Cross-Domain Recommendation for Mapping Sentiment Review Pattern

Conference Paper · October 2019

4 authors, including:

Yang Xu
Shandong Normal University

Yupeng Hu
Shandong University

Cross-Domain Recommendation for Mapping Sentiment Review Pattern

Yang Xu, Zhaohui Peng, Yupeng Hu, Xiaoguang Hong(✉), and Wenjing Fu

School of Computer Science and Technology, Shandong University, Jinan, People's Republic of China
{xuyang0211, pzh, huyupeng, hxg}@sdu.edu.cn, fuwenjing1214@mail.sdu.edu.cn

Abstract. Cross-domain algorithms which aim to transfer knowledge available in the source domains to the target domain are gradually becoming more attractive as an effective approach to help improving quality of recommendations and to alleviate the problems of cold-start and data sparsity in recommendation systems. However, existing works on cross-domain algorithm mostly consider ratings, tags and the text information like reviews, cannot use the sentiments implicated in the reviews efficiently. In this paper, we propose a Sentiment Review Pattern Mapping framework for cross-domain recommendation, called SRPM. The proposed SRPM framework can model the semantic orientation of the reviews of users, and transfer sentiment review pattern of users by using a multi-layer perceptron to capture the nonlinear mapping function across domains. We evaluate and compare our framework on a set of Amazon datasets. Extensive experiments on each cross-domain recommendation scenarios are conducted to prove the high accuracy of our proposed SRPM framework.

Keywords: Cross-domain recommendation · Sentiment review pattern
Pattern mapping

1 Introduction

Cross-domain recommendation systems are gradually becoming more attractive as a practical approach to improve quality of recommendations and to alleviate cold-start problem, especially in small and sparse datasets. These algorithms mine knowledge on users and items in a source domain to improve the quality of the recommendations in a target domain. They can also provide joint recommendations for items belonging to different domains by the linking information among these domains [1]. Most existing works about cross domain recommendation tend to aggregate knowledge from different domains from the perspective of explicitly specified common information [2–4] or transferring latent features [5, 8, 9, 13]. However, the aggregated knowledge merely based on ratings, tags, or the text information like reviews, ignores the sentiments implicated in the reviews. After watching a popular film, using a novel electronic product or playing a video game, users often rate them and submit reviews to share their feelings, which could convey fairly rich sentiment information.

W. Liu et al. (Eds.): KSEM 2018, LNAI 11061, pp. 324–336, 2018.
https://doi.org/10.1007/978-3-319-99365-2_29

For all we know, existing cross-domain recommendation algorithms which utilize user reviews didn't take full advantage of the sentiment information of these reviews. They implement knowledge transfer by mixing positive and negative reviews together, which will weaken and even lose some sentiment information of the users, especially the negative sentiment. For instance, a user may deeply care about the plot of a novel, and he made positive comments on the plots of some novels in the domain of electronic book (source domain) while made negative comments on the plots of some other novels. If we transfer the knowledge gained from user reviews from the source domain to the target domain without distinguishing the sentiment polarity of these reviews, some latent factors such as "plot", "positive sentiment" and "negative sentiment" of the reviews will be mixed up as "users' feature" to be transferred to the target domain. In the domain of movie (target domain), a movie with poor plots, namely the movie whose latent factors "plot" and "negative sentiment" take higher weight, will produce a match with the users' feature transferred to the target domain. Nevertheless, the user may not be fond of this movie.

To address this problem, in this paper, we propose a new cross-domain recommendation framework called SRPM. Under SRPM, we can effectively identify the sentiment orientation of user reviews and adapt topic modeling approach to deduce the sentiment review pattern (SRP) from user reviews. To achieve the goal of transferring knowledge, we propose an MLP based mapping method to transfer sentiment information of users from source domain to target domain. Then we can get an affine SRP for a cold-start user in the target domain and predict the cold-start user's rating of items in the target domain. Through transferring SRP of users, the SRPM method gets a superior performance in cross-domain recommendation.

To summarize, the major contributions of this paper are as follows:

– We consider sentiment information in the cross-domain recommendation task and propose a novel cross-domain recommendation framework named SRPM. SRPM can be used to transfer sentiment review pattern from source domain to target domain and make recommendation for cold-start users in target domain.
– In SRPM, we design a sentiment information extracting approach, and propose the modeling and mapping method of sentiment review pattern.
– We systematically compare the proposed SRPM approach with other algorithms on the Amazon dataset. The results confirm that our new method substantially improves the performance of cross-domain recommendation.

SRPM is applicable for User-Item overlap scenarios in which users or items are found to be in common in both domains. In this paper, we introduce SRPM under the user overlap scenario.

The rest of this paper is organized as follows. Section 2 presents some notations and the problem formulation. Section 3 introduce the modeling method of SRP and Sect. 4 details the mapping method of SRP and the cross-domain recommendation approach. Experiments and discussion are given in Sect. 5. Section 6 reviews the related works on cross-domain recommendation and sentiment analysis in recommendation system. Conclusions are drawn in Sect. 7.

2 Preliminaries

In this section, we first introduce some notations in SRPM cross-domain recommendation framework and then present the SRPM framework to solve the cold-start recommendation problem.

2.1 Notations

Objects to be recommended in the cross-domain recommendation system are referred to as *items*. Let $\mathcal{U} = \{u_1, u_2, \ldots, u_{|\mathcal{U}|}\}$ denote the set of common users in both domains and $\mathcal{I}_S = \{i_1, i_2, \ldots, i_{|\mathcal{I}_S|}\}, \mathcal{I}_T = \{\imath_1, \imath_2, \ldots, \imath_{|\mathcal{I}_T|}\}$ are the sets of items (e.g. movies, books, or electronics) in source domain and in target domain respectively. The user review dataset is represented as $SR_U = \left\{r_{u_1}, r_{u_2}, \ldots, r_{u_{|\mathcal{U}|}}\right\}$ in source domain and TR_U in target domain, where r_{u_i} is all of reviews of user u_i in the corresponding domain. Similarly, we let $TR_I = \left\{r_{\imath_1}, r_{\imath_2}, \ldots, r_{\imath_{|\mathcal{I}_T|}}\right\}$ denote the item review dataset in target domain, where $r_{\imath_j}$ is all of the reviews which item $\imath_j$ acquired in target domain.

In the SRPM framework, the sentiment analysis algorithm is employed on the review datasets mentioned above to divide them into corresponding positive review datasets (e.g. $SR_U^{pos}, TR_U^{pos}\ and\ TR_I^{pos}$) and negative review datasets (e.g. $SR_U^{neg}, TR_U^{neg}\ and\ TR_I^{neg}$). $S_U^{pos} = \left\{\theta_{S,u_1}^{pos}, \theta_{S,u_2}^{pos}, \ldots, \theta_{S,u_{|U|}}^{pos}\right\}$ represents the positive review pattern matrix in latent space of all users in source domain, where θ_{S,u_i}^{pos} is the positive review pattern of user u_i in source domain, similarly to $S_U^{neg}, T_U^{pos}\ and\ T_U^{neg}$. In addition, $T_I^{pos} = \left\{\theta_{T,\imath_1}^{pos}, \theta_{T,\imath_2}^{pos}, \ldots, \theta_{T,\imath_{|I|}}^{pos}\right\}$ denotes the positive review topic distribution matrix in latent space of all items in target domain, where $\theta_{T,\imath_j}^{pos}$ is the positive review topic distribution of item $\imath_j$ in target domain, similarly to T_I^{neg}. For a cold-start user u' in target domain, $\theta_{S,u'}^{pos}$ ($\theta_{S,u'}^{neg}$) denotes the positive (negative) review pattern of user u' in source domain, and $\widehat{\theta}_{T,u'}^{pos}$ ($\widehat{\theta}_{T,u'}^{neg}$) represents the affine positive (negative) review pattern of user u' in target domain.

2.2 Problem Formulation

Given two domains which share the same users U. Users appearing in only one domain can be regarded as the cold-start users U' in the other domain. Without loss of generality, one domain is referred to as the source domain and the other as the target domain. The most common cross-domain recommendation approaches focus on transferring information based on ratings, tags and reviews from source domain to target domain, without accounting for any emotional information implicated the reviews.

We are tackling cross-domain recommendation task for cold-start users by modeling the Sentiment Review Pattern (SRP) of users and transferring them from source domain to target domain.

To achieve this purpose, we propose a cross-domain recommendation framework called SRPM. This framework contains three major steps, i.e., sentiment review pattern modeling, sentiment review pattern mapping and cross-domain recommendation, as illustrated in Fig. 1.

In the first step, we apply SO-CAL [7] to analyze the Semantic Orientation (SO) of each sentence of user reviews in both domains. Then, the original review datasets of both domains are divided into corresponding positive review datasets and negative review datasets respectively. Next, we employ Smoothed Latent Dirichlet Allocation (SLDA) on the sentiment tagged datasets to find the sentiment review pattern of users. In the second step, we model the cross-domain relationships of users through a mapping function based on Multi-Layer Perceptron (MLP) [6]. We assume that there is an underlying mapping relationship between the user's SRPs of the source and target domains, and further use a mapping function to capture this relationship. Finally, in the third step, we make recommendation for cold-start user in the target domain. We can get an affine SRP for cold-start user in the target domain, with the SRP learned for him/her in the source domain and the MLP-based mapping functions between the source and target domain. In the rest of this paper, we will introduce each step of SRPM in details.

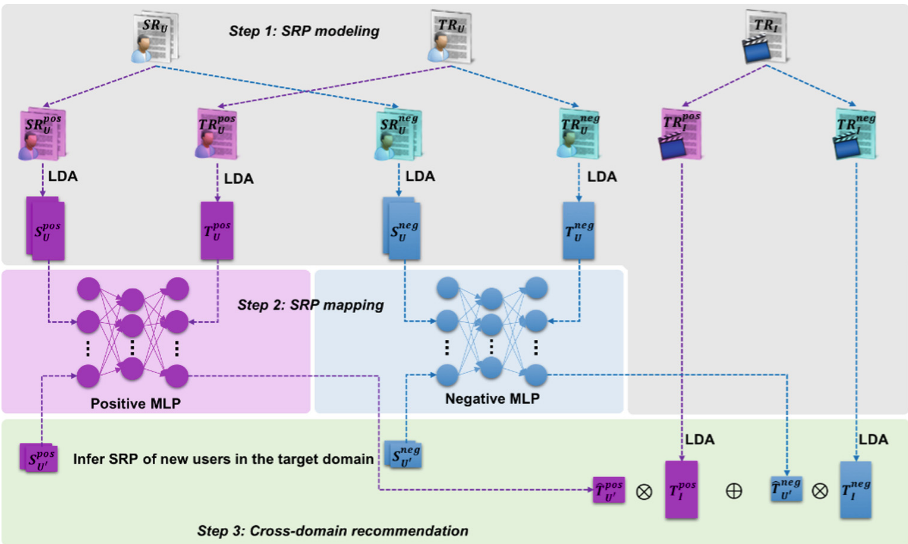

Fig. 1. Overview of the SRPM cross-domain recommendation framework

3 Sentiment Review Pattern Modeling

As discussion in the previous section, in order to transfer sentiment review pattern in the source domains to the target domain, the first phase of SRPM is to model the sentiment review pattern of common users in both domains. The key challenge is how to extract the user's focus on the item and the positive or negative emotions expressing in the user reviews. To address this challenge, we propose a sentiment review pattern modeling method based on sentence-level sentiment analysis approach and smoothed LDA.

3.1 Sentiment Analysis

The sentiment analysis problem in SRPM can be formulated as follows: Given a set of reviews R, a sentiment classification algorithm classifies each sentence of a piece of review $r \in R$ into one of the two classes, positive R^{pos} and negative R^{neg}. For this purpose, we apply the sentiment analysis algorithm SO-CAL [7] to analyze the semantic orientation of each sentence of user reviews. Since the sentiment analysis algorithm is not the focus of this paper, we refer the readers to the related literature such as [7, 20] for details.

We employ SO-CAL on original review sets SR_U, TR_U, TR_I to divide them into positive review subsets $SR_U^{pos}, TR_U^{pos}, TR_I^{pos}$ and negative review subsets SR_U^{neg}, TR_U^{neg}, TR_I^{neg} on the sentence-level respectively.

3.2 Sentiment Review Pattern

Sentiment review pattern indicates the user's focus on the item and the positive or negative emotions expressed in a sentence. In this paper, we use the smoothed LDA topic model [10] to extract review pattern. In our formulation, "document" is a collection of positive (negative) reviews of a user u or an item i in a certain domain, which represented as $r_u^{pos}\left(r_u^{neg}\right)$ for the user u and $r_i^{pos}(r_i^{neg})$ for the item i. In the higher level, "corpus" is a collection of positive (negative) "documents" of the user set U or the item set I in that domain.

Here, we employ SLDA on the positive (negative) corpus of U in source domain SR_U^{pos} (SR_U^{neg}) to find the positive (negative) topic distribution $\theta_{S_u}^{pos}$ $(\theta_{S_u}^{neg})$ of each user u in source domain and the per-topic word distribution $\beta_{S_{U,k}}^{pos}$ $(\beta_{S_{U,k}}^{neg})$.

Similarly, we employ SLDA on $TR_U^{pos}, TR_U^{neg}, TR_I^{pos}, TR_I^{neg}$ to find topic distributions $\theta_{T_u}^{pos}, \theta_{T_u}^{neg}, \theta_{T_i}^{pos}, \theta_{T_i}^{neg}$ and per-topic word distributions $\beta_{T_{U,k}}^{pos}, \beta_{T_{U,k}}^{neg}, \beta_{T_{I,k}}^{pos}, \beta_{T_{I,k}}^{neg}$ respectively. In this paper, the topic distribution of user u's positive (negative) reviews is called as user u's Positive Review Pattern PRP_u (or Negative Review Pattern NRP_u). Then, the sentiment review pattern of user u is denoted as $SRP_u = (PRP_u, NRP_u)$.

4 Sentiment Review Pattern Mapping and Cross-Domain Recommendation

4.1 Sentiment Review Pattern Mapping

In this paper, we utilize an MLP-based method to tackle the SRP mapping problem, as shown in Fig. 1. To avoid mutual interference between positive and negative emotion factors during the process of knowledge transfer, two MLP models, the positive MLP model and the negative MLP model, were employed to map PRP and NRP from source domain to target domain respectively. Next, we will introduce the proposed mapping algorithm under *PRP* mapping scenario, and the mapping algorithm under *NRP* mapping scenario is similar.

In our proposed mapping algorithm, only the common users with sufficient review data are used to learn the mapping function in order to guarantee its robustness to noise caused by review data sparsity and imbalance in both domains. We use entropy and statistical method to measure the cross-domain degree of common users. Formally, the cross-domain degree is defined as follows:

$$\begin{aligned} c(u) &= \left(-p_{u,s}\log_2 p_{u,s} - p_{u,t}\log_2 p_{u,t}\right)\frac{\mathcal{N}(u,s)+\mathcal{N}(u,t)}{\sum_{u_i\in U_c}\mathcal{N}(u_i,s)+\mathcal{N}(u_i,t)} \\ where \quad p_{u,s} &= \frac{\mathcal{N}(u,s)}{\mathcal{N}(u,s)+\mathcal{N}(u,t)} \;, p_{u,t} = \frac{\mathcal{N}(u,t)}{\mathcal{N}(u,s)+\mathcal{N}(u,t)} \end{aligned} \tag{1}$$

$\mathcal{N}(u,s)$ is the number of reviews in source domain of common user u, and $\mathcal{N}(u,t)$ is that in target domain. U_c denotes the set of common users between both domains. The common users with $c(u) > threshold\,\gamma$ are selected to learn the mapping function.

Let $\theta^S = \left\{\theta_1^S, \theta_2^S, \ldots, \theta_N^S\right\}$ denotes the set of *PRP* s in the source domain, and $\theta^T = \left\{\theta_1^T, \theta_2^T, \ldots, \theta_N^T\right\}$ represents the set of *PRP* s in the target domain. N is the number of common users in both domains. Under the MLP model setting, we formulate the *PRP* mapping problem as: Given N training instance $\left(\theta_i^S, \theta_i^T\right)$, $\theta_i^S, \theta_i^T \in R^M$, $(i = 1, 2, \ldots, N)$, where $\theta_i^S = \left(\theta_{i1}^S, \theta_{i2}^S, \ldots, \theta_{iM}^S\right)$ is the *PRP* of common user u_i in the source domain and $\theta_i^T = \left(\theta_{i1}^T, \theta_{i2}^T, \ldots, \theta_{iM}^T\right)$ is that in the target domain, our task is to learn an MLP mapping function to map the *PRP* from the source domain to the target domain.

In a feedforward MLP model, the output o_{ik} is formulated as

$$y_{ik} = \sum_{j=1}^{L} c_{jk}a_j, \quad o_{ik} = g(y_{ik}) \tag{2}$$

where c_{jk} represents the weight of the j'th input of the output layer neuron k and L is the number of hidden neurons in each hidden layer. $g(y)$ is the activation function of the output layer, which is set to be the softmax function in this study. a_j denotes the j'th hidden neuron activation of lower hidden layer, which can be defined as

$$y_j = \sum_{p=1}^{P} w_{pj}a_p, \quad a_j = f\left(y_j\right) \tag{3}$$

where w_{pj} is the weight of the p'th input of the hidden layer neuron j (the hidden bias can be included in the input weights) and a_p is the input θ_{ip}^S or the p'th hidden neuron activation of the lower hidden layer. P represents the number of inputs or neurons in the lower layer. $f(y)$ is the hidden layer activation function, which is set to be the ReLU function in this study.

Considering that input and output of MLP model in this study are all topic distributions, the error between θ_i^T and $o_i = (o_{i1}, o_{i2}, \ldots, o_{iM})$ is measured by KL divergence:

$$E = \sum_{i=1}^{N} \sum_{k=1}^{M} o_{ik} \log \frac{o_{ik}}{\theta_{ik}^T} \tag{4}$$

To obtain the MLP mapping function, we utilize stochastic gradient descent to learn the weights. We refresh the weights of the MLP by looping through the training instances. The back-propagation algorithm is adopted to calculate the gradients of the weights, thus we can get the positive MLP mapping function $f_{pmlp}(\cdot;\theta_p)$, where θ_p is its weights set. Similarly, we can learn the negative MLP mapping function $f_{nmlp}(\cdot;\theta_n)$ by employing the above learning algorithm.

4.2 Cross-Domain Recommendation

Given a cold-start user in the target domain, we do not have sufficient information to estimate his/her preference features to make recommendation directly in the target domain. However, we can get the affine SRP for him/her in the target domain, with the SRP learned in the source domain and the MLP mapping functions from the source domain to the target domain. In this section, we will introduce how to predict the cold-start user's ratings on the specific items in the target domain.

Given a cold-start user u' in the target domain, we can extract user u''s positive review pattern $\theta_{S,u'}^{pos}$ and negative review pattern $\theta_{S,u'}^{neg}$ from S_U^{pos} and S_U^{neg} in the source domain respectively. Then the affine positive review pattern $\widehat{\theta}_{T,u'}^{pos}$ and the affine negative review pattern $\widehat{\theta}_{T,u'}^{neg}$ can be obtained by the following equations:

$$\widehat{\theta}_{T,u'}^{pos} = f_{pmlp}\left(\theta_{S,u'}^{pos};\theta_p\right), \quad \widehat{\theta}_{T,u'}^{neg} = f_{nmlp}\left(\theta_{S,u'}^{neg};\theta_n\right) \tag{5}$$

Next, the similarity of each pair of topics between the corresponding emotional review dataset are defined as:

$$SIM^{pos} = \left\{sim_{i,j}^{pos}\right\}, where\, sim_{i,j}^{pos} = \cos\left(\beta_{T_{U,i}}^{pos}, \beta_{T_{I,j}}^{pos}\right), \quad i,j = 1,2,\ldots,M \tag{6}$$

$$SIM^{neg} = \left\{sim_{i,j}^{neg}\right\}, where\, sim_{i,j}^{neg} = \cos\left(\beta_{T_{U,i}}^{neg}, \beta_{T_{I,j}}^{neg}\right), \quad i,j = 1,2,\ldots,M \tag{7}$$

Then, the predicted emotional rating between cold-start user u' and item $\imath_j \in \mathcal{J}_T$ in the target domain is calculated as:

$$E(u') = \left\{e_{u',j}\right\} = \widehat{\theta}_{T,u'}^{pos} \cdot SIM^{pos} \cdot T_I^{pos\mathrm{T}} - \widehat{\theta}_{T,u'}^{neg} \cdot SIM^{neg} \cdot T_I^{neg\mathrm{T}} \tag{8}$$

Finally, we combine a baseline estimate function and the predicted emotional rating to predict the overall rating between u' and $\imath_j$, which is formulated as:

$$S(u', \imath_j) = b_{\mathcal{I}_T} + b_{u'} + b_{\imath_j} + e_{u',j} \tag{9}$$

where $b_{\mathcal{J}_T}$ denoted the overall average ratings of all items in the target domain. The parameter $b_{u'}$ is the user rating bias in the source domain and $b_{\imath_j}$ is the item rating bias in the target domain, which indicate the observed deviations of user u' and item $\imath_j$ from the average.

5 Experiments

We have conducted a set of experiments to examine the performance of our cross-domain recommendation method compared with the baselines. In this section, we first introduce the experimental settings, and then analyze the evaluation results.

5.1 Experimental Settings

Data Description. We employ Amazon cross-domain dataset [11] in our experiment. This dataset contains product reviews and star ratings with 5-star scale from Amazon, including 142.8 million reviews spanning May 1996 – July 2014. We select the top three domains with the most widely used in previous studies to employ in our cross-domain experiment. In our experiments, source domains are selected by calculating their relevance to the target domain. The relevance between two domains is defined as the ratio of the number of overlapped users' reviews in the target domain to the number of reviews in the target domain. The global statistics of these domains used in our experiments are shown in Table 1.

Table 1. Characteristics of datasets

Dataset	Books	Electronics	Movies & TV
# of Users	603,668	192,403	123,960
# of Items	367,982	63,001	50,052
# of Reviews	8,898,041	1,689,188	1,697,533
Density	0.004%	0.014%	0.027%

Experiment Setup. The domains in the Amazon dataset only have user overlaps. Thus, we evaluate the validity and efficiency of SRPM on the cross-domain recommendation task under the user overlap scenario. We randomly remove all the rating information of a certain proportion of users in the target domain and take them as cross-domain cold-start users for making recommendation. For the sake of stringency of the experiments, we set different proportions of cold-start users, namely, $\phi = 20\%$, 50% and 70%. Moreover, we repeatedly sample users for 10 times to generate different sets to balance the effect of different sets of cold-start users on the final recommendation results. We report the average results and standard deviations over these 10 different sets. Dimension of latent factor used in the compared method and the number of topics in LDA are set as $M = 20$, 50 and 100. For the mapping function, we set the structure of MLP as three hidden layers with $2M$ nodes in each hidden layer.

Compared Methods. We examine the performance of the proposed SRPM framework by comparing it with the following baseline methods:

– MF: This is the single-domain matrix factorization algorithm proposed in [12]. Comparing MF with the cross-domain methods will show us whether adding the extra information from source domains will increase recommendation accuracy.
– AVG: It predicts ratings by the following equation: $r_{ui} = b_T + b_u + b_i$ where b_T is the overall average ratings of all items in the target domain, b_u denotes the user rating bias in the source domain and b_i represents item bias in the target domain.
– CMF: This is a cross-domain recommendation method proposed in [14]. In CMF, the latent factors of users are shared between source domain and target domain.
– MF-MLP: This is a cross-domain recommendation framework based on MF and MLP, which is proposed by [15]. In our experiments, for MF-MLP, the structure of the MLP is set as one-hidden layer, and the number of nodes in the hidden layer is set as $2M$.

Evaluation Metric. We adopt the metrics of Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) to evaluate our method. They are defined as:

$$RMSE = \sqrt{\sum_{r_{ui} \in \mathcal{O}_{test}} \frac{(\hat{r}_{ui} - r_{ui})^2}{|\mathcal{O}_{test}|}}, \quad MAE = \frac{1}{|\mathcal{O}_{test}|} \sum_{r_{ui} \in \mathcal{O}_{test}} |\hat{r}_{ui} - r_{ui}| \tag{10}$$

where $\mathcal{O}_{test}$ is the set of test ratings, r_{ui} denotes an observed rating in the test set, and $\hat{r}_{ui}$ represents the predictive value of r_{ui}. $|\mathcal{O}_{test}|$ is the number of test ratings.

Table 2. Recommendation performance in terms of MAE on the “Books-Movies&TV”

	ϕ	MF	CMF	MF_MLP	AVE	SRPM
K = 20	20%	0.8783	0.8472	0.8421	0.8371	0.8016
	50%	0.8809	0.8620	0.8619	0.8810	0.8442
	70%	0.8824	0.8649	0.8638	0.9243	0.8671
K = 50	20%	0.9047	0.8902	0.8896	0.8371	0.8013
	50%	0.9372	0.9092	0.9008	0.8810	0.8388
	70%	0.9548	0.9329	0.9017	0.9243	0.8512
K = 1000	20%	1.0241	0.9088	0.9008	0.8371	0.8024
	50%	1.0764	0.9267	0.9322	0.8810	0.8429
	70%	1.1453	0.9702	0.9631	0.9243	0.8698

Table 3. Recommendation performance in terms of RMSE on the "Books-Movies&TV"

	ϕ	MF	CMF	MF_MLP	AVE	SRPM
K = 20	20%	1.3417	1.2035	1.1739	1.2280	1.1554
	50%	1.3443	1.2356	1.2090	1.2595	1.1890
	70%	1.3481	1.2360	1.2094	1.2835	1.2027
K = 50	20%	1.3698	1.2712	1.2371	1.2280	1.1223
	50%	1.4032	1.3488	1.2530	1.2595	1.1645
	70%	1.4256	1.3912	1.2544	1.2835	1.1982
K = 1000	20%	1.5069	1.4709	1.2791	1.2280	1.1776
	50%	1.5634	1.5106	1.2850	1.2595	1.1964
	70%	1.6310	1.5561	1.3282	1.2835	1.2035

Table 4. Recommendation performance in terms of MAE on the "Electronics-Movies&TV"

	ϕ	MF	CMF	MF_MLP	AVE	SRPM
K = 20	20%	0.9152	0.7091	0.8746	0.8778	0.8494
	50%	0.9175	0.7429	0.9297	0.9369	0.8729
	70%	0.9234	0.7633	0.9416	0.9461	0.9040
K = 50	20%	1.0651	0.8092	0.9699	0.8778	0.8534
	50%	1.1375	0.8495	0.9765	0.9369	0.8922
	70%	1.1958	0.9030	0.9544	0.9461	0.9270
K = 100	20%	1.2751	0.9022	1.0632	0.8778	0.8698
	50%	1.3622	0.9462	1.0316	0.9369	0.8938
	70%	1.4387	0.9883	1.0048	0.9461	0.9285

Table 5. Recommendation performance in terms of RMSE on the "Electronics-Movies&TV"

	ϕ	MF	CMF	MF_MLP	AVE	SRPM
K = 20	20%	1.4663	1.3083	1.3994	1.2403	1.1908
	50%	1.4703	1.3334	1.4447	1.2641	1.2145
	70%	1.4794	1.3509	1.4572	1.2861	1.2434
K = 50	20%	1.6300	1.3332	1.4996	1.2403	1.1948
	50%	1.7071	1.3778	1.5021	1.2641	1.2246
	70%	1.7682	1.4047	1.5113	1.2861	1.2591
K = 100	20%	1.7494	1.3422	1.5338	1.2403	1.1991
	50%	1.8300	1.3893	1.5726	1.2641	1.2248
	70%	1.8950	1.4347	1.5929	1.2861	1.2612

5.2 Performance Comparison

Recommendation Performance. Experimental results of MAE and RMSE on the two pair of domains "Books-Movies&TV" and "Electronics-Movies&TV" are presented in Tables 2, 3, 4 and 5, respectively. The domain "Books" and "Electronics" are chosen as the target domain in each pair of domains because they are extremely sparse.

We respectively evaluate all the methods under different K and ϕ in both pair of domains. From these tables, we can see that the proposed SRPM outperforms all baseline models in terms of both MAE and RMSE metrics. With the proportion of cold-start users becoming higher, the performance of single domain method MF will become progressively worse while the cross-domain methods keep satisfactory results, which shows the effectiveness of knowledge transfer. Compared with CMF and MF_MLP, our method SRPM gets an improvement of 5% to 10% both in RMSE and MAE. These results demonstrate that the SRPM is more suitable for making recommendations to cold-start users compared to other cross-domain baseline methods, especially in the dataset with high sparsity. SRPM performs better than AVG especially in higher ϕ, which demonstrates that the SRP transferred from the source domain is highly effective. And MF_MLP outperforms MF, indicating that the MLP based mapping function is feasible in knowledge transfer. For the proposed SRPM method, the optimal value of K is nearly 50 in "Books-Movie&TV" and nearly 20 in "Electronics-Movies&TV".

6 Related Work

Existing works about cross-domain algorithm mostly extract domain-specific information from ratings [5, 12], tags [2] and the text information like reviews [16]. Ren [8] proposed the PCLF model to learn the shared common rating pattern across multiple rating matrices and the domain-specific rating patterns from each domain. Fang [2] exploited the rating patterns across multiple domains by transferring the tag co-occurrence matrix information. Xin [16] exploited review text by learning a non-linear mapping on users' preferences on different topics across domains. On the whole, the main difference between our work and the previous approaches is the utilization of sentiment analysis method and mapping function which can predict the sentiment review pattern of cold-start users in the target domain and make cross-domain recommendations.

Sentiment analysis is widely used in recommendation systems. Computing the sentiment orientation of a user review has been studied by several researchers. Diao [17] built a language model component in their proposed JMARS model to capture aspects and sentiments hidden in reviews. Zhang [18] extracted explicit product features and user opinions by phrase-level sentiment analysis on user reviews to generate explainable recommendation results. Li [19] proposed a SUIT model to simultaneously utilize the textual topic and user item factors for sentiment analysis. In this paper, we employ sentiment analysis on cross-domain recommendation task and focus on discovering user's sentiment review pattern and mapping it from the source domain to the target domain.

7 Conclusions

The user reviews contain plenty of sentiment information. We proposed a novel framework for cross domain recommendation that establishes linkages between the source and target domains by using sentiment review pattern of users. In this paper, a sentiment review pattern extracting algorithm was proposed. We employed smoothed LDA and MLP based mapping method to model user's SRP and map it to the target domain to make recommendations for cold-start users. In different scenarios, that is to say, experiments convincingly demonstrate that the proposed SRPM framework can significantly improve the quality of cross-domain recommendation and SRPs extracted from reviews are important links between each domain.

Acknowledgments. This work is supported by NSF of Shandong, China (Nos. ZR2017MF065, ZR2018MF014), the Science and Technology Development Plan Project of Shandong, China (No. 2016GGX101034).

References

1. Cremonesi, P., Tripodi, A., Turrin, R.: Cross-domain recommender systems. In: ICDM 2012, pp. 496–503 (2012)
2. Fang, Z., Gao, S., Li, B., Li, J.: Cross-domain recommendation via tag matrix transfer. In: ICDM 2015, pp. 1235–1240 (2015)
3. Chen, W., Hsu, W., Lee, M.L.: Making recommendations from multiple domains. In: KDD 2013, pp. 892–900 (2013)
4. Yang, D., He, J., Qin, H., Xiao, Y., Wang, W.: A graph-based recommendation across heterogeneous domains. In: CIKM 2015, pp. 463–472 (2015)
5. Li, B., Yang, Q., Xue, X.: Transfer learning for collaborative filtering via a rating-matrix generative model. In: ICML 2009, pp. 617–624 (2009)
6. Ruck, D.W., Rogers, S.K., Kabrisky, M., Oxley, M.E., Suter, B.W.: The multilayer perceptron as an approximation to a bayes optimal discriminant function. IEEE Trans. Neural Netw. **1**(4), 296–298 (1990)
7. Taboada, M., Brooke, J., Tofiloski, M., Voll, K., Stede, M.: Lexicon-based methods for sentiment analysis. Comput. Linguist. **37**(2), 267–307 (2011)
8. Ren, S., Gao, S., Liao, J., Guo, J.: Improving cross-domain recommendation through probabilistic cluster-level latent factor model. In: AAAI 2015, pp. 4200–4201 (2015)
9. Gao, S., Luo, H., Chen, D., Li, S., Gallinari, P., Guo, J.: Cross-domain recommendation via cluster-level latent factor model. In: Blockeel, H., Kersting, K., Nijssen, S., Železný, F. (eds.) ECML PKDD 2013. LNCS (LNAI), vol. 8189, pp. 161–176. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40991-2_11
10. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet allocation. J. Mach. Learn. Res. **3**, 993–1022 (2003)
11. He, R., Mcauley, J.: Ups and downs: modeling the visual evolution of fashion trends with one-class collaborative filtering. In: WWW 2016, pp. 507–517 (2016)
12. Koren, Y., Bell, R., Volinsky, C.: Matrix factorization techniques for recommender systems. Computer **42**(8), 30–37 (2009)
13. Pan, W., Liu, N.N., Xiang, E.W., Yang, Q.: Transfer learning to predict missing ratings via heterogeneous user feedbacks. In: AAAI 2011, pp. 2318–2323 (2011)
14. Singh, A.P., Kumar, G., Gupta, R.: Relational learning via collective matrix factorization. In: KDD 2008, pp. 650–658 (2008)
15. Man, T., Shen, H., Jin, X., Cheng, X.: Cross-domain recommendation: an embedding and mapping approach. In: IJCAI 2017, pp. 2464–2470 (2017)
16. Xin, X., Liu, Z., Lin, C., Huang, H., Wei, X., Guo, P.: Cross-domain collaborative filtering with review text. In: IJCAI 2015, pp. 1827–1833 (2015)

17. Diao, Q., Qiu, M., Wu, C.Y., Smola, A.J., et al.: Jointly modeling aspects, ratings and sentiments for movie recommendation (JMARS). In: KDD 2014, pp. 193–202 (2014)
18. Zhang, Y., Lai, G., Zhang, M., et al.: Explicit factor models for explainable recommendation based on phrase-level sentiment analysis. In: SIGIR 2014, pp. 83–92 (2014)
19. Li, F., Wang, S., Liu, S., Zhang, M.: SUIT: a supervised user-item based topic model for sentiment analysis. In: AAAI 2014, pp. 1636–1642 (2014)
20. Baccianella, S., Esuli, A., Sebastiani, F.: SentiWordNet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In: International Conference on Language Resources and Evaluation, vol. 2542, pp. 83–90 (2010)