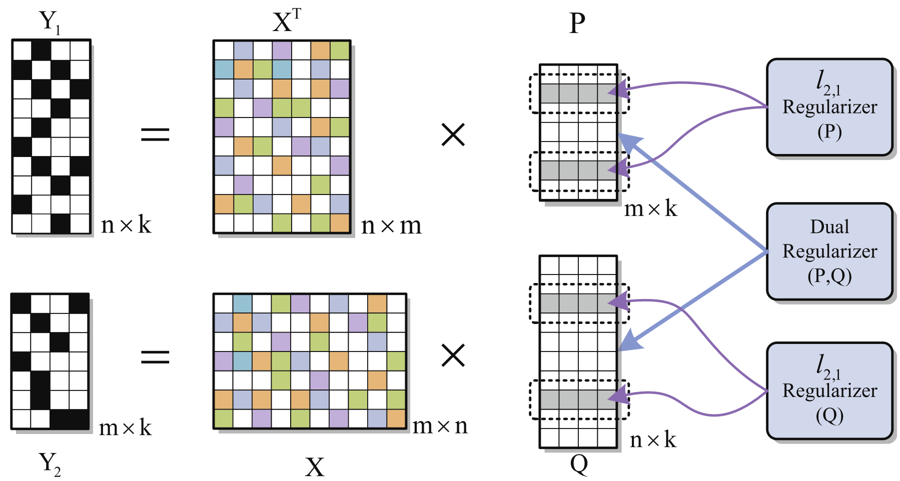


Fig. 2. Illustration of the DualDS framework.

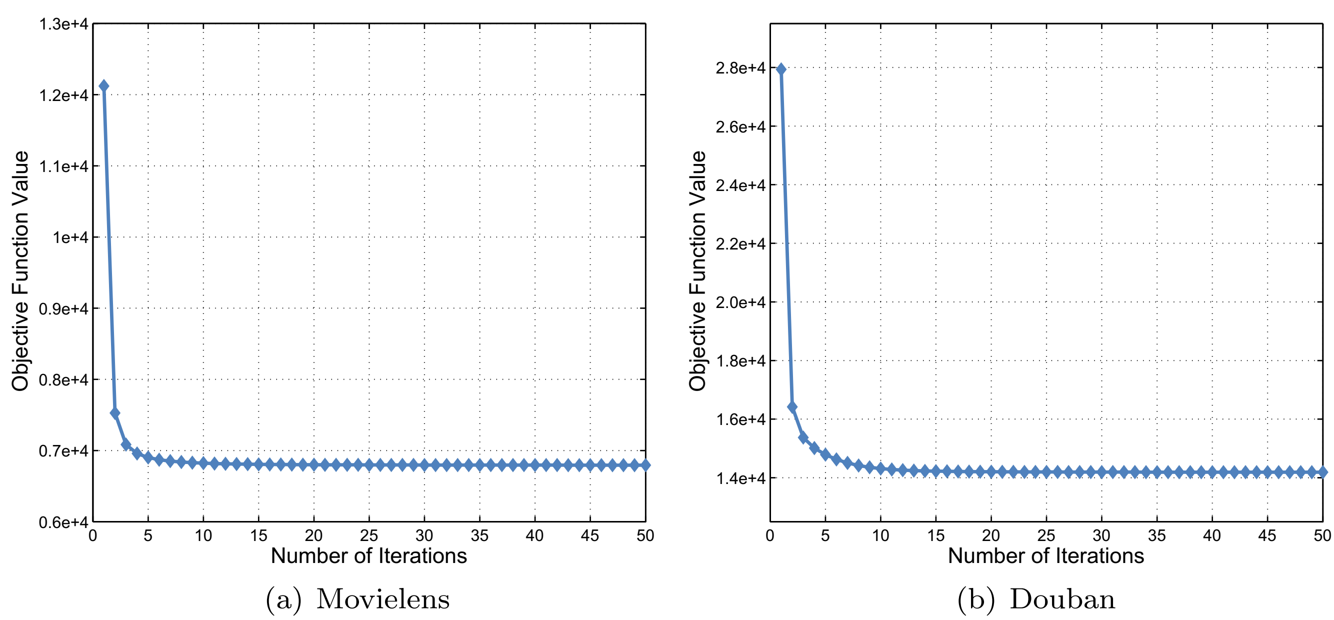


Fig. 3. Empirical demonstration of convergence for DualDS on Movielens and Douban.

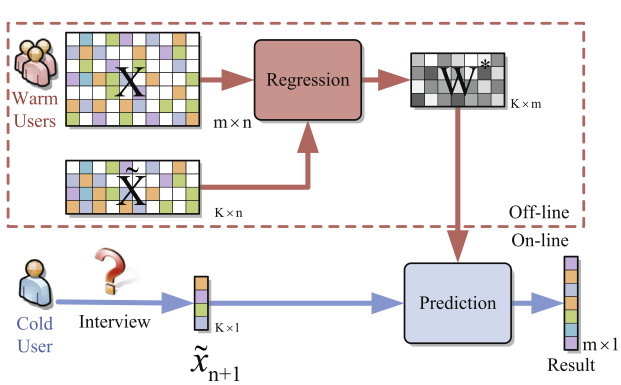


Fig. 4. Flowchart of recommendation for cold user.

respectively. To demonstrate the effectiveness of the proposed framework, DualDS is evaluated in comparison with several representative rating elicitation approaches on two real-world datasets.

5.1. Datasets

We examine our method on two movie rating datasets: MovieLens[3] and Douban.[4] In both datasets, category information about movies is available. The statistics of the two datasets are shown in Table 1.

5.1.1. Movielens

Movielens is a well-known benchmark movie recommendation dataset. It includes 1,000,209 ratings ranged from 1 to 5 points which are given by 6040 users to 3952 movies. There are 18 predefined categories that used as labels in our work.

5.1.2. Douban

Douban is a widely used Chinese Web 2.0 site which allows users to express their opinions on movies, books, and music

[3] http://www.grouplens.org/.
[4] http://www.douban.com/.

Table 1
Statistics of the datasets.

	Movielens	Douban
# Of users	6040	8629
# Of items	3952	4688
# Of categories	18	19
# Of ratings	1,000,209	1,859,339
Rating sparsity	95.81%	95.40%

Table 2
Statistics of label similarity to support hypothesis of dual regularizer.

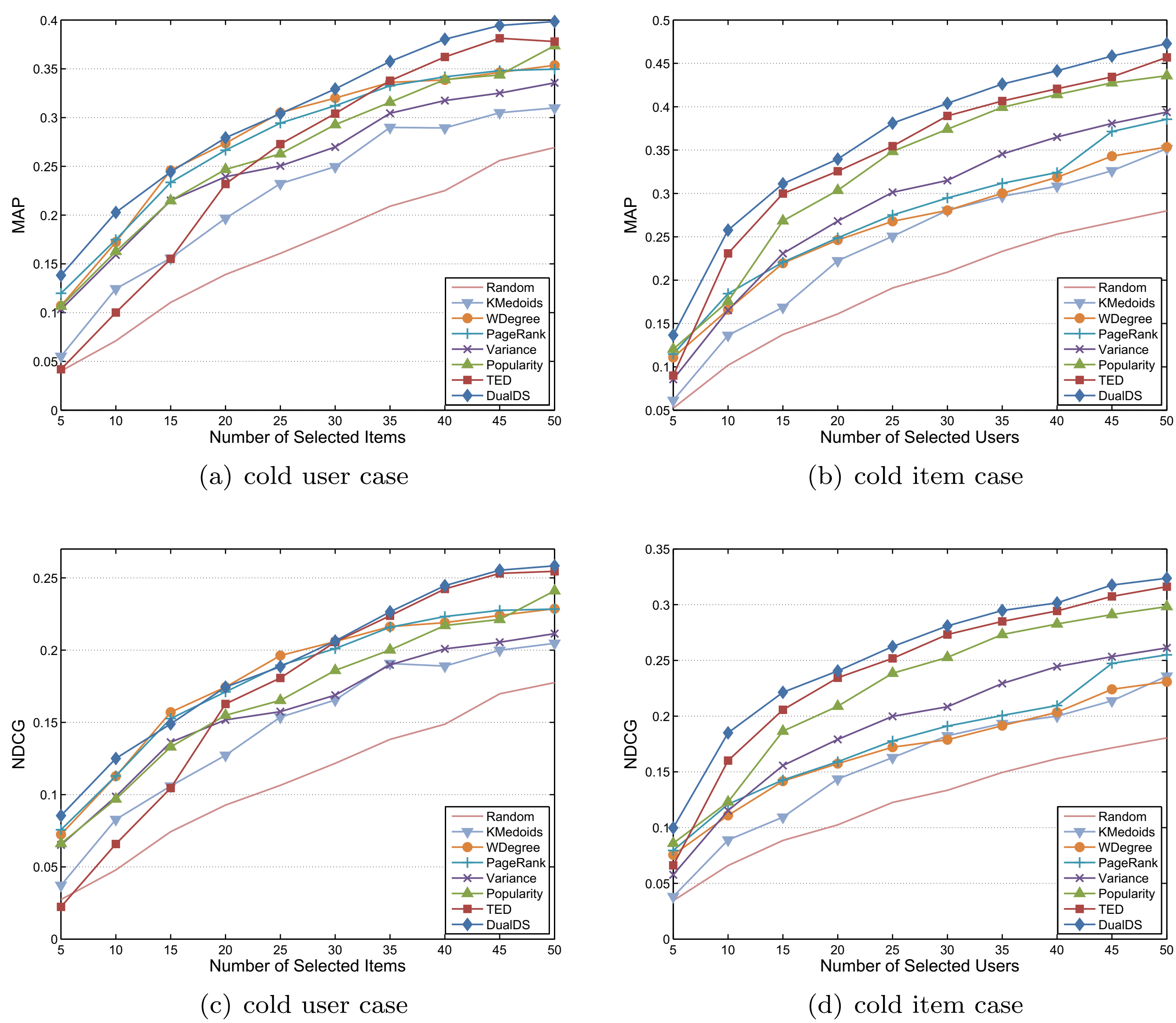
	Movielens	Douban
p-Value	1.0574e−191*	9.1245e−60*

through ratings or comments. The rating scale is also from 1 to 5. We crawled a dataset from Douban in the month of June, 2013 and then collect a movie subset which contains 1,859,339 ratings from 8629 users to 4688 movies. Similar to Movielens, the categories in the dataset are defined as "Action", "Comedy", "Romance", and so forth. There are 19 categories in total.

5.2. Preliminary verification

One of the major contributions of our framework is that it can jointly select optimal user sets and item sets. The pivotal component to bridge the two tasks is dual regularizer, which is based on a fundamental hypothesis. Before conducting experiments for cold start recommendation, it is necessary to validate if the hypothesis is reasonable. To analyze the relation (similarity of category labels) between items and users, we first define a label distance for an item-user pair (v_i, u_j). As aforementioned, Y_1 and Y_2 are indicator matrix to store the label of users and items. Let $Y_2(i,:)$ is the label vector for item v_i, and $Y_1(j,:)$ is the label vector for user u_j. The label distance between v_i and u_j is defined as

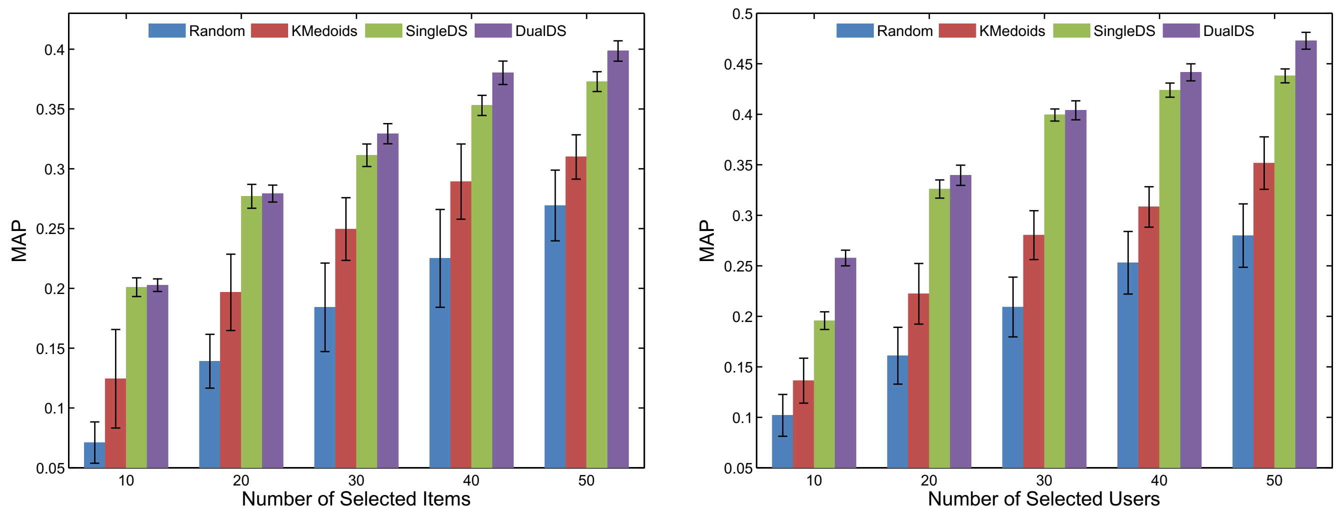
$$\mathcal{T}_{dist}(v_i, u_j) = \|Y_2(i,:) - Y_1(j,:)\|_2$$



(a) cold user case (b) cold item case

(c) cold user case (d) cold item case

Fig. 5. Cold start recommendation performance measured by MAP and NDCG with different rating elicitation approaches on Movielens ($\lambda_1 = 0.5, \lambda_2 = 0.5, \alpha = 500, \beta = 1$).



(a) cold user case (b) cold item case

Fig. 6. Performance comparison measured by MAP with respects to two baselines as well as two variants of our model SingleDS and DualDS on Movielens ($\pm std$, 10 repeated times).

For each item v_i, $\mathbf{s}_t(i)$ and $\mathbf{s}_r(i)$ is calculated as follows: the former is to compute the average distance $\mathcal{T}_{dist}$ between v_i and users with high rating on v_i,[5] and the latter is to compute the average distance $\mathcal{T}_{dist}$ between v_i and randomly picked users without high rating on v_i. We control the number of randomly chosen users to keep the same length for vectors $\mathbf{s}_t$ and $\mathbf{s}_r$.

With the defined vectors $\mathbf{s}_t$ and $\mathbf{s}_r$, we form a null hypothesis H_0: there is no difference between relational data and random data, that is $\mathbf{s}_t = \mathbf{s}_r$; and an alternative hypothesis, H_1: the average label distance with a relation of high rating is less than without, that is $\mathbf{s}_t < \mathbf{s}_r$. By performing a two-sample t-test on $\{\mathbf{s}_t, \mathbf{s}_r\}$, we can see that there is strong evidence (p-value< 0.01) to reject the null hypothesis. In Table 2, it can be observed that p-values on both Movielens and Douban are 1.0574e-191 and 9.1245e-60. The evidence suggests that: *with high probability, item-user pairs with relation of high ratings have smaller $\mathcal{T}_{dist}$ than those without*. Based on the hypothesis testing, we further check how the dual regularizer assists recommendation.

5.3. Experimental setup

In our experiments, the performance of rating elicitation for cold user case and cold item case is tested separately. As the cold start entity only exists in practice, we first have to partition the observed data to simulate cold users or items. In both recommendation tasks, 80% users and 80% items are picked as warm data, and the rest is cold data. Moreover, a 50–50 disjoint split is used for each cold entity, where 50% ratings are randomly chosen as response set to acquire ratings during interviews, the other 50% ratings used for evaluation. The experiments are repeated 10 times by randomly sampling the response and evaluation set, and the average performance are reported.

Then we determine the parameters through cross-validation. Particularly, the resulting parameters are: $\{\lambda_1 = 0.5, \lambda_2 = 0.5, \alpha = 500, \beta = 1\}$ for Movielens, and $\{\lambda_1 = 0.6, \lambda_2 = 0.4, \alpha = 500, \beta = 1\}$ for Douban. The impact of $\ell_{2,1}$-norm regularizer parameter α and dual regularizer parameter β to our full model DualDS will be further discussed in Section 5.6.

[5] the value of rating is higher than 3 point.

5.4. Baselines and metrics

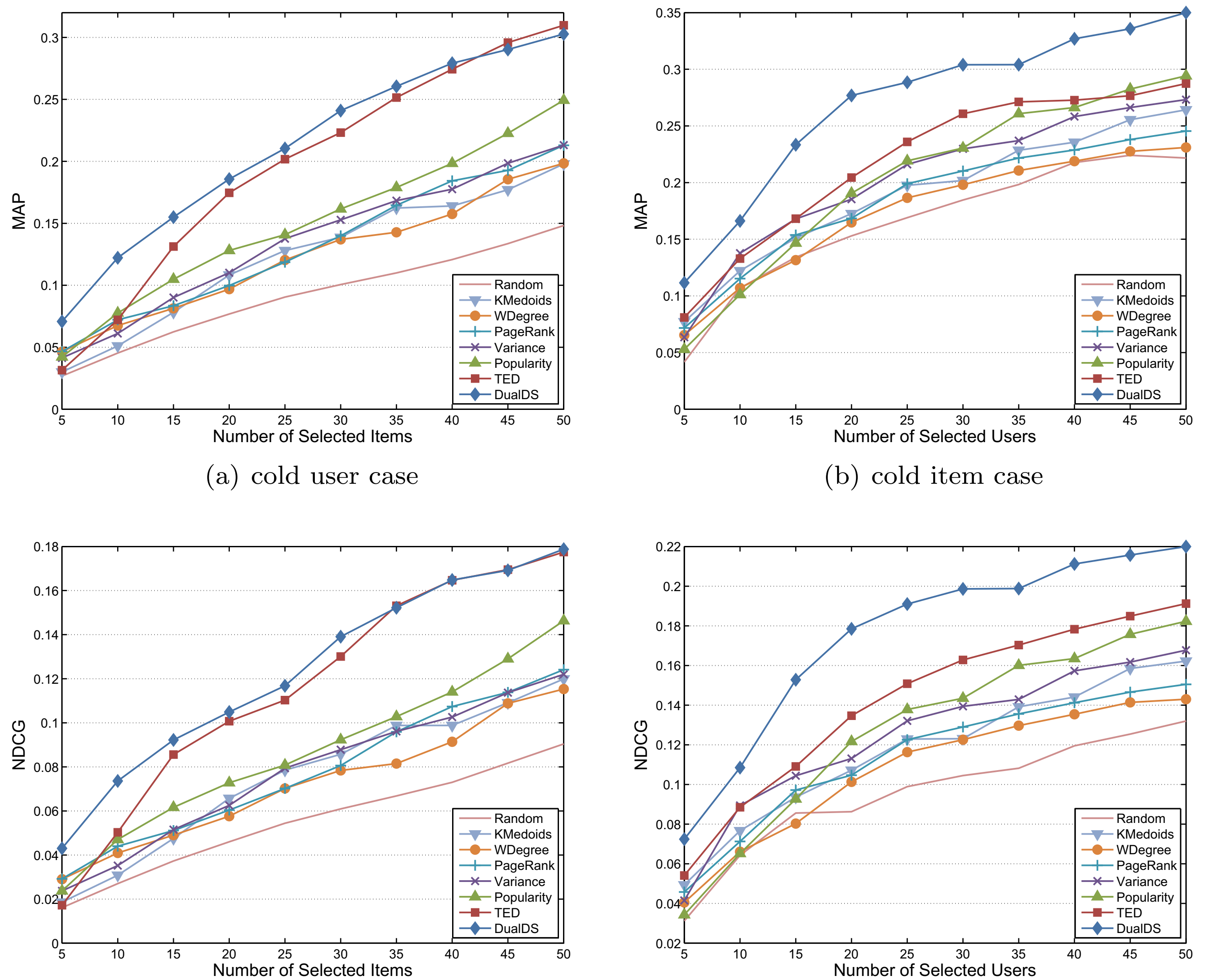
DualDS is compared with following representative rating elicitation methods. To make a fair comparison, we predict ranking scores for all the strategies by the same cold start recommendation approach used in Section 4.

To make it is clear, we introduce the baselines from the perspective of cold user recommendation. **Random**: randomly draws items from the whole warm item set $\mathcal{V}$ to constitute a query set. **KMedoids**: items into K groups and then select the cluster centers as representative queries. **Popularity**: the category information is embedded in the popularity criterion. The query list is comprised of popular items extracted form multiple categories and thus chosen queries are diverse. **Variance** [9]: assumes items with more diverse ratings are more useful. According to [9], the diversity is combined with popularity by multiplying variance with the square root of popularity. **WDegree** [27]: first constructs an item network by rating similarity matrix $\mathbf{A}$, the *WDegree* value for each item is defined as $WDegree(\mathcal{V}_i) = \sum_j^m A_{ij}$. The larger value of *WDegree* means the more influential the item is. **PageRank** [27]: Similar to WDegree, an item network is generated and the top-K items are computed by PageRank. Since we do not use time information, the network is undirected. **TED**: Transductive experimental design(TED) [26] is an active learning method which can be used in questionnaire design for cold start recommendation via select informative products.

To evaluate the predicted ranking scores, two classical evaluation metrics in top-N recommendation are employed: MAP (Mean Average Precision), NDCG (Normalized Discounted Cumulative Gain). For scenarios of cold users and cold items recommendation, user-oriented and item-oriented evaluations are utilized. Here we present user-oriented definition of these measures, and the item-oriented metrics is defined by a similar way.

Mean Average Precision (MAP). For each user, Average Precision(AP) is first defined as

$$\mathrm{AP}(u) = \frac{\sum_{i=1}^{N} \mathrm{prec}(i) \times \mathrm{pref}(i)}{\#\ \text{of preferred items}}$$



(a) cold user case (b) cold item case

(c) cold user case (d) cold item case

Fig. 7. Cold start recommendation performance measured by MAP and NDCG with different rating elicitation approaches on Douban ($\lambda_1 = 0.6, \lambda_2 = 0.4, \alpha = 500, \beta = 1$).

where prec(i) is precision and pref(i) is a binary preference indicator at ranked position i. MAP is computed based on AP by the following equation

$$\mathrm{MAP} = \frac{1}{|\mathcal{U}|}\sum_{u\in\mathcal{U}}\mathrm{AP}(u)$$

Normalized Discounted Cumulative Gain (NDCG). For a ranked list of N item, NDCG is computed by

$$\mathrm{NDCG} = \frac{1}{\mathrm{IDCG}} \times \sum_{i=1}^{N} \frac{2^{\mathrm{pref}(i)-1}}{\log_2(i+1)}$$

where IDCG denotes DCG of a perfect ranking algorithm. We report the results when the length of returned list N equals 5 in this paper.

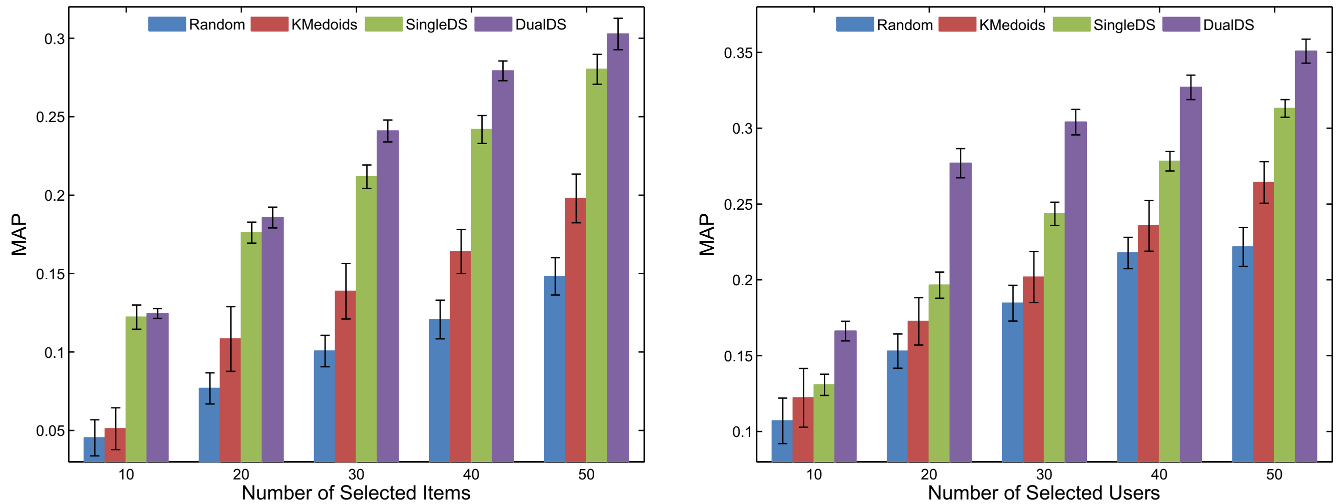
5.5. Performance evaluation

In this subsection, we discuss recommendation qualities for cold users and cold items respectively. In each recommendation task, we keep increasing the number of selected users or items K from 5 to 50 so as to avoid boring interviews, and compare the recommendation accuracy under different rating elicitation approaches.

5.5.1. Performance on movielens

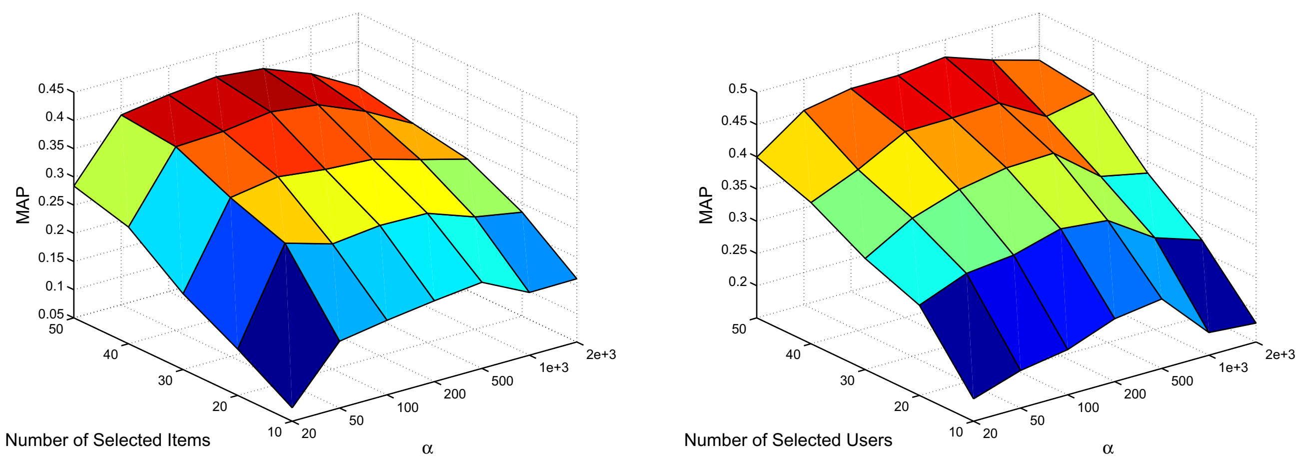
The comparison results for cold user and cold item recommendation task with two different metrics MAP and NDCG are shown in Fig. 5. From the figures, the overall trends of performance along with the length of interview can be observed for all compared methods. Basically, when the number of selected items or users K is getting larger, the accuracy increases rapidly, which is in accordance with a simple principle: the longer the interview is, the more we know about the new entity and the better recommendation we can give. Also, it is easy to observe that different selection method results in entirely different accuracy, which indicates that designing a rating elicitation approach is imperative.

For a detailed comparison in Fig. 5, our proposed model DualDS yields the best performance under almost all of the evaluation conditions, whereas Random strategy performs severely worse than other approaches. First, among the baselines, TED, which is an active learning method to find representative set, has high performance when there is a long interview, while fails to cope with



(a) cold user case (b) cold item case

Fig. 8. Performance comparison measured by MAP with respects to two baselines as well as two variants of our model SingleDS and DualDS on Douban ($\pm std$, 10 repeated times).

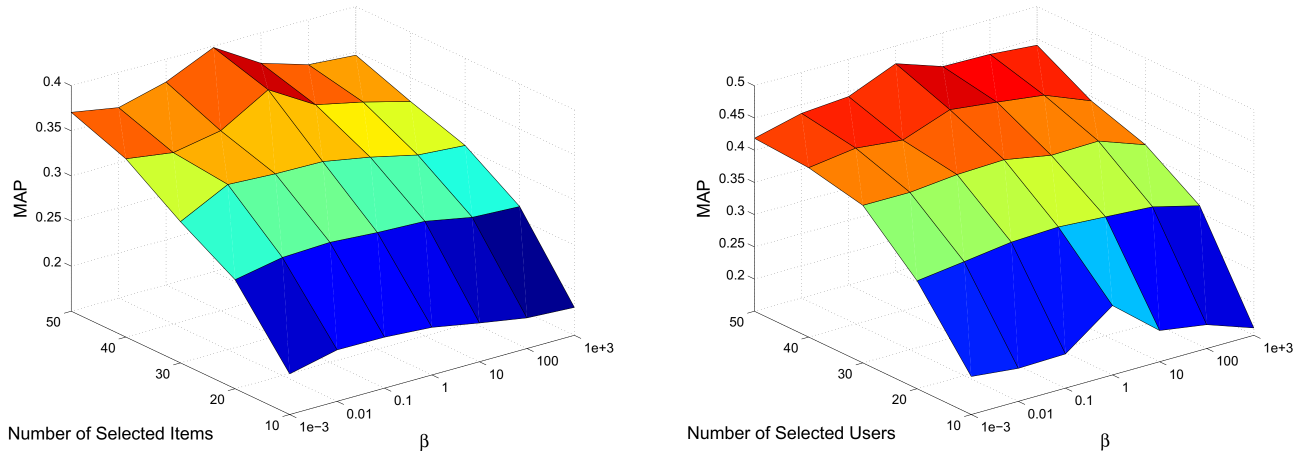


(a) cold user case (b) cold item case

Fig. 9. Performance variation of DualDS with respect to weight of $\ell_{2,1}$-norm regularizer α.

the circumstance of a short interview ($K = 5$) especially for cold user case. As we know, it is quite difficult to capture the multiple facets of user preference by merely asking five questions. Compared with TED, DualDS can always keep better performance, suggesting that the supervised guidance does benefit on detecting the most discriminative queries to overcome the difficult circumstance. Second, by comparing with the Popularity strategy where the category labels are also used, the consistently better performance of DualDS illustrates that our designed model is more effective than simply drawing the most popular representatives from each category. Third, both of WDegree and PageRank select representatives based on their influence abilities, therefore they always behave comparably. For Movielens, they can handle cold user case better than cold item case, which infers that the constructed item network on this dataset is more helpful to mine the influential nodes than user network. Finally, we compare Variance and Popularity to certify that the selected set should be multi-faceted. As commonly used approaches in cold user case, Variance concerns not only popularity but also rating diversity. But only choosing queries whose ratings are controversial is not sufficient to ensure that the whole query set is diversity. In contrast, Popularity integrates multiple category information and performs better than Variance.

Now we show the benefit of the unified model DualDS by comparing it with SingleDS (Single Discriminative Selection model) in Fig. 6. SingleDS, which is a variant of DualDS without dual regularizer, utilizes the formulation introduced in Section 3.2 and Section 3.3 to deal with cold user case and cold item case separately. Experimental results show that DualDS outperforms SingleDS. For example, DualDS is better than SingleDS by 6.87% in cold user case and by 7.96% in cold item case at the end of the interview ($K = 50$). Therefore, it can be empirically verified that the dual regularizer make the two selection tasks mutually enhance each other and thus bring about accuracy improvement. Meanwhile, the comparison of standard deviation with respect to two baselines and our methods is shown in Fig. 6. For our methods and most of the baselines, standard deviation is mainly caused by the random split of datasets. However, random strategy is inner randomness and KMedoids heavily relies on initialization, which lead to their standard deviation over 10 runs is much larger than the two variant of our methods.



(a) cold user case (b) cold item case

Fig. 10. Performance variation of DualDS with respect to weight of dual regularizer β.

5.5.2. Performance on Douban

As for Douban, Fig. 7 shows that DualDS outstands among all the comparison methods in most experimental cases. In particular, although DualDS and TED behave comparably when K is large in cold user case, DualDS outperforms all of the baselines with a significant margin in cold item case. Different from the performance on Movielens, KMedoids works better than WDegree and PageRank in most of the time. One possible reason is that there exists more representative users and items on the Website and can be easily discovered by clustering algorithm. At last, Fig. 8 illustrates the comparison between SingleDS and DualDS on Douban, and DualDS still delivers a superior performance. At the end of the interview, DualDS outperforms SingleDS by 8.03% and by 12.08% in cold user and item case respectively.

5.6. Parameter analysis

To study the two important parameters in DualDS: $\ell_{2,1}$-norm regularizer weight α and dual regularizer weight β, we analyze each of them by keeping the others fixed. The analysis of α and β on Movielens is reported here and the parameter selection process for Douban is conducted in a similar way.

Firstly, setting $\beta = 1$, we vary α as $\alpha \in \{20, 50, 100, 200, 500, 1000, 2000\}$. The performance variation of MAP with respect to α is presented in Fig. 9. In cold user case, with the increase of α, the performance first increases rapidly, then reaches the highest performance at $\alpha = 200$ and $\alpha = 500$, and then decreases. In cold item case, it is clear that DualDs achieve the peak point with $\alpha = 500$. Since the model is supposed to be trained once for the both cases, we choose $\alpha = 500$ eventually. Besides, the significance performance change indicates that the sparseness regularizer $\|\mathbf{P}\|_{2,1}$ and $\|\mathbf{Q}\|_{2,1}$ have a great influence on our full model.

Next, fixing α as $\alpha = 500$, we vary β as $\beta \in \{0.001, 0.01, 0.1, 1, 10, 100, 1000\}$, which balances the impact of dual regularizer. The results are shown in Fig. 10. In both cold user and item cases, as we increase β, MAP also raises and reaches the highest point at $\beta = 1$. A large β strengthens the effect of dual regularizer. Yet, when β is increasingly large, the dual regularizer would overwhelm the discriminative term as well as the $\ell_{2,1}$-norm regularizer, and thus the performance drops. When $\beta = 0$, the model boils down to SingleDS. From Fig. 10 we can see that, dual regularizer is effective for improving the final performance, which supports the result of hypothesis testing.

6. Conclusions

In this paper, we study rating elicitation method for cold-start recommendation. By identifying the main problem as user and item selection, a novel Dual Discriminative Selection (DualDS) framework is proposed. The DualDS is capable to mine a set of items which can reflect user preferences as well as a set of users which can describe item characteristics. With the designed dual regularizer, the two selection tasks are integrated into one unified model. An alternating minimization algorithm is derived to optimize the dual regularized problem. Finally, we demonstrate the effectiveness of our model in two cold start cases with real-world datasets. The evaluation shows that, compared with competitors, our model can offer a more accurate recommendation for cold entity.

Acknowledgements

This work was supported in part by 863 Program (Grant No. 2014AA015100), National Natural Science Foundation of China (Grant Nos. 61170127, 61332016).

Appendix A

A.1. Proof of Theorem 3.3

Proof. For $\mathbf{P}$ and $\mathbf{Q}$, we update one meanwhile keep the other fixed. With $\mathbf{Q}^t$ fixed, we first verify

$$\mathcal{L}(\mathbf{P}^{t+1}, \mathbf{Q}^t) \leqslant \mathcal{L}(\mathbf{P}^t, \mathbf{Q}^t) \tag{19}$$

which equals to $\mathcal{L}_1(\mathbf{P}^{t+1}) \leqslant \mathcal{L}_1(\mathbf{P}^t)$. It can easily to prove that Eq. (12) is the solution to the following problem

$$\min_{\mathbf{P}} tr(\mathbf{P}^T(\mathbf{G}_1 + \alpha\mathbf{D}_P)\mathbf{P} - 2\mathbf{H}_1\mathbf{P})$$

That is, in the t iteration,

$$\mathbf{P}^{t+1} = \arg\min_{\mathbf{P}} tr(\mathbf{P}^T(\mathbf{G}_1 + \alpha\mathbf{D}_P^t)\mathbf{P} - 2\mathbf{H}_1\mathbf{P})$$

which means that

$$tr((\mathbf{P}^{t+1})^T(\mathbf{G}_1 + \alpha\mathbf{D}_P^t)\mathbf{P}^{t+1} - 2\mathbf{H}_1\mathbf{P}^{t+1})$$
$$\leqslant tr((\mathbf{P}^t)^T(\mathbf{G}_1 + \alpha\mathbf{D}_P^t)\mathbf{P}^t - 2\mathbf{H}_1\mathbf{P}^t)$$

$$\Rightarrow tr((\mathbf{P}^{t+1})^T\mathbf{G}_1\mathbf{P}^{t+1} - 2\mathbf{H}_1\mathbf{P}^{t+1}) + \lambda_1 \cdot \alpha\sum_i \frac{\|\mathbf{P}^{t+1}(i,:)\|_2^2}{2\|\mathbf{P}^t(i,:)\|_2}$$
$$\leqslant tr((\mathbf{P}^t)^T\mathbf{G}_1\mathbf{P}^t - 2\mathbf{H}_1\mathbf{P}^t) + \lambda_1 \cdot \alpha\sum_i \frac{\|\mathbf{P}^t(i,:)\|_2^2}{2\|\mathbf{P}^t(i,:)\|_2}$$

Then we have

$$tr((\mathbf{P}^{t+1})^T\mathbf{G}_1\mathbf{P}^{t+1} - 2\mathbf{H}_1\mathbf{P}^{t+1}) + \lambda_1 \cdot \alpha\sum_i\|\mathbf{P}^{t+1}(i,:)\|_2$$
$$- \lambda_1 \cdot \alpha\left(\sum_i\|\mathbf{P}^{t+1}(i,:)\|_2 - \sum_i \frac{\|\mathbf{P}^{t+1}(i,:)\|_2^2}{2\|\mathbf{P}^t(i,:)\|_2}\right)$$
$$\leqslant tr((\mathbf{P}^t)^T\mathbf{G}_1\mathbf{P}^t - 2\mathbf{H}_1\mathbf{P}^t) + \lambda_1 \cdot \alpha\sum_i\|\mathbf{P}^t(i,:)\|_2$$
$$- \lambda_1 \cdot \alpha\left(\sum_i\|\mathbf{P}^t(i,:)\|_2 - \sum_i \frac{\|\mathbf{P}^t(i,:)\|_2^2}{2\|\mathbf{P}^t(i,:)\|_2}\right)$$

According to the inequality $\sqrt{a} - \frac{a}{2\sqrt{b}} \leqslant \sqrt{b} - \frac{b}{2\sqrt{b}}$ verified in [28], the following inequality holds

$$\sum_i\|\mathbf{P}^{t+1}(i,:)\|_2 - \sum_i \frac{\|\mathbf{P}^{t+1}(i,:)\|_2^2}{2\|\mathbf{P}^t(i,:)\|_2} \leqslant \sum_i\|\mathbf{P}^t(i,:)\|_2 - \sum_i \frac{\|\mathbf{P}^t(i,:)\|_2^2}{2\|\mathbf{P}^t(i,:)\|_2}$$

Therefore, we have the following inequality

$$tr((\mathbf{P}^{t+1})^T\mathbf{G}_V\mathbf{P}^{t+1} - 2\mathbf{H}_V\mathbf{P}^{t+1}) + \lambda_1 \cdot \alpha\|\mathbf{P}^{t+1}\|_{2,1}$$
$$\leqslant tr((\mathbf{P}^t)^T\mathbf{G}_V\mathbf{P}^t - 2\mathbf{H}_V\mathbf{P}^t) + \lambda_1 \cdot \alpha\|\mathbf{P}^t\|_{2,1}$$

which indicates that $\mathcal{L}_1(\mathbf{P})$ monotonically decreases during iterations. Similarly, $\mathcal{L}_2(\mathbf{Q}^{t+1}) \leqslant \mathcal{L}_2(\mathbf{Q}^t)$ can also be proved in the same way. With $\mathbf{P}^t$ fixed, the objective function holds

$$\mathcal{L}(\mathbf{P}^t, \mathbf{Q}^{t+1}) \leqslant \mathcal{L}(\mathbf{P}^t, \mathbf{Q}^t) \tag{20}$$

Combining Eqs. (19) and (20), we eventually obtain

$$\mathcal{L}(\mathbf{P}^{t+1}, \mathbf{Q}^{t+1}) \leqslant \mathcal{L}(\mathbf{P}^{t+1}, \mathbf{Q}^t) \leqslant \mathcal{L}(\mathbf{P}^t, \mathbf{Q}^t)$$

As a result, the objective function of Eq. (7) monotonically decreases using the update rules of Algorithm 1 in each iteration. Because the value of $\mathcal{L}(\mathbf{P}, \mathbf{Q})$ for DualDS is positive, the iterative algorithm can converge to an optimal solution. □

Appendix B. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.knosys.2014.09.015.

References

[1] R. Salakhutdinov, A. Mnih Probabilistic matrix factorization, in: NIPS, 2008.
[2] Y. Koren, Factorization meets the neighborhood: a multifaceted collaborative filtering model, in: KDD, 2008, pp. 426–434.
[3] A. Krohn-Grimberghe, L. Drumond, C. Freudenthaler, L. Schmidt-Thieme, Multi-relational matrix factorization using bayesian personalized ranking for social network data, in: WSDM, 2012.
[4] L. Tang, H. Liu, Relational learning via latent social dimensions, in: KDD, 2009.
[5] Z. Gantner, L. Drumond, C. Freudenthaler, S. Rendle, L. Schmidt-Thieme, Learning attribute-to-feature mappings for cold-start recommendations, in: ICDM, 2010.
[6] C. Wang, D.M. Blei, Collaborative topic modeling for recommending scientific articles, in: KDD, 2011.
[7] A.S. Harpale, Y. Yang, Personalized active learning for collaborative filtering, in: SIGIR, 2008.
[8] N.N. Liu, X. Meng, C. Liu, Q. Yang, Wisdom of the better few: cold start recommendation via representative based rating elicitation, in: ACM Conference on Recommender Systems, 2011.
[9] N. Golbandi, Y. Koren, R. Lempel, On bootstrapping recommender systems, in: CIKM, 2010.
[10] K. Zhou, S.-H. Yang, H. Zha, Functional matrix factorizations for cold-start recommendation, in: SIGIR, 2011.
[11] J. Tang, H. Liu, Unsupervised feature selection for linked social media data, in: KDD, 2012.
[12] A.P. Singh, G.J. Gordon, Relational learning via collective matrix factorization, in: KDD, 2008.
[13] H. Ma, H. Yang, M.R. Lyu, I. King, Sorec: social recommendation using probabilistic matrix factorization, in: CIKM, 2008.
[14] M. Jamali, M. Ester, A matrix factorization technique with trust propagation for recommendation in social networks. in: ACM Conference on Recommender Systems, 2010.
[15] P. Massa, P. Avesani, Trust-aware recommender systems, in: ACM Conference on Recommender Systems, 2007.
[16] S. Purushotham, Y. Liu, C.-C.J. Kuo, Collaborative topic regression with social matrix factorization for recommendation systems.
[17] Q. Gu, J. Zhou, C. Ding, Collaborative filtering: weighted nonnegative matrix factorization incorporating user and item graphs, 2010.
[18] A.M. Rashid, I. Albert, D. Cosley, S.K. Lam, S.M. McNee, J.A. Konstan, J. Reidl Konsan, Getting to know you: learning new user preferences in recommender systems, in: IUI, 2002, pp. 127–134.
[19] G. Carenini, J. Smith, D. Poole, Towards more conversational and collaborative recommender systems, in: IUI, 2003.
[20] C. Boutilier, R.S. Zemel, B. Marlin, Active collaborative filtering, in: UAI, 2003.
[21] R. Jin, L. Si, A bayesian approach toward active learning for collaborative filtering, in: UAI, 2004.
[22] N. Golbandi, Y. Koren, R. Lempel, Adaptive bootstrapping of recommender systems using decision trees, in: WSDM, 2011.
[23] A.M. Rashid, G. Karypis, J. Riedl, Learning preferences of new users in recommender systems: an information theoretic approach, SIGKDD Explor. Newsl.
[24] M. Sun, F. Li, J. Lee, K. Zhou, G. Lebanon, H. Zha, Learning multiple-question decision trees for cold-start recommendation, in: WSDM, 2013.
[25] S.A. Goreinov, I.V. Oseledets, D.V. Savostyanov, E.E. Tyrtyshnikov, N.L. Zamarashkin, How to Find a Good Submatrix, Tech. rep., 2008.
[26] K. Yu, J. Bi, V. Tresp, Active learning via transductive experimental design, in: ICML, 2006.
[27] Q. Liu, B. Xiang, E. Chen, Y. Ge, H. Xiong, T. Bao, Y. Zheng, Influential seed items recommendation, in: ACM Conference on Recommender Systems, 2012.
[28] F. Nie, H. Huang, X. Cai, C. Ding, Efficient and robust feature selection via joint l2,1-norms minimization, in: NIPS, 2010.
[29] Y. Yang, H.T. Shen, Z. Ma, Z. Huang, X. Zhou, l2,1-norm regularized discriminative feature selection for unsupervised learning, in: IJCAI, 2011.
[30] J. Tang, H. Liu, Feature selection with linked data in social media, in: SIAM International Conference on Data Mining, 2012.