Weighted Item Ranking for Pairwise Matrix Factorization

Haiyang Zhang[1], Ivan Ganchev[1,2], Nikola S. Nikolov[1,3], Máirtín O'Droma[1]
[1] Telecommunications Research Centre (TRC), University of Limerick, Ireland
[2] Department of Computer Systems, University of Plovdiv "Paisii Hilendarski", Plovdiv, Bulgaria
[3] Department of Computer Science and Information Systems, University of Limerick, Ireland
E-mail: {Haiyang.Zhang;Ivan.Ganchev;Nikola.Nikolov; Mairtin.ODroma }@ul.ie

***Abstract*— Recommendation systems employed on the Internet aim to serve users by recommending items which will likely be of interest to them. The recommendation problem could be cast as either a rating estimation problem which aims to predict as accurately as possible for a user the rating values of items which are yet unrated by that user, or as a ranking problem which aims to find the top-*k* ranked items that would be of most interest to a user, which s/he has not ranked yet. In contexts where explicit item ratings of other users may not be available, the ranking prediction could be more important than the rating prediction. Most of the existing ranking-based prediction approaches consider items as having equal weights which is not always the case. Different weights of items could be regarded as a reflection of items' importance, or desirability, to users. In this paper, we propose to integrate variable item weights with a ranking-based matrix factorization model, where learning is driven by Bayesian Personalized Ranking (BPR). Two ranking-based models utilizing different-weight learning methods are proposed and the performance of both models is confirmed as being better than the standard BPR method.**

Keywords*— *collaborative filtering; matrix factorization; Bayesian Personalized Ranking; implicit feedback; item recommendation

I. Introduction

With the rapid growth of the Web, people are inundated with massive information. Recommendation systems have become essential tools to assist users to discover the information they may be interested in and as such have been widely accepted by them [1]. The objective of the recommendation task can be treated either as a rating estimation problem which aims to predict the ratings for unrated user-item pairs as accurate as possible, or as a ranking problem which aims to find unrated items with the highest prediction score to a given user [2]. Both rating and ranking methods are widely studied in the literature. However, many current works have switched from developing more accurate rating prediction models to ranking-oriented models, especially in the case when explicit feedback of numerical rating is not available [3].

No matter rating-based or ranking-based, the recommendation models rely on users' past feedbacks, which are either explicit (ratings, reviews, etc.) or implicit (clicks, browsing history, etc.). In many recommendation scenarios, explicit rating data is sparse or even non-existent, since users are usually reluctant to spend extra time or effort on supplying that information. In such cases, the preference of users could be approximated by their implicit feedback. Compared with explicit feedback information, implicit feedback is closer to the real-industry perception of the problem and potential recommendation solutions, where the feedback data may be collected automatically at a much larger and faster scale with no user efforts needed [3].

Matrix Factorization (MF) has been a prior choice for solving recommendation problems due to its simplicity and outstanding performance in its capacity to yield maximum likelihood estimations of rankings or ratings for unconsumed user-item pairs. The MF based recommendation algorithms could be classified into rating-based or ranking-based approaches. The rating-based approaches – the so called pointwise regression methods – try to learn vectors of latent features of both users and items by minimizing a pointwise square error to approximate the rating matrix [4]. The ranking-based approaches – known as pairwise ranking methods – assume that a user prefers consumed items to unconsumed ones; thus they take pairs of items (consumed and unconsumed) for a user as basic unit and then try to maximize the likelihood of pairwise preference over consumed items and unconsumed items for that user [5]. Another advantage of the MF methods, whether rating-based or ranking-based, is their ability to incorporate additional information [6]. Many studies in the literature have been done to integrate various types of information with rating-based MF models, e.g. [6-9]. However, few studies have been done to exploit additional information for ranking-based MF.

Bayesian Personalized Ranking (BPR) is the most widely used ranking-based optimization approach and it has been experimentally proven to outperform the pointwise methods [5]. However, one problem with the BPR is that it assumes that all items have the same weight when the pairwise differentials between items for a certain user are computed. It is common knowledge that in recommendation scenarios the items importance may vary relative to one another due to their popularity or other features. Based on these observations, in this paper, relying on implicit feedback data and its exploitation, we propose to combine the pairwise MF based on BPR [5] with the item weights, which are learnt using item bias and PageRank, in a unified model. We compare the proposed unified models with the traditional BPR approach and the experimental results show that both proposed models have better performance.

The rest of the paper is organized as follows. Section II introduces the BPR-based MF approach and reviews the collaborative filtering that utilizes item weights. Section III describes the proposed recommendation models, whereas Section IV presents their experimental evaluation. Finally, Section V concludes the paper.

II. RELATED WORK

A. *Matrix Factorization Based on Bayesian Personalized Ranking*

BPR is the state-of-the-art ranking model for implicit feedbacks, which (experimentally) has shown much better performance than the pointwise methods [5]. It assumes that a user prefers a consumed/viewed item to an unconsumed/unviewed item, trying to maximize the following posterior probability [5]:

$$p(\Theta \mid R) \propto p(R \mid \Theta) p(\Theta) \quad (1)$$

where R is the rating matrix, $p(R \mid \Theta)$ represents the likelihood of the desired preference structure for all users according to R, and Θ represents the parameters vector of an arbitrary model. Thus, BPR is based on pairwise comparisons between a small set of positive items and a very large set of negative items from the users' histories. It estimates parameters by minimizing the loss function defined as follows [5]:

$$O = -\sum_{u \in U} \sum_{i \in R_u^+, j \in R_u^-} \ln \sigma(\hat{r}_{ui} - \hat{r}_{uj}) + \lambda \|\Theta\|^2 \quad (2)$$

where $\sigma(x)$ is the sigmoid function, λ is the regularization parameter. $\hat{r}_{ui}$ and $\hat{r}_{uj}$ denote the rating score for positive item i and negative item j using the selected rating model, respectively.

For item ranking from implicit feedbacks, the MF using BPR learns the rank approximation to the original implicit feedback matrix, resulting in two low-rank matrices P and Q. Therefore, the rating score of a user u to an item i is computed as:

$$\hat{r}_{ui} = p_u q_i^T \quad (3).$$

And the objective function for the BPR-based MF is:

$$O = -\sum_{u \in U} \sum_{i \in R_u^+, j \in R_u^-} \ln \sigma(\hat{r}_{ui} - \hat{r}_{uj}) + \lambda(\|P\|^2 + \|Q\|^2) \quad (4)$$

where $\sigma(x) = 1/(1+\exp(-x))$.

B. *Collaborative Filtering with Additional Information*

Collaborative filtering (CF) is the most popular recommendation approach, and it is classified into two primary areas: latent factor models and neighborhood-based methods [6]. Matrix factorization (MF) is the most widely used latent factor model due to its competitive advantage which is the ability to incorporate additional information. Different type of information has been exploited for integration with the MF model. Koren, et al. [6] propose to add user- and item bias information to the rating-based MF model. These authors also propose to consider both the explicit- and implicit influence of user-item ratings in the rating-based MF model, called SVD++, which achieves more accurate results than majority of the recommendation models [7]. Neighborhood information has been also considered for incorporation into the MF models. Guo, et al. [10] propose a neighborhood-based MF technique which integrates item neighborhood with the standard rating-based MF. Similarly, Zheng, et al. [11] propose to consider user neighborhood for integration with the rating-based MF.

Different types of information have also been exploited for integration with the neighborhood-based CF approaches. Gao, et al. [12] claim that the importance of user's recommendation should vary to one another, since some users may be more important than others in a social group. These authors propose a user rank approach, which incorporates user weights into the computation of item similarities and differential for item-based K-nearest neighbor (KNN) and Slope One, respectively. User's trust information has also been utilized for incorporation with the item-based CF [13], where trust information is used in the process of item similarity computation.

In this paper, we propose to integrate item weights, which are learnt only based on the user-item interactions, with the ranking-based MF model to improve the performance of item ranking.

III. LEARNING TO RANK WITH ITEM WEIGHTS

Most of the current ranking-based MF models treat items as having equal weights. However, it is common sense that some items are more popular than others and thus are of higher importance. In this section, we explore the item weight influence for MF using the BPR model. In the BPR-based MF model, for a given user, it is assumed that the rating score of an observed item is higher than that of unobserved ones, and an attempt is made to maximize the predicted rating differential in the model learning process. This predicted rating differential should be influenced by not only the user- and item factors differential but also by the item weight differential. Based on this hypothesis, two item-weighted BPR-based MF models for implicit feedbacks are proposed in the following subsections.

A. *Weighted item ranking with item bias*

Inspired by the pointwise MF technique [6], we first propose to utilize item bias to learn item weights. Therefore, the predicted rating for user u to item i is given as:

$$\hat{r}_{ui} = p_u q_i + b_i \quad (5)$$

where b_i denotes the item weight. Thus, the objective function of the BPR optimization method is presented as:

$$O = -\sum_{u \in U} \sum_{i \in R_u^+, j \in R_u^-} \ln \sigma(\hat{r}_{ui} - \hat{r}_{uj}) + \lambda(\sum_{u \in U} \|p_u\|^2 + \sum_{i \in R_u^+} (\|q_i\|^2 + \|b_i\|^2) + \sum_{j \in R_u^-} (\|q_j\|^2 + \|b_j\|^2))) \quad (6).$$

We employ stochastic gradient descent (SGD) [14] to estimate the parameters, which are updated as:

$$x_{uij} \leftarrow \hat{r}_{ui} - \hat{r}_{uj}$$
$$b_i \leftarrow b_i + \gamma(\frac{e^{-xuij}}{1+e^{-xuij}} - regB \cdot b_i)$$
$$b_j \leftarrow b_j + \gamma(-\frac{e^{-xuij}}{1+e^{-xuij}} - regB \cdot b_j)$$
$$p_u \leftarrow p_u + \gamma(\frac{e^{-xuij}}{1+e^{-xuij}}(q_i - q_j) - regU \cdot p_u) \quad (7)$$
$$q_i \leftarrow q_i + \gamma(\frac{e^{-xuij}}{1+e^{-xuij}} p_u - regI \cdot q_i)$$
$$q_j \leftarrow q_j + \gamma(-\frac{e^{-xuij}}{1+e^{-xuij}} p_u - regI \cdot q_j)$$

where γ is the learning rate and $regB$, $regU$, $regI$ are the regularized term for the biased factor, user factors, and item factors, respectively.

B. Weighted item ranking using PageRank

PageRank [15] is a very successful technique for ranking nodes in a certain graph, such as a webgraph or a social network [16]. It computes the importance score for each node in the graph according to the graph connectivity. PageRank has also been used as a weight learning method in recommendation systems, based on different degrees of correlation between users or items [12, 17]. Here, we propose to leverage PageRank to compute item importance first, and then to incorporate it into the MF model with given weights. The predicted model is represented as:

$$\hat{r}_{ui} = \alpha p_u q_i + (1-\alpha) w_i \quad (8)$$

where α is used to balance the influence of the original rating learnt from the MF model and the item weight, w_i is the normalized weight for item i learnt using the PageRank technique, with $w(i) = (PR(i) - \min(PR)) / (\max(PR) - \min(PR))$, where PR is the set of vectors representing item ranking using PageRank and $PR(i)$ is the ranking score for item *i*. PR is calculated as:

$$PR = \alpha \cdot CM \cdot PR + (1-\alpha) \cdot \frac{1}{|V|} \cdot 1_{|V|} \quad (9)$$

where CM is the item correlation matrix, *V* is the set of items, and α is a decay factor. Details of the PageRank computing process can be found in [18].

The objective function using the BPR optimization model therefore is given as:

$$O = -\sum_{u \in U} \sum_{i \in R_u^+, j \in R_u^-} \ln \sigma(\hat{r}_{ui} - \hat{r}_{uj}) + \lambda(\sum_{u \in U} \|p_u\|^2 + \sum_{i \in R_u^+} \|q_i\|^2 + \sum_{j \in R_u^-} \|q_j\|^2) \quad (10).$$

The parameters are updated using SGD [14] as:

$$x_{uij} \leftarrow \hat{r}_{ui} - \hat{r}_{uj}$$
$$p_u \leftarrow p_u + \gamma(\frac{e^{-xuij}}{1+e^{-xuij}} \alpha(q_i - q_j) - regU \cdot p_u)$$
$$q_i \leftarrow q_i + \gamma(\frac{e^{-xuij}}{1+e^{-xuij}} \alpha p_u - regI \cdot q_i) \quad (11)$$
$$q_j \leftarrow q_j + \gamma(-\frac{e^{-xuij}}{1+e^{-xuij}} \alpha p_u - regI \cdot q_j)$$

IV. Experimets

A. Experiment Setup

We evaluated the performance of the two proposed models on three different public datasets, namely FilmTrust [19], MovieLens-100k (ML-100k)[1] and MovieLens-1M (ML-1M)[2], with details presented in Table 1.

Since the FilmTrust dataset is very sparse, we regard every user–movie pair with rating as an observed interaction. However, for ML-100k and ML-1M, only ratings higher than 3 were kept as part of the observed positive feedback. After this pre-processing, the statistics of the three datasets are shown in

Features	***FilmTrust***	***ML-100k***	***ML-1M***
Users	1508	943	6040
Items	2071	1682	3952
Ratings	35497	100000	1000209
Density	1.14%	6.31%	4.19%

TABLE II. Statistics (*initial*) of the three public datasets used in the experiments

Table 2. In each dataset, we took 60% as a training set and the rest – as a test set. 50 triples $(u,i,j):(u,i) \succ (u,j)$ were randomly generated for each user in the training set.

B. Evaluation Metrics and Comparative Approaches

To evaluate the effectiveness of item ranking, the standard evaluation metrics for rating-based approaches, such as mean absolute error (MAE) and root mean square error (RMSE) are not suitable. In this study, we deploy the well-studied top-k ranking metrics used for information retrieval evaluation,

Features	***FilmTrust***	***ML-100k***	***ML-1M***
Users	1508	943	6040
Items	2071	1682	3952
Interactions	35497	**71623**	**836478**
Density	1.14%	**4.52%**	**3.5%**

TABLE I. Statistics (after pre-processing) of the three public datasets used in the experiments.

namely *precision*, *recall*, *F1-measure*, and *normalized discounted cumulated gain* (NDCG) [20] to the ranked lists found by our proposed models.

[1] http://grouplens.org/datasets/movielens/100k/

[2] https://grouplens.org/datasets/movielens/1m/

FilmTrust	Prec@10	Rec@10	F1@10	NCDG@5	NCDG@10
BPR-MF	0.4278	0.4968	0.4597	0.4849	0.4483
WBPR-Bias	0.4300	0.4993	0.4618	0.4867	0.4511
WBPR-P2	0.4298	0.4993	0.4619	0.4905	0.4547

ML-100k	Prec@10	Rec@10	F1@10	NCDG@5	NCDG@10
BPR-MF	0.2222	0.1367	0.1692	0.2745	0.2417
WBPR-Bias	0.2226	0.1418	0.1733	0.2760	0.2424
WBPR-P2	0.2368	0.1461	0.1807	0.2864	0.2548

ML-1M	Prec@10	Rec@10	F1@10	NCDG@5	NCDG@10
BPR-MF	0.2212	0.0775	0.1148	0.2585	0.2368
WBPR-Bias	0.2222	0.0780	0.1155	0.2587	0.2372
WBPR-P2	0.2221	0.0778	0.1152	0.2586	0.2376

TABLE III. PERFORMANCE COMPARISON USING FILMTRUST AND MOVIELENS DATASETS (d=10, iterations=30, LRATE=0.02 , REGU=REGI=0.1, α=0.8).

- Prec@k: *Precision* indicates how many items are actually relevant among all recommended items. For a given user *u*, the precision value of a ranked recommendation list at position *k* is defined as:

$$Prec@k = \frac{1}{k}\sum_{i=1}^{k}\delta(x_u(i) \in I_u^{test}) \ ,$$

where $x_u(i)$ is the predicted ranking list for user *u* and I_u^{test} is the set of preferred items by user *u* in the test set. $\delta(y) = 1$ if y is true; otherwise $\delta(y) = 0$.

- Rec@k: *Recall* gives the number of recommended items among all relevant items. For a given user *u*, the recall at position *k* of a ranked recommendation list is defined as:

$$Rec@k = \frac{1}{\left|I_u^{test}\right|}\sum_{i=1}^{k}\delta(x_u(i) \in I_u^{test})$$

where $\left|I_u^{test}\right|$ is the number of preferred items by user *u* in the test set.

- F1@k: The F1-measure is the harmonic mean of *precision* and *recall* [21], which is defined as:

$$F1@k = \frac{2 * Prec@k * Rec@k}{Prec@k + Rec@k}.$$

- NDCG@k: NDCG measures the quality of a recommendation model based on the graded relevance of the ranked items [22]. The NDCG value at position *k* of a ranked item list for a given user *u* is defined as:

$$nDCG@k = \frac{1}{Z_u}\sum_{i=1}^{k}\frac{2^{\delta(x_u(i) \in I_u^{test})} - 1}{\log(i+1)} \ ,$$

where $Z_u = \sum_i^k \frac{1}{\log(i+1)}$ is the normalization term.

To demonstrate the effectiveness of the proposed models, we compared their performance to the traditional BPR [5], which has been experimentally shown to have better performance than some well-known pointwise methods. We name our proposed models WBPR-bias (Section III.A) and WBPR-P2 (Section III.B), respectively.

C. Results and Discussion

The experimental results for the proposed recommendation models and the traditional BPR model are presented in Table 3, with learning rate equal to 0.02 and α equal to 0.8 for the WBPR-P2 model. As one can observe, both proposed models outperform the standard BPR-based MF model for all evaluation metrics, which indicates that the consideration of item weights has positive effect on the model performance.

Figure 1 shows the experimental results for precision vs. recall, based on the Movielens-100k dataset, for different

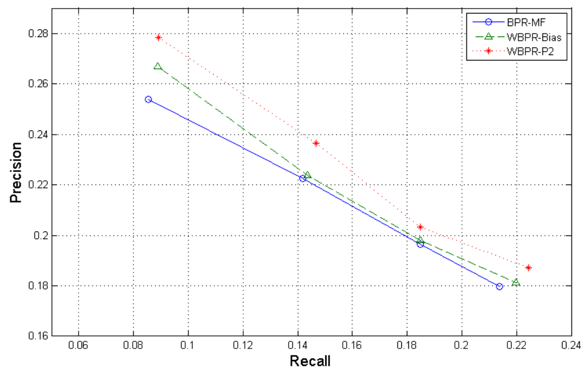

Fig. 1. Precision vs. recall for the proposed weighted BPR models and the traditional BPR-MF model for the ML-100l dataset (d=10 and iterations=30).

values of k, where k denotes the size of the ranked list. The results clearly demonstrate that both proposed item-ranking-based models show better performance than the standard BPR-based MF model, over a range of top-k values. For this dataset, the PageRank-based approach performs slightly better than the item bias-based approach.

V. CONCLUSIODN

This paper has proposed to incorporate item weights into the ranking-based matrix factorization (MF) approach for improving the performance of item ranking. Two weighted-item ranking recommendation models, based on the Bayesian Personalized Ranking (BPR), have been implemented and trialled. Experimental results confirm that the consideration of item weights in the BPR model helps improve the recommendation accuracy.

ACKNOWLEDGMENT

This publication has been supported by the Chinese Scholarship Council (CSC), the Telecommunications Research Centre (TRC), University of Limerick, Ireland, and the NPD of the University of Plovdiv “Paisii Hilendarski”, Bulgaria, under Grant No. ФП17-ФМИ-008.

REFERENCE

[1] M. Gao, K. Liu, and Z. Wu, “Personalisation in web computing and informatics: Theories, techniques, applications, and future research,” Information Systems Frontiers, vol. 12, no. 5, Nov. 2010, pp. 607-629.

[2] T. Zhao, J. McAuley, and I. King, “Leveraging social connections to improve personalized ranking for collaborative filtering,” in Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, ser. CIKM ’14. New York, NY, USA: ACM, 2014, pp. 261-270.

[3] W. Pan, H. Zhong, C. Xu, and Z. Ming, Adaptive bayesian personalized ranking for heterogeneous implicit feedbacks, Knowledge-Based Systems, vol. 73, 2015, pp. 173 - 180.

[4] Y. Hu, Y. Koren, and C. Volinsky, Collaborative filtering for implicit feedback datasets, in Proceedings of the 8th IEEE International Conference on Data Mining, Dec 2008, pp. 263-272.

[5] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme, ”BPR: Bayesian personalized ranking from implicit feedback,” in Proceedings of the 25th conference on uncertainty in artificial intelligence. AUAI Press, 2009, pp. 452-461.

[6] Y. Koren, R. Bell, and C. Volinsky, ”Matrix Factorization Techniques for Recommender Systems,” Computer, vol. 42, 2009, pp. 30-37.

[7] Y. Koren, ”Factorization meets the neighborhood: a multifaceted collaborative filtering model,” in Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, 2008, pp. 426-434.

[8] G. Guo, J. Zhang, and N. Yorke-Smith, ”A Novel Recommendation Model Regularized with User Trust and Item Ratings,” IEEE Transactions on Knowledge and Data Engineering, vol. 28, 2016, pp. 1607-1620.

[9] B. Yang, Y. Lei, J. Liu, and W. Li, Social collaborative filtering by trust, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PP, no. 99, 2017, pp. 114.

[10] M.-j. Guo, J.-g. Sun, and X.-f. Meng, ”A Neighborhood-based Matrix Factorization Technique for Recommendation,” Annals of Data Science, vol. 2, 2015, pp. 301-316.

[11] Z. Zheng, H. Ma, M. R. Lyu, and I. King, ”Collaborative web service qos prediction via neighborhood integrated matrix factorization,” IEEE Transactions on Services Computing, vol. 6, 2013, pp. 289-299.

[12] M. Gao, Z. Wu, and F. Jiang, ”Userrank for item-based collaborative filtering recommendation,” Information Processing Letters, vol. 111, 2011, pp. 440-446.

[13] H. Zhang, I. Ganchev, N. S. Nikolov, and M. O. Droma, ”A trustenriched approach for item-based collaborative filtering recommendations,” in Proceedings of the IEEE 12th International Conference on Intelligent Computer Communication and Processing (ICCP), 2016, pp. 65-68.

[14] M. Zinkevich, M. Weimer, L. Li, and A. J. Smola, Parallelized stochastic gradient descent, in Advances in neural information processing systems, 2010, pp. 2595-2603.

[15] L. Page, S. Brin, R. Motwani, and T. Winograd, The pagerank citation ranking: Bringing order to the web.Stanford Digital Library Technologies Project. 1999.

[16] Pedroche F. Ranking nodes in Social Network Sites using biased PageRank. Instituto de Matem atica Multidisciplinaria, Universidad Politcnica de Valencia 2010; E-46022 Valencia.

[17] M. Gori and A. Pucci, Itemrank: A random-walk based scoring algorithm for recommender engines, in Proceedings of the 20th International Joint Conference on Artifical Intelligence, ser. IJCAI07. San Francisco, CA, USA, 2007, pp.2766-2771.

[18] H. Zhang, I. Ganchev, N. S. Nikolov, Z. Ji, and M. O. Droma, ”Weighted Matrix Factorization with Bayesian Personalized Ranking,,” accepted for the SAI Computing Conference, London, UK, 2017.

[19] J. Golbeck, ”Filmtrust: movie recommendations from semantic webbased social networks,” in Proceedings of the 3rd IEEE Consumer Communications and Networking Conference,2006, pp. 1314-1315.

[20] H. Chen and D. R. Karger, Less is more: Probabilistic models for retrieving fewer relevant documents, in Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, ser. SIGIR 06. New York, NY, USA: ACM, 2006, pp. 429-436.

[21] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl, ”Evaluating collaborative filtering recommender systems,” ACM Transactions on Information Systems (TOIS), vol. 22, 2004, pp. 5-53.

[22] Y. Wang, L. Wang, Y. Li, D. He, W. Chen, and T.-Y. Liu, ”A theoretical analysis of NDCG ranking measures,” in Proceedings of the 26th, Annual Conference on Learning Theory 2013.pp. 2554.