COMMUNITY DISCOVERING GUIDED COLD-START RECOMMENDATION: A DISCRIMINATIVE APPROACH

Shuang Qiu, Jian Cheng, Xi Zhang, Biao Niu, Hanqing Lu

National Lab of Pattern Recognition, CASIA, Beijing, China, 100190
{shuang.qiu, jcheng, xi.zhang, bniu, luhq}@nlpr.ia.ac.cn

ABSTRACT

Recommendation for new users is a key challenge due to the lack of prior information from them, which is the well-known cold-start problem. Preference elicitation has been proposed as an efficient strategy for eliciting new users preference through an initial interview where new users are queried by elaborately selected items. In this paper, we propose a novel community discovering guided discriminative selection (CDDS) model for constructing query set. We exploit the community as an effective information which is not fully used in existing approaches. By integrating item selection and community discovery into one framework, our model selects most discriminative items for preference elicitation, with guidance of unsupervised community discovering process. To perform community discovering process, the model utilizes rating similarity graph and social network as a graph regularization. Experimental results on real-world datasets Flixster and Douban demonstrate that the proposed method outperforms traditional preference elicitation methods for cold-start recommendation.

***Index Terms*—** cold-start recommendation, community discovering, preference elicitation

1. INTRODUCTION

Recommender systems(RS), as an indispensable service for Internet users' online life, are increasingly playing an critical role in coping with information overload. Thanks to RS, the "best" items matching with individual tastes will be provided once users' historical behaviors are recorded and analyzed, which can enhance user satisfaction and boost the sales on websites. Due to the practical importance, recommendation algorithms, especially Collaborative Filtering(CF) based methods, have been widely studied in many literatures [1, 2]. By utilizing observed past ratings or purchases as collaborative information, the CF based systems can perform an effective recommendation. However, it fails when a new user comes to the system with no available preference information, which is the well-known cold start problem.

Providing effective cold-start recommendation is crucial to commercial websites. Many cold-start recommendation methods have been proposed to alleviate this urgent problem. Some methods employ meta data(e.g. user profile) for infer the taste of new users by similar users [3, 4]. Nevertheless, many users are not willing to make their real private information open to the public in consideration of privacy protection in a real-world scenario, which causes meta data not always available. As an alternative, many recommender system resort to a strategy named *Preference Elicitation* [5, 6, 7] which attempts to ask new users to provide ratings for the elaborately selected *seeds items*(the initial items offered to new users) through an interview process. By the interview process with seed items, a new user's personalized preference could be elicited and his ratings on overall items can be predicted. Therefore, the cold-start problem is alleviated via preference elicitation.

The challenging problem on how to discover the taste of new users accurately using fewer seed items for interviews has attracted many studies. Some naive selection criteria such as random, popularity, entropy [5] are proposed to determine the seed items for users to rate. Moreover, decision tree-based frameworks are introduced to devise preference elicitation methods [5, 8]. A new user will enter into a leaf node revealing his preference with responding to questions along the query tree. Methods based on active learning [6] are also proposed, where a personalized query items are generated by different active learning strategies. Recently, the method in [7] focus on a new perspective of consumption behaviors and find the most influential items by PageRank on a constructed item network.

Although the past preference elicitation methods make full use of the observed rating behaviors, they have ignored an essential component, the *Community*, as a latent pattern commonly existed in user-based online systems. As we know, community is normally defined as a group of users whose tastes are inherently similar to each other. For any users, their tastes could be discovered once we know which communities they belong to. Hence community information is valuable for selecting seed items which can largely reflect a user's tastes. Nevertheless, users' communities are not explicitly provided in real-world systems. Therefore, we propose a framework which can select seed items guided by a community discovering process. In other words, our proposed method

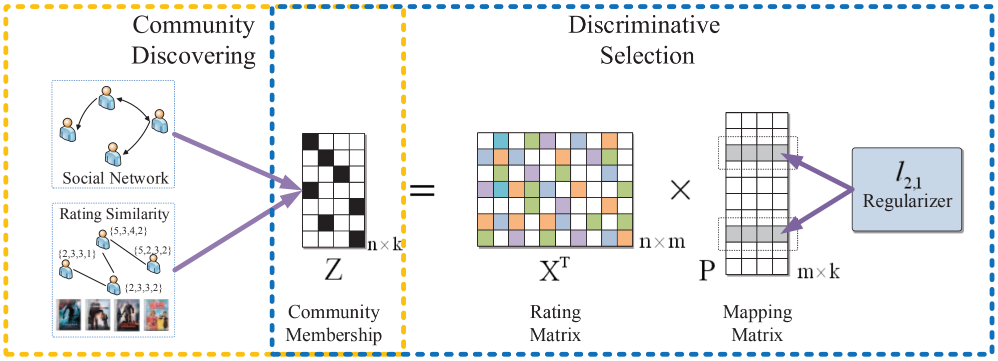


Fig. 1: Illustration of CDDS Framework

integrates community discovering with seed item selection together and mutually enhance each other to select appropriate item as queries.

Additionally, the community membership of users can be regarded as a sort of pseudo label to guide query selection. With these label guidance, our framework is capable to select the most discriminative items which distinguish new users' tastes broadly. To the best of our knowledge, the discriminative approach based on community mining has not been considered in previous preference elicitation works.

With above concerns, a novel **C**ommunity **D**iscovering guided **D**iscriminative **S**election(CDDS) framework is developed to elicit preference for cold-start recommendation. Our framework of CDDS is conceptually illustrated in Fig. 1. Specifically, for the discriminative item selection, a least square loss function with users' community membership as guidance is exploited. We utilize user social network and rating similarity graph for community discovery. Guided by the community discovering process, we adopt $\ell_{2,1}$-norm into the objective function for reducing the redundant items. Combining the $\ell_{2,1}$ regularized discriminative select, a unified framework CDDS is obtained.

2. COMMUNITY DISCOVERING GUIDED DISCRIMINATIVE SELECTION MODEL

2.1. Problem Statement

In recommender systems we have a set of user $\mathcal{U} = \{u_1, u_2, ..., u_n\}$ where n is the number of users and a set of items $\mathcal{V} = \{v_1, v_2, ..., v_m\}$ where m is the number of items. Given the observed rating matrix $\mathbf{R} = (\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_n) \in \mathbb{R}^{m \times n}$, each element $\mathbf{R}_{ij}$ is the rating of user u_j to item v_i. In a real world scenario, the rating matrix $\mathbf{R}$ is highly sparse and most items are rated by only a few users. Thereby as defined in [9], a candidate pool $\mathcal{V}_p = \{v_1, ..., v_{m_p}\}$ is extracted from $\mathcal{V}$ with $m_p < m$ by filtering out the long-tailed items. A submatrix $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n) \in \mathbb{R}^{m_p \times n}$ of $\mathbf{R}$ is also constructed where each element $\mathbf{X}_{ij}$ denotes the rating of user u_j to item v_i in candidate pool $\mathcal{V}_p$.

A social trust network usually exists in the form of a directed graph with 'trust' and 'trusted' behaviors among users. We denote the social network $\mathbf{T} \in \{0, 1\}^{n \times n}$ as an asymmetric matrix where $\mathbf{T}_{ij} = 1$ if u_i trusts u_j or u_j is trusted by u_i and 0 otherwise.

Based on the defined annotation above, the problem of selecting discriminative items can be stated as: given a warm data including rating matrix $\mathbf{R}$ for user set $\mathcal{U}$ and social network indicator matrix $\mathbf{T}$, the goal is to select most representative items set $\mathcal{V}_s$ from candidate pool $\mathcal{V}_p$ guided by a process of grouping users into different communities, where the size of $\mathcal{V}_s$ is d. The item selection model f can be formulated as

$$f : \{\mathcal{V}_p; \mathbf{X}, \mathbf{T}\} \rightarrow \{\mathcal{V}_s\}$$

2.2. Discriminative Selection

Here we present discriminative item selection in detail. In order to capture the preference of new users, the most discriminative items need to be extracted from the candidate pool $\mathcal{V}_p$. Through ratings of the most discriminative items by new users, their tastes can best embodied or, from the perspective of communities, they can be best placed in different communities representing different user preference. However, the communities that a user belongs to is not easy to obtain and even more serious problems exist that communities in a real websites are heavily noisy and overlapped.

To ensure the selection is discriminative, pseudo labels defined by community membership of users is used. Specifically, we introduce the community indicator matrix $\mathbf{Z} \in \mathbb{R}^{n \times k}$ to model concept of community where k is the predefined number of communities and $\mathbf{Z}_{ij} = 1$ if u_i is in the j^{th} community, otherwise $\mathbf{Z}_{ij} = 0$. The constraints on $\mathbf{Z}$ are written as

$$\mathbf{Z} \in \{0, 1\}^{n \times k}, \quad \|\mathbf{Z}(i, :)\|_0 = 1, \quad \forall i, \ 1 \leq i \leq n$$

where $\| \cdot \|_0$ is zero norm counting the number of nonzeros in a vector, $\mathbf{Z}(i, :)$ denotes the i^{th} row of matrix $\mathbf{Z}$ and $\|\mathbf{Z}(i, :)\|_0 = 1$ is utilized to set the label of the most likely community a user u_i belongs to as 1 and others as 0. By introducing the the $\ell_{2,1}$-norm regularization for discriminative item selection, we can derive our formula as follows

$$\begin{aligned} \min_{\mathbf{P},\mathbf{Z}} \quad & \left\|\mathbf{X}^T\mathbf{P}-\mathbf{Z}\right\|_F^2+\beta\left\|\mathbf{P}\right\|_{2,1} \\ s.t. \quad & \|\mathbf{Z}(i,:)\|_0=1, \quad \forall i,\ 1\le i\le n \\ & \mathbf{Z}\in\{0,1\}^{n\times k} \end{aligned} \tag{1}$$

where $\|\cdot\|_F$ is Frobenius norm as $\|\mathbf{A}\|_F=\sqrt{\sum_{i,j}\mathbf{A}_{ij}^2}$ and $\mathbf{P}\in\mathbb{R}^{m_p\times k}$ is a mapping matrix. $\mathbf{P}$ tries to project user rating matrix to a k-dimensional space such that the disagreement of predicted community membership $\mathbf{X}^T\mathbf{P}$ and community indicator matrix $\mathbf{Z}$ is minimized. $\|\mathbf{P}\|_{2,1}$ is defined as

$$\|\mathbf{P}\|_{2,1}=\sum_{i=1}^{m_p}\sqrt{\sum_{j=1}^{k}\mathbf{P}_{ij}^2}=\sum_{i=1}^{m_p}\|\mathbf{P}(i,:)\|_2 \tag{2}$$

where $\ell_{2,1}$-norm controls the sparsity of item dimension, which is consistent with intuitive explanation of selection process. The items corresponding to the zero rows of $\mathbf{P}$ will be discarded such that the discriminative items could be selected [9]. Different with [9], we need to deal with an unsupervised scenario that the selection guidance $\mathbf{Z}$ is unknown. Consequently, it is necessary to employ the data of user behaviors to discover the community guidance.

2.3. Community Discovering

Intuitively, users with similar rating or trust behaviors should have a similar community indicator. With above assumption, a user graph regularization can be embedded to constrain the community indicator matrix $\mathbf{Z}$, which is formulated as the following minimization problem on the basis of the spectral analysis [10]

$$\min_{\mathbf{Z}}\sum_{i,j=1}^{n}\mathbf{S}_{ij}\|\frac{\mathbf{Z}(i,:)}{\sqrt{\mathbf{D}_{ii}}}-\frac{\mathbf{Z}(j,:)}{\sqrt{\mathbf{D}_{jj}}}\|_2^2=Tr(\mathbf{Z}^T\mathbf{L_s}\mathbf{Z}) \tag{3}$$

where $\mathbf{S}$ is user similarity matrix, $\mathbf{D}_{ii}=\sum_{j=1}^{n}\mathbf{S}_{ij}$ is a diagonal matrix and $\mathbf{L_s}=\mathbf{D}^{-1/2}(\mathbf{D}-\mathbf{A})\mathbf{D}^{-1/2}$ is normalized laplacian matrix. Hence, the users can be clustered by Eq.(3) into different communities according to similarity matrix $\mathbf{S}$.

In our formulation, user similarity can be expressed from two perspectives of user behaviors, user rating similarity and social trust network. For user rating similarity, we assume that similar users have similar ratings. On the one hand, we construct an undirected graph $\mathcal{G}$ with a similarity matrix $\mathbf{S_R}$ by users' ratings. The nodes in $\mathcal{G}$ is corresponding to user rating vectors $\{\mathbf{x}_1,\mathbf{x}_2,...,\mathbf{x}_n\}$. Thus $\mathbf{S_R}$ is computed by cosine similarity as follows

$$\mathbf{S_R}_{ij}=\begin{cases} cos(\mathbf{x}_i,\mathbf{x}_j), & \mathbf{x}_i\in\mathcal{N}(\mathbf{x}_j)\ or\ \mathbf{x}_j\in\mathcal{N}(\mathbf{x}_i) \\ 0, & otherwise \end{cases} \tag{4}$$

where $\mathcal{N}(\mathbf{x}_i)$ is the nearest neighbor set of user u_i. On the other hand, the homophily property social network assumes that similar users will build a trust relation in social network. Several real-world social networks are constructed with a directed graph $\mathcal{T}$ which is asymmetric. For simplicity, we compute the social similarity matrix $\mathbf{S_T}$ as follows

$$\mathbf{S_T}_{ij}=\begin{cases} 1, & u_i\ has\ bidirectional\ trust\ with\ u_j \\ \frac{1}{2}, & u_i\ has\ unidirectional\ trust\ with\ u_j \\ 0, & u_i\ has\ no\ trust\ with\ u_j \end{cases} \tag{5}$$

Once $\mathbf{S_R}$ and $\mathbf{S_T}$ are computed, we can obtain the following user similarity matrix with a coefficient μ to balance the two perspectives of similarity,

$$\mathbf{S}=\mu\mathbf{S_T}+(1-\mu)\mathbf{S_R} \tag{6}$$

2.4. The Unified Framework: CDDS

Now we consider to integrate the discriminative selection with community discovering process as guidance into a unified framework. With above preliminary formulation, a unified model is written as the following minimization problem.

$$\begin{aligned} \min_{\mathbf{P},\mathbf{Z}} \quad & Tr(\mathbf{Z}^T\mathbf{L_s}\mathbf{Z})+\alpha(\left\|\mathbf{X}^T\mathbf{P}-\mathbf{Z}\right\|_F^2+\beta\left\|\mathbf{P}\right\|_{2,1}) \\ s.t. \quad & \|\mathbf{Z}(i,:)\|_0=1, \quad \forall i,\ 1\le i\le n \\ & \mathbf{Z}\in\{0,1\}^{n\times k} \end{aligned} \tag{7}$$

Because of the constraints in Eq.(7), the minimization problem is mixed with integer programming, which is difficult to solve. Inspired by [11], scaled community indicator matrix can be introduced in our framework such that constraints in Eq.(7) can relaxed as $\mathbf{Z}^T\mathbf{Z}=\mathbf{I}$ with $\mathbf{Z}\ge 0$. We can eventually obtain our CDDS model as follows

$$\begin{aligned} \min_{\mathbf{P},\mathbf{Z}} \quad & Tr(\mathbf{Z}^T\mathbf{L_s}\mathbf{Z})+\alpha(\left\|\mathbf{X}^T\mathbf{P}-\mathbf{Z}\right\|_F^2+\beta\left\|\mathbf{P}\right\|_{2,1}) \\ s.t. \quad & \mathbf{Z}^T\mathbf{Z}=\mathbf{I}, \qquad \mathbf{Z}\ge 0 \end{aligned} \tag{8}$$

2.5. Optimization Algorithm for CDDS

It is difficult to optimize the components $\mathbf{P}$ and $\mathbf{Z}$ simultaneously in Eq.(8). To solve the problem, we apply an alternating optimization scheme to update $\mathbf{P}$ and $\mathbf{Z}$ iteratively and alternatingly to find an optimal solution.

Given Z, Optimizing P: Fixing Z, the constraints in Eq.(8) are independent on $\mathbf{P}$ and the optimal $\mathbf{P}$ can be obtained by solving the following equivalent subproblem,

$$\min_{\mathbf{P}}\mathcal{J}(\mathbf{P})=\left\|\mathbf{X}^T\mathbf{P}-\mathbf{Z}\right\|_F^2+\beta\left\|\mathbf{P}\right\|_{2,1} \tag{9}$$

Setting $\frac{\partial\mathcal{J}(\mathbf{P})}{\partial\mathbf{P}}=0$, the updating rule of $\mathbf{P}$ are obtained as

$$\mathbf{P}=(\mathbf{X}\mathbf{X}^T+\beta\mathbf{D_P})^{-1}\mathbf{X}\mathbf{Z} \tag{10}$$

Algorithm 1 The CDDS Algorithm

Input: $\{\mathbf{X}, \mathbf{T}, \alpha, \beta, \mu, k, d\}$

Output: d most discriminative items

Construct the laplacian matrix $\mathbf{L_s}$ by $\mathbf{X}$, $\mathbf{T}$ and μ;
Construct $\mathbf{L_s}^+$ and $\mathbf{L_s}^-$
Initialize $\mathbf{D_P}$ as an identity matrix;
Initialize $\mathbf{Z}$ as $\mathbf{Z}^T\mathbf{Z} \approx \mathbf{I}$ with $\mathbf{Z} \geq 0$;
while not convergent **do**
  Update $\mathbf{P}$ by $\mathbf{P} = (\mathbf{X}\mathbf{X}^T + \beta\mathbf{D_P})^{-1}\mathbf{X}\mathbf{Z}$;
  Construct $\mathbf{Q} = \mathbf{X}^T\mathbf{P}$, $\mathbf{Q}^+$ and $\mathbf{Q}^-$;
  Set $\mathbf{\Lambda} = \alpha(\mathbf{Z}^T\mathbf{Q} - \mathbf{I}) - \mathbf{Z}^T\mathbf{L_s}\mathbf{Z}$, $\mathbf{\Lambda}^+$ and $\mathbf{\Lambda}^-$;
  Update $\mathbf{Z}$ by Eq.(15);
  Update $\mathbf{D_P}$ by $\mathbf{D_P}_{ii} = \frac{1}{2\|\mathbf{P}(i,:)\|_2}$;
end while
Sort items according to $\|\mathbf{P}(i,:)\|_2$ in descending order and select the top-d ranked ones;

where $\mathbf{D_P}$ is a diagonal matrix with $\mathbf{D_P}_{ii} = \frac{1}{2\|\mathbf{P}(i,:)\|_2}$.

Given P, Optimizing Z: Fixing $\mathbf{P}$, we can update $\mathbf{Z}$ by solving the following optimization subproblem ,

$$\begin{aligned} &\min_{\mathbf{Z}} \mathcal{J}(\mathbf{Z}) = Tr(\mathbf{Z}^T\mathbf{L_s}\mathbf{Z}) + \alpha\|\mathbf{Q} - \mathbf{Z}\|_F^2 \\ &s.t. \quad \mathbf{Z}^T\mathbf{Z} = \mathbf{I}, \qquad \mathbf{Z} \geq 0 \end{aligned} \tag{11}$$

where we set $\mathbf{Q} = \mathbf{X}^T\mathbf{P}$ and discard irrelevant term $\|\mathbf{P}\|_{2,1}$.

An optimization strategy is proposed to solve Eq.(11). The lagrangian function of Eq.(11) is

$$\begin{aligned} \mathcal{L}(\mathbf{Z}) =& Tr(\mathbf{Z}^T\mathbf{L_s}\mathbf{Z}) + \alpha Tr(\mathbf{Z}^T\mathbf{Z} - 2\mathbf{Q}^T\mathbf{Z}) \\ &+ Tr(\mathbf{\Lambda}(\mathbf{Z}^T\mathbf{Z} - \mathbf{I})) - Tr(\mathbf{\Phi}\mathbf{Z}^T) \end{aligned} \tag{12}$$

where $\mathbf{\Lambda}$ and $\mathbf{\Phi}$ are lagrangian multipliers.

Setting $\frac{\partial\mathcal{L}(\mathbf{Z})}{\partial\mathbf{Z}} = 0$, we can obtain

$$\frac{\partial\mathcal{L}(\mathbf{Z})}{\partial\mathbf{Z}} = 2\mathbf{L_s}\mathbf{Z} + 2\alpha(\mathbf{Z} - \mathbf{Q}) + 2\mathbf{Z}\mathbf{\Lambda} - \mathbf{\Phi} = 0 \tag{13}$$

Using the Karush-Kuhn-Tucker complementary condition for the nonnegativity of $\mathbf{Z}$, $\mathbf{\Phi}_{ij}\mathbf{Z}_{ij} = 0$, we get

$$(\mathbf{L_s}\mathbf{Z} + \alpha(\mathbf{Z} - \mathbf{Q}) + \mathbf{Z}\mathbf{\Lambda})_{ij}\mathbf{Z}_{ij} = 0 \tag{14}$$

which leads to the following updating rule according to [12],

$$\mathbf{Z}_{ij} \leftarrow \mathbf{Z}_{ij}\sqrt{\frac{(\mathbf{L_s^-}\mathbf{Z} + \alpha\mathbf{Q}^+ + \mathbf{Z}\mathbf{\Lambda}^-)_{ij}}{(\mathbf{L_s^+}\mathbf{Z} + \alpha(\mathbf{Z} + \mathbf{Q}^-) + \mathbf{Z}\mathbf{\Lambda}^+)_{ij}}} \tag{15}$$

Since $\mathbf{L_s}$, $\mathbf{Q}$ and $\mathbf{\Lambda}$ may take any signs, we decompose them as $\mathbf{A} = \mathbf{A}^+ - \mathbf{A}^-$, where $\mathbf{A}_{ij}^+ = (|\mathbf{A}_{ij}| + \mathbf{A}_{ij})/2$ and $\mathbf{A}_{ij}^- = (|\mathbf{A}_{ij}| - \mathbf{A}_{ij})/2$. Inspired by [13], we approximate $\mathbf{\Lambda}$ by ignoring the non-negative constraint lagrangian multiplier $\mathbf{\Phi}$ in Eq.(13), which leads to $\mathbf{\Lambda} = \alpha(\mathbf{Z}^T\mathbf{Q} - \mathbf{I}) - \mathbf{Z}^T\mathbf{L_s}\mathbf{Z}$.

The entire optimization algorithm is presented in Algorithm 1 and the convergence analysis is presented in the supplemental material.

Table 1: Descriptions of the Datasets

Dataset	# User	# Item	# Ratings	# Trusts
Flixster	5396	3234	563566	56930
Douban	6034	3427	299812	52116

3. COLD-START RECOMMENDATION

After the discriminative items are selected from the candidate pool $\mathcal{V}_p$, an interview process is performed for eliciting preference. A new user needs to response to a series of queries about what his ratings to these selected items are. Then a personalized ranking scores of the whole items in $\mathcal{V}$ is predicted by the following ranking estimation approach, which can provide high quality cold-start recommendation.

A new user can be denoted as u_{n+1}. Once the interview is conducted, a column vector $\tilde{\mathbf{x}}_{n+1}$ is generated whose elements are rating responses of u_{n+1} to queries on selected items. For users in $\mathcal{U}$, a rating matrix $\widetilde{\mathbf{X}}$ can be constructed by $\widetilde{\mathbf{X}} = \mathbf{R}(\mathcal{V}_s, :) \in \mathbb{R}^{d\times n}$ each column of whose is corresponding to ratings of one user to selected items. To estimate personalized ranking scores, a coefficient matrix $\mathbf{W}$ is obtained through $\widetilde{\mathbf{X}}$ to reconstruct the original observed rating matrix $\mathbf{R}$. Thus we introduce the ridge regression to learn coefficient matrix $\mathbf{W}^* \in \mathbb{R}^{d\times m}$ via

$$\mathbf{W}^* = \arg\min_{\mathbf{W}} \frac{1}{2}\|\mathbf{R} - \mathbf{W}^T\widetilde{\mathbf{X}}\|_F^2 + \frac{\nu}{2}\|\mathbf{W}\|_F^2 \tag{16}$$

The optimum is obtained by $\mathbf{W}^* = (\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^T + \nu\mathbf{I})^{-1}\widetilde{\mathbf{X}}\mathbf{R}^T$, where $\mathbf{I} \in \mathbb{R}^{d\times d}$ is an identity matrix. The personalized ranking scores can be computed with $\mathbf{W}^{*T}\tilde{\mathbf{x}}_{n+1}$ for new user u_{n+1}. We should notice that $\mathbf{W}^*$ is pre-computed offline, which means there is no need to retrain the entire model for cold-start recommendation. Thus our model is consistent with online-updating principle of real-world system.

4. EXPERIMENTS

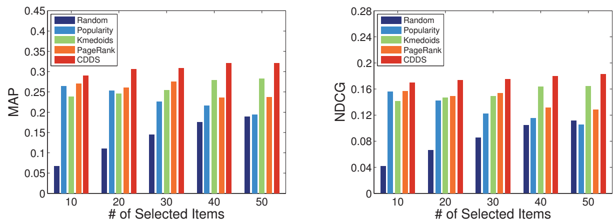
4.1. Datasets and Experimental Setup

To evaluate CDDS on cold-start recommendation, we perform experiments on two real-world datasets: Flixster [1] and Douban[2], with all ratings scaled from 1 to 5. The social network in Douban is directed while the network is undirected in Flixster which can be viewed as a bidirectional directed network. The detailed information of the two datasets is shown in Table 1.

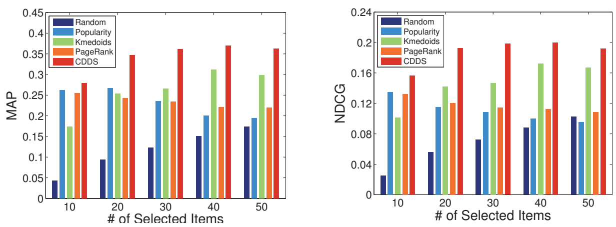
In our experiments, we randomly select 20% users from the user set to simulate cold users and the rest 80% users for training models. In order to test the cold-start recommendation results, we apply an 80-20 split for each cold user, where

[1] http://www.cs.sfu.ca/ sja25/personal/datasets/

[2] We crawled http://www.douban.com/ for movie ratings and users' trusts.



(a) MAP on Flixster (b) NDCG on Flixster



(c) MAP on Douban (d) NDCG on Douban

Fig. 2: Cold-start Recommendation Performance with Different Seed Item Selection Approaches on Flixster and Douban

80% ratings are randomly picked as responses to the queries during interview process and the rest 20% ratings constitute the test set. We set the parameter $\nu = 0.1$ via cross-validation. In addition, the size of candidate pool $\mathcal{V}_p$ and user nearest neighbor set are determined as 500 and 15 respectively. With above experimental settings, We repeat the entire training and test process 5 times to get better statistics of the performance.

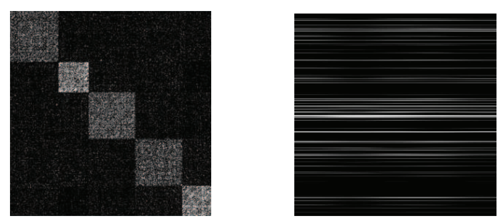
4.2. Baselines and Metrics

To demonstrate the effectiveness of our CDDS method, several traditional item selection methods are used to make comparisons with it. (1)Random: Randomly select d items as queries. (2)Popularity: Select top-d rated items as queries. (3)Kmedoids [9]: Select representative items of d item clusters. (4)PageRank [7]: Utilize PageRank to select top-d most influential items based on undirected item network, where we miss temporal information and construct the network by rating similarity.

When performing these preference elicitation methods, we can obtain ranking scores on entire items for each user. To measure the performance of our model, we adopt two widely used evaluation metrics in top-N recommendation: MAP(Mean Average Precision) and NDCG(Normalized Discounted Cumulative Gain). A higher MAP and NDCG corresponds to a better cold-start recommendation performance.

4.3. Results and Analysis

Firstly, we compare performance of our proposed CDDS with other methods for cold-start recommendation. Considering the real-world scenarios where new users are willing to response to fewer queries, we conduct our experiments on different interview query numbers $d = \{10, 20, 30, 40, 50\}$. The experimental results on Flixster and Douban datasets are shown in Fig.2. The parameters for CDDS are determined by cross-validation. The resulting parameters are: $\{\alpha = 100, \beta = 1, \mu = 0.7, k = 4\}$ for Flixster and $\{\alpha = 1, \beta = 10, \mu = 0.4, k = 5\}$ for Douban. We observe that our proposed CDDS framework consistently outperforms all the other baselines. Note that Random strategy has a worse performance, which shows the necessity of designing an appropriate selection model. A trend of Popularity and PageRank methods can be observed that their curves decrease at larger d, which can be explained by information overlap and noise among seed items selected by them [14]. Moreover, since Kmedoids method can select representative items, information overlap could be reduced and it performs well at larger d. However, our method can always keep the best performance, indicating that the community-discovering guided selection process can largely avoid information redundancy by filtering out noisy items.

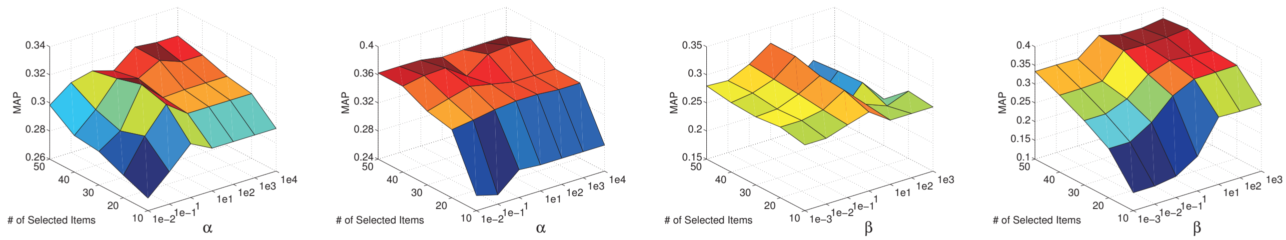


(a) Community Pattern (b) Learned $\mathbf{P}$

Fig. 3: Visualization of Community Pattern and Learned $\mathbf{P}$

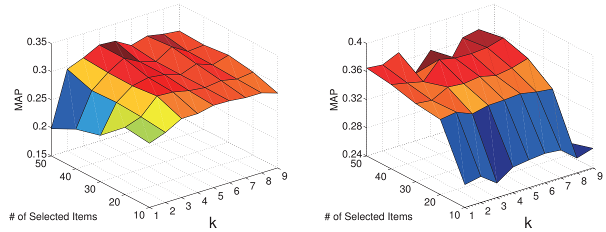
Next, we present the visualization of community discovering and seed item selection on Douban dataset. Fig. 3a shows a block structure of similarity matrix $\mathbf{S}$ after we reorder the index of users according to the learned community indicator matrix $\mathbf{Z}$, where we can see 5 communities clearly. This demonstrates our approach can effectively discover the communities. In Fig. 3b, brighter lines represent the rows with higher norm values in $\mathbf{P}$, which is corresponding to most discriminative items. It indicates that CDDS can perform an effective item selection with a community discovering process.

Finally, we demonstrate the parameters studies for three important parameters: $\ell_{2,1}$-norm regularizer weight α, graph regularizer weight β and number of communities k in our model. We analyze α, β, k respectively by fixing the other two parameters with variation of numbers of selected items. To save the space, we only report the analysis based on MAP. In Fig. 4a,4b, we can observe that the performance will peak at 100 for Flixster and at 1 for Douban. In Fig. 4c,4d, the performance will peak at 1 for Flixster and at 10 for Douban. We can also get the best k at 4 and 5 for Flixster and Douban respectively in Fig. 5.



(a) α on Flixster ($\beta = 1, k = 4$) (b) α on Douban ($\beta = 10, k = 5$) (c) β on Flixster ($\alpha = 100, k = 4$) (d) β on Douban ($\alpha = 1, k = 5$)

Fig. 4: Parameter Study of α, β w.r.t Numbers of Selected Items



(a) k on Flixster ($\alpha = 100, \beta = 1$) (b) k on Douban ($\alpha = 1, \beta = 10$)

Fig. 5: Parameter Study of k w.r.t Numbers of Selected Items

5. CONCLUSIONS

In this paper, we propose a novel query selection framework CDDS for preference elicitation. Guided by a community discovering process, our model tends to select most discriminative items which can best elicit user tastes via interview process. Experimental results on real-world datasets Flixster and Douban show our proposed framework performs a more accurate cold-start recommendation than baseline methods.

6. ACKNOWLEDGEMENTS

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

7. REFERENCES

[1] Andriy Mnih and Ruslan Salakhutdinov, "Probabilistic matrix factorization," in *NIPS*, 2007.

[2] Yehuda Koren, "Factorization meets the neighborhood: a multifaceted collaborative filtering model," in *KDD*, 2008.

[3] Zeno Gantner, Lucas Drumond, Christoph Freudenthaler, Steffen Rendle, and Lars Schmidt-Thieme, "Learning attribute-to-feature mappings for cold-start recommendations," in *ICDM*, 2010.

[4] Seung-Taek Park, David Pennock, Omid Madani, Nathan Good, and Dennis DeCoste, "Naïve filterbots for robust cold-start recommendations," in *KDD*, 2006.

[5] Al Mamunur Rashid, Istvan Albert, Dan Cosley, Shyong K Lam, Sean M McNee, Joseph A Konstan, and John Riedl, "Getting to know you: learning new user preferences in recommender systems," in *IUI*, 2002.

[6] Abhay S Harpale and Yiming Yang, "Personalized active learning for collaborative filtering," in *SIGIR*, 2008.

[7] Qi Liu, Biao Xiang, Enhong Chen, Yong Ge, Hui Xiong, Tengfei Bao, and Yi Zheng, "Influential seed items recommendation," in *RecSys*, 2012.

[8] Nadav Golbandi, Yehuda Koren, and Ronny Lempel, "Adaptive bootstrapping of recommender systems using decision trees," in *ICDM*, 2011.

[9] Xi Zhang, Jian Cheng, Ting Yuan, Biao Niu, and Hanqing Lu, "Semi-supervised discriminative preference elicitation for cold-start recommendation," in *CIKM*, 2013.

[10] Ulrike Von Luxburg, "A tutorial on spectral clustering," *Statistics and computing*, vol. 17, no. 4, pp. 395–416, 2007.

[11] Yi Yang, Heng Tao Shen, Feiping Nie, Rongrong Ji, and Xiaofang Zhou, "Nonnegative spectral clustering with discriminative regularization," in *AAAI*, 2011.

[12] Chris HQ Ding, Tao Li, and Michael I Jordan, "Convex and semi-nonnegative matrix factorizations," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 32, no. 1, pp. 45–55, 2010.

[13] Jiliang Tang, Xia Hu, Huiji Gao, and Huan Liu, "Unsupervised feature selection for multi-view data in social media," in *SDM*, 2013.

[14] Kai Yu, Jinbo Bi, and Volker Tresp, "Active learning via transductive experimental design," in *ICML*, 2006.