Personalized Semantic Ranking for Collaborative Recommendation

Song Xu
Institute of Automation,
Chinese Academy of Sciences
Beijing, China
song.xu@nlpr.ia.ac.cn

Shu Wu
Institute of Automation,
Chinese Academy of Sciences
Beijing, China
shu.wu@nlpr.ia.ac.cn

Liang Wang
Institute of Automation,
Chinese Academy of Sciences
Beijing, China
wangliang@nlpr.ia.ac.cn

ABSTRACT

Recently a ranking view of collaborative recommendation has received much attention in recommendation systems. Most of existing ranking approaches are based on pairwise assumption, i.e., everything that has not been selected is of less interest for a user. However it is usually not proper in many cases. To alleviate the limitation of this assumption, in this work, we present a unified framework, named Personalized Semantic Ranking (PSR). PSR models the personalized ranking and the user-generated content (UGC) simultaneously, and the semantic information extracted from UGC can make a remedy for the pairwise assumption. Moreover, utilizing the semantic information, PSR can capture the more subtle information of the user-item interaction and alleviate the overfitting problem caused by insufficient ratings. The learned topics in PSR can also serve as proper explanations for recommendation. Experimental results show that the proposed PSR yields significant improvements over the competitive compared methods on two typical datasets.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*Information Filtering*

Keywords

Learning to rank; recommendation; user-generated content

1. INTRODUCTION

In recent years there have been growing concerns over the problem of information overload. To cater the users' needs for finding right information, a lot of researches have been done on recommender systems (RS). Personalized recommendation involves a process of learning users' preferences by analyzing their feedback, either in explicit or implicit form, and delivering the right items to each user. As one of the most well-known approaches in RS, collaborative filtering
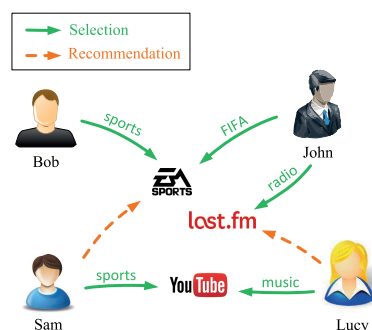
Figure 1: An illustration of user-generated content in website recommendation system. All the three websites, YouTube.com, last.fm and EASports.com, are represented by their logos respectively. The green solid lines represent "Selection", and the orange dashed lines denote "Recommendation". For example, Sam tags the website "YouTube" with "sports" and is recommended with "EASports".

(CF) exploits the similarity among users and recommend items the similar users liked.

In most of real-world circumstances, implicit feedback, such as clicks, purchases, is more frequently available. A characteristic of implicit feedback is that it is one-class, i.e., only positive examples are observed. Motivated by the work in the domain of learning-to-rank, a ranking view to recommendation [4, 7] provides us a good way to handle the implicit feedback. One of the most successful methods is Bayesian Personalized Ranking (BPR), which assumes that everything that has not been selected is of less interest for a user [7]. However this assumption is usually not proper in many cases. In order to overcome the limitation mentioned above, we utilize semantic information extracted from UGC to construct personalized ranking for better recommendation. In the recommendation scenarios, the user-generated content is assigned by certain user to certain item in the process of user-item interaction (such as rating, purchasing, etc.). In these scenarios, UGC provides us a further clue not only on the user interest but also on the item characteristic. By incorporating UGC, recommendation systems have the potential to generate more meaningful and effective recommendations for users.

For ease of explanation, we take the tags in website recommendation systems (e.g. Delicious[1]) as a specific example to illustrate some benefits of UGC. (1) As shown in Figure 1, we may infer that Bob prefers "EASports" over "Lastfm" following the pairwise assumption in BPR. Nevertheless it is obviously unreasonable to making a judgment that Bob prefers "EASports" over "YouTube". The semantic similarity (e.g. common tags) between "EASports" and "YouTube" can make a remedy for the pairwise assumption, whereas the semantic difference between "EASports" and "Lastfm" strengthen the pairwise hypothesis. (2) Utilizing the UGC is a good choice to solve the rating sparsity problem. In Figure 1, both Sam and Lucy have selected the website "YouTube". Merely from this behavior, we may infer that they have the same interest. However, through further observation of their tags on "YouTube", we realize Sam prefers "sports" video while Lucy prefers "music" video. These behaviors cannot differentiate users' interests or item properties while the UGC can. From the above example, we notice that UGC is capable to reflect more subtle information, which may not be revealed by very sparse rating behavior. (3) As the UGC captures both the user interest and item characteristic explicitly, it can help us understand user preferences explicitly, which makes a great difference in recommendation systems. After capturing the user preference, RS can demonstrate the topics user interested in, rather than describe redundantly and blindly with a lot of annoying details. In the case described in Figure 1, we can recommend Sam "EAsports" with proper explanation, such as, "it is a popular sports site". This adequate and persuasive explanation would make the recommendation more acceptable.

In this paper, we present a novel probabilistic framework, named Personalized Semantic Ranking (PSR), which jointly modeling the personalized ranking and UGC in a common subspace. It alleviates the rigid pairwise assumption in BPR by seamlessly incorporating semantic information for recommendation. Our method captures both the collaborative preference and the content-based preference, wick can also alleviate the overfitting problem caused by sparse rating data. PSR can also exhibit the user interests and item characteristics in an explicit manner, which can be interpreted as a vector of topics discovered from UGC.

2. RELATED WORK

Collaborative filtering with implicit feedbacks has been steadily receiving more attention. For item prediction, [3] and [5] propose a regularized least-square optimization with case weights (WR-MF). The case weights can be used to reduce the impact of negative examples. Despite its popularity, the above approaches often performs poorly compared to more recent models based on ranked loss minimization. The main reason is that all these approaches focus on the task of accurately predicting the exact rating values, while ranked loss minimization focus on accurately picking the top N recommended items, which is a more practical goal. As one of the most popular methods, Bayesian Personalized Ranking (BPR) is still under a rigid assumption, that everything that has not been selected is of less interest for a user [7]. Besides, all the above approaches suffer from the problem of poor explanation for recommendation.

---

[1] http://www.delicious.com

There are several methods which integrate attribute information about users and items into recommender systems. For example fLDA [1] and CTR [9] utilize topic models [2] for recommendation. Instead of learning one representation from a single document in these methods, PSR learns each user/item latent vector from all the relevant documents, which is more appropriate for UGC. Even some general frameworks like [6] can not well utilize the UGC.

3. PRELIMINARY

Notations In our scenario, all the observations can be denoted by $\mathcal{G} = (\mathcal{U}, \mathcal{V}, \mathcal{E}^+, \mathcal{W})$, where $\mathcal{U} = \{u_1, ..., u_N\}$ is the user set, $\mathcal{V} = \{v_1, ..., v_M\}$ is the item set. The positive feedback is given by user-item interaction set $\mathcal{E}^+ \subseteq \mathcal{U} \times \mathcal{V}$, and the element $(u, i) \in \mathcal{E}^+$ represents that user $u \in \mathcal{U}$ has purchased/clicked on item $i \in \mathcal{V}$. All the UGC is a collection of documents denoted by $\mathcal{W} = \{\mathbf{w}_{u,i} | u \in \mathcal{U}, i \in \mathcal{V}\}$. The document $\mathbf{w}_{u,i}$ is the content posted by user u to item i, and its t-th word is $W_{u,i,t}$, where $t \in \{1, ..., T_{u,i}\}$. If there is no content between user u and item i, word count $T_{u,i} = 0$.

Personalized Ranking from Implicit Feedback For each user, we transform the implicit feedback $\mathcal{E}^+$ to personalized ranking $\mathcal{R}$. Instead of modeling the implicit feedback $\mathcal{E}^+$ directly, we adopt a popular pairwise approach mentioned in BPR [7]. The idea is to discriminate the positive items $\mathcal{V}_u^+ = \{i | (u, i) \in \mathcal{E}^+\}$ from the remaining items $\mathcal{V} \backslash \mathcal{V}_u^+$ for each user u. Let "user u prefers item i to item j" be specified as $i \succ_u j$. If an item i has been selected by user u, then we assume that the user prefers this item over all other non-observed items. This property is formulated as follows:

$$\forall i, j \in \mathcal{V} : i \in \mathcal{V}_u^+ \wedge j \in \mathcal{V} \backslash \mathcal{V}_u^+ \Leftrightarrow i \succ_u j. \quad (1)$$

For convenience, we employ a random variable $R_{u,i,j}$ to denote the above pairwise property

$$R_{u,i,j} = \begin{cases} 1 & \text{if } i \succ_u j \\ 0 & \text{else} \end{cases} \quad (2)$$

For any two different items $i, j \in \mathcal{V}$, we have $R_{u,i,j} = 1 \Leftrightarrow R_{u,j,i} = 0$ (or $R_{u,i,j} = 0 \Leftrightarrow R_{u,j,i} = 1$). In our experiments, we just take the $R_{u,i,j} = 1$ ($R_{u,j,i} = 0$) into consideration. The set of all pairwise preferences can be represented as $\mathcal{T} = \{(u, i, j) | i \in \mathcal{V}_u^+ \wedge j \in \mathcal{V} \backslash \mathcal{V}_u^+\}$, and the ranking set is $\mathcal{R} = \{R_{u,i,j} | (u, i, j) \in \mathcal{T}\}$. Our model is expected to accurately preserve the ranking order.

4. PERSONALIZED SEMANTIC RANKING

Probabilistic Framework PSR maps both users and items to a common latent factor space $\mathbb{R}^K$ (each dimension represents a latent factor). Each user u is associated with a factor vector $\mathbf{U}_u \in \mathbb{R}^K$, which encodes the user preferences, and each item i is associated with a factor vector $\mathbf{V}_i \in \mathbb{R}^K$, representing the item characteristics. Both $\mathcal{W}$ and $\mathcal{R}$ can be generated from the latent factors.

For constructing the ranking $\mathcal{R}$, we first assume that each user-item pair has a rating score, which reflects the preference degree. Thus we have a preference score $x_{u,i}$ for user-item pair (u, i) and another score $x_{u,j}$ for a different pair (u, j). The probability that user u prefers item i to item j is $p(R_{u,i,j} = 1)$, which is parameterized by the relative score $x_{u,i,j} = x_{u,i} - x_{u,j}$. Here, we adopt a scoring function, which is widely used in factor models, $x_{u,i} = \mathbf{U}_u^T \mathbf{V}_i$.
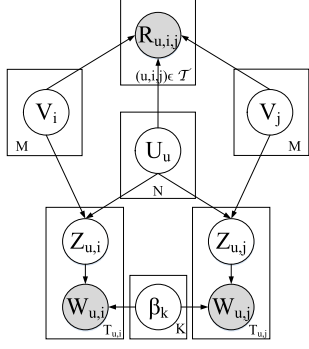
Figure 2: Graphical model of PSR. There are N users and M items. R is the personalized ranking. For some of the user-item pairs, there are documents extracted from the UGC. The Gaussian prior parameters for U and V have been omitted for simplicity in this figure.

For constructing the content $\mathcal{W}$, a latent representation for $\mathbf{w}_{u,i}$ is introduced as $\boldsymbol{\theta}_{u,i}$. It is also called topic proportions in topic models [2], where $\mathbf{z}_{u,i}$ is the topic assignments for each word in document $\mathbf{w}_{u,i}$ and each topic $\boldsymbol{\beta}_k$ is a distribution over words. $\boldsymbol{\theta}_{u,i}$ encodes not only the latent semantic of document $\mathbf{w}_{u,i}$, but also the user u's topic preferences on item i. Each component of $\boldsymbol{\theta}_{u,i}$ is expected to be positively correlated with both the corresponding user factor and item factor. That is, if an item has higher value of a certain factor, or a user values higher on a certain topic, the corresponding topic is more likely to appear in the UGC. In order to preserve such dependent correlation, we introduce a simple way to build a direct connection between the user-item pair and document, $\boldsymbol{\theta}_{u,i} = \pi(\mathbf{U}_u + \mathbf{V}_i)$. For a K-dimensional vector $\mathbf{x}$, $\pi(\mathbf{x})$ is logistic transformation function $\pi(\mathbf{x}) = \exp(\mathbf{x})/\sum_k \exp(x_k)$, which makes it possible to bound $\mathbf{x}$ within the range $[0, 1]$.

In order to balance the model's complexity against overfitting on the training data, Gaussian priors are also imposed on the factors $\mathbf{U}$ and $\mathbf{V}$. The graphical model is depicted in Figure 2. The generative process for all the textual documents $\mathcal{W}$ and personalized ranking $\mathcal{R}$ is as follows:

1. For each user u, draw user factors $\mathbf{U}_u \sim \mathcal{N}(\boldsymbol{\mu}_U, \boldsymbol{\Sigma}_U)$

2. For each item i, draw item factors $\mathbf{V}_i \sim \mathcal{N}(\boldsymbol{\mu}_V, \boldsymbol{\Sigma}_V)$

3. For each word $W_{u,i,t}$ / $W_{u,j,t}$
   (a) Draw topic assignment $Z_{u,i,t}|\mathbf{U}_u, \mathbf{V}_i \sim Mult(\boldsymbol{\theta}_{u,i})$
   (b) Draw word $W_{u,i,t}|Z_{u,i,t} \sim Mult(\boldsymbol{\beta}_{Z_{u,i,t}})$

4. For each triplet $(u, i, j) \in \mathcal{T}$, draw the ranking $R_{u,i,j} \sim Ber(1/(1 + \exp(-x_{u,i,j})))$

For ease of exposition, let $\Theta = [\boldsymbol{\mu}_U, \boldsymbol{\mu}_V, \boldsymbol{\Sigma}_U, \boldsymbol{\Sigma}_V, \boldsymbol{\beta}]$ denote the model parameters, $\Delta = [\mathbf{U}, \mathbf{V}, \mathbf{Z}]$ denote the latent variables. The joint probability distribution for the observed variables can be written as

$$p(\mathcal{R}, \mathcal{W}|\Theta) = \iint p(\mathbf{U})p(\mathbf{V})p(\mathcal{R}|\mathbf{U}, \mathbf{V})p(\mathcal{W}|\mathbf{U}, \mathbf{V})d\mathbf{U}d\mathbf{V}, \quad (3)$$

where Θ is omitted in the right for brevity.

Maximum A Posterior The objective of our model is to find the optimal latent factors $\mathbf{U}$ and $\mathbf{V}$ for accurately modeling ranking and UGC. Maximum a posterior (MAP) estimation can be used to obtain a point estimate of the posterior based on the observed data, $[\hat{\mathbf{U}}, \hat{\mathbf{V}}] = \arg\max_{\mathbf{U}, \mathbf{V}} p(\mathbf{U}, \mathbf{V}|\mathcal{W}, \mathcal{R}; \Theta)$. The negative log likelihood can be used as loss function

$$\mathcal{L}_{MAP} = -\log p(\mathbf{U}, \mathbf{V}|\mathcal{W}, \mathcal{R}; \Theta) = \overbrace{-\log p(\mathcal{R}|\mathbf{U}, \mathbf{V})}^{\mathcal{L}_R}$$
$$\overbrace{-\log p(\mathcal{W}|\mathbf{U}, \mathbf{V})}^{\mathcal{L}_W} \overbrace{-\log p(\mathbf{U}) - \log p(\mathbf{V})}^{\mathcal{L}_{reg}}. \quad (4)$$

In Equation (4), we notice that loss function consists of three parts, ranking loss $\mathcal{L}_R$, content loss $\mathcal{L}_W$ and regularization loss $\mathcal{L}_{reg}$. The ranking loss $\mathcal{L}_R$, which ensures successfully preserving the ranking, will be discussed in the next section. The content loss can be denoted by $\mathcal{L}_W = -\sum_{u,i} \sum_t n_{u,i,t} \sum_k \log \theta_{u,i,k} \beta_{k, W_{u,i,t}}$, where $n_{u,i,t}$ denotes how often the word $W_{u,i,t}$ occurred in document $\mathbf{w}_{u,i}$. If we place zero-mean spherical Gaussian priors on $\mathbf{U}$ and $\mathbf{V}$, $\mathcal{L}_{reg}$ works like $L2$ regularization which prevent models from overfitting [8]. Given the model parameters Θ, the optimal factors $[\mathbf{U}, \mathbf{V}]$ should be obtained by minimization of all three losses. Then each score $x_{u,i}$ can be predicted by $\mathbf{U}_u^T \mathbf{V}_i$. After sorting all the scores for each user, the top N items will be recommended to the user.

5. EXPERIMENTS

Datasets and Settings We evaluate our model on two published datasets: social bookmarking dataset from Delicious and scientific literature from Citation-network. In Delicious dataset, each user bookmarks a set of webpages which can be regarded as positive examples of the user's interests. Here, the bookmark is a kind of UGC, and the task is to recommend webpages to users. In our experiments, we draw a subsample such that each user has bookmarked at least 5 webpages, each webpage has been selected by at least 3 users and each bookmark appears at least 5 times. Citation-network $V1$ is a dataset released by Arnetminer. In this circumstance, we treat the paper title as a kind of UGC between the author and the publication, and our task is to recommend publications to authors. Removing papers which have no publication information, we obtain a subset of papers published from 2001 to 2010.

Baselines and Evaluation Metric We conduct comparisons with several state-of-the-art methods, including SVD, WR-MF [3, 5], BPR-MF [7] and FM [6]. In SVD, the missing is treated as negative examples. In FM, the UGC are directly treated as additional dimension in feature vector. In order to evaluate the qualities of the recommendations, we randomly select 20% of the implicit ratings to be the test set, and use the remaining entries for training. Two metrics, F1-score and Mean Average Precision (MAP), which are widely used for recommendation evaluation, are adopted here.

Result Analysis Table 1 gives the performance comparison with a variety of dimensionalities on the Delicious and the Citation-network datasets. From that we can see, under different settings of K, our models consistently outperform the compared methods. Figure 3a is the precision-recall curve on Delicious dataset. From the results we notice that the personalized ranking methods have great superiority over rating prediction methods. Owing much to the se-

Table 1: Performance comparison on the Delicious and Citation-network datasets with three different factor dimensionalities $K = 10, 30, 50$. The predictive accuracy is measured by MAP and F1-score. The values are under the percentage scale.

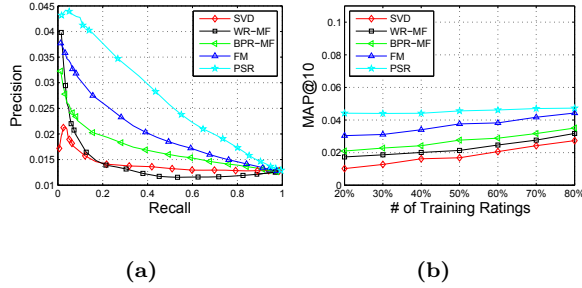| Method	Delicious						Citation-network					
	F1@10			MAP@10			F1@10			MAP@10		
	K=10	K=30	K=50	K=10	K=30	K=50	K=10	K=30	K=50	K=10	K=30	K=50
SVD	2.52	2.73	2.73	1.91	2.24	1.99	0.91	0.91	1.07	3.17	3.33	3.42
WR-MF	3.03	3.17	4.34	2.65	2.90	2.97	1.27	1.28	1.32	4.02	4.15	4.80
BPR-MF	3.35	3.50	3.57	2.54	2.87	2.71	1.53	1.51	1.57	6.08	5.80	6.12
FM	5.57	5.52	5.61	4.43	4.63	4.72	1.72	1.75	1.74	7.82	7.67	7.88
PSR	6.12	6.35	6.41	4.72	4.93	5.23	2.39	2.26	2.50	11.4	11.2	12.2

(a) **(b)**

Figure 3: Left panel: precision-recall curves of the Delicious dataset. Right panel: MAP with different amounts of training ratings on Delicious dataset. (K=30)

Table 2: Five topics discovered by PSR on the Citation-network dataset. Each topic is shown with the top 8 words, top 4 authors and top 5 publications. (K=30)

topic 00 "CV"	Topic 02 "Wireless"	topic 05 "Education"	topic 013 "Graphics"	topic 020 "Grid Computing"
learning	network	computer	motion	grid
analysis	sensor	science	surface	control
recognition	wireless	teaching	interactive	access
image	hoc	student	simulation	web
algorithm	routing	design	dynamic	computing
neural	energy	education	texture	resource
classification	mobile	assessment	mesh	management
face	protocol	study	human	composition
Jennifer C.	Joseph P.	Brian H.	Andrew S.	Chunlin L.
Guoliang X.	Sha L.	Jorma S.	Bryan E.	Layuan L.
Radha P.	Ajit W.	Marja K.	Eran G.	Brahim M.
Loukas L.	Antonio G.	Raymond L.	Yiying T.	Steve B.
ACCV	IJSNet	ICER	SIGGRAPH	IJWGS
SSVM	TOSN	ACE	I3D	CCGrid
PR	EWSN	SIGCSE	SGP	HPDC
TIP	SECON	JERIC	SPM	JSSPP
IJCV	WiSec	FECS	TOG	AusGrid

mantic analysis from UGC, PSR increases the performance by more than 20% compared to SVD and WR-MF. PSR achieves great improvement even compared with BPR-MF, which largely due to the semantic offset from naive pairwise assumption in BPR-MF. PSR works better than the general framework FM for well utilizing the UGC.

We further evaluate the ability of compared methods in handling different amount of training ratings ($20\% - 80\%$). The performance comparison with $K = 30$ on the Delicious dataset is shown in Figure 3b. The experimental results indicate that PSR can obtain the best performance with varying amounts of training ratings. These improvements are largely due to the incorporation of UGC. Moreover, our method outperforms other methods more significantly when fewer ratings are provided. When we use 80% of the ratings as training set, PSR increases the performance of SVD with 130%, while given 20% ratings, PSR enhances the performance by more than 360%. The evidences show that when the rating matrix is sparse, the impact of UGC is very significant in PSR.

Another advantage of PSR is that it can explain the user and item latent space using the topics discovered from the user-generated content. For user u, we can find the top matched topics by ranking the entries of factor vector $\mathbf{U}_u$. For item i, we also can rank the entries of factor vector $\mathbf{V}_i$. The top matched topics serve as an explanation of user interest and item characteristic. Some of the topics discovered in the Citation-network are displayed in Table 2. For example, the table illustrates that the top words in topic 02 are "network" , "sensor" etc., the most related authors of this topic are "Joseph P.", "Sha L. " etc, and the most related publications are "IJSNet", "TOSN" etc. Owing to the

good properties of interpretation, user experience would be greatly improved in real-world recommender systems.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] D. Agarwal and B.-C. Chen. flda: Matrix factorization through latent dirichlet allocation. In *WSDM*, 2010.

[2] D. M. Blei, A. Y. Ng, and M. Jordan. Latent dirichlet allocation. *JMLR*, 2003.

[3] Y. Hu, Y. Koren, and C. Volinsky. Collaborative filtering for implicit feedback datasets. In *ICDM*, 2008.

[4] J. Lee, S. Bengio, S. Kim, G. Lebanon, and Y. Singer. Local collaborative ranking. In *WWW*, 2014.

[5] R. Pan, Y. Zhou, B. Cao, N. N. Liu, R. Lukose, M. Scholz, and Q. Yang. One-class collaborative filtering. In *ICDM*, 2008.

[6] S. Rendle. Factorization machines. In *ICDM*, 2010.

[7] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme. Bpr: Bayesian personalized ranking from implicit feedback. In *UAI*, 2009.

[8] R. Salakhutdinov and A. Mnih. Probabilistic matrix factorization. In *NIPS*, 2007.

[9] C. Wang and D. M. Blei. Collaborative topic modeling for recommending scientific articles. In *ACM SIGKDD*, 2011.