RNR: A Generic Bayesian-based Framework for Enhancing Top-N Recommender Systems

Rong-qin Xu*
Shenzhen University
Shenzhen, China
xurongqin2018@email.szu.edu.cn

Mingyang Zhou
Shenzhen University
Shenzhen, China
zmy@szu.edu.cn

Hao liao
Shenzhen University
Shenzhen, China
jamesliao520@gmail.com

ABSTRACT

In personalized top-N recommender systems, a core task is to design effective methods to measure user-item preference scores and then to suggest, for each user, a small set of personalized items with high scores. However, little attention was paid to the recommendation of low-score user-item links. In this work, based on the Bayesian estimation theory, we propose a novel metric RNR (**R**ecall-to-**N**oise **R**atio) to characterize the ability to recommend both high-score and low-score user-item links. Then we propose a generic framework that leverages **RNR** to transfer the link scores of the state-of-the-art recommendation methods. Empirical experiments show that the proposed framework could optimally improve the recommendation accuracy.

KEYWORDS

Top-N recommender systems, Bayesian theory, Accuracy

ACM Reference Format:
Rong-qin Xu, Mingyang Zhou, and Hao liao. 2020. RNR: A Generic Bayesian-based Framework for Enhancing Top-N Recommender Systems. In *Companion Proceedings of the Web Conference 2020 (WWW '20 Companion), April 20–24, 2020, Taipei, Taiwan.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3366424.3382719

1 INTRODUCTION

In most top-N recommender systems, we usually assign a specific preference score for pairwise user and item, and then suggest a set of N items with the highest scores to each user. The key issue is to design effective algorithms to measure the user-item likelihood. Over decades, there has been different kinds of recommendation methods to improve recommendation accuracy, including collaborative filtering, matrix factorization, neural network based methods and so on [6]. However, most recommendation methods only recommend items with high scores, neglecting the low-score ones which are in line with the particular interests of users. Nevertheless, since irrelevant items also have low scores, it remains a great challenge to precisely recognize the low-score potential items in the discipline. Unlike most of previous approaches that design different preference score models, we explore the validity of the fundamental underlying hypothesis that whether a user is inclined to consume items with high preference scores to him/her. Inspired by the Bayesian estimation theory and signal-to-noise curves [2], we propose a novel metric, named RNR (**R**ecall-to-**N**oise **R**atio), to characterize the ability to recommend potential items. Based on RNR, a novel and generic framework is put forwarded to redefine the preference scores of an arbitrary existing method by an one-variable transferring function. The proposed framework benefits the recommendation of low-score user-item links, which will further benefit improving recommendation quality.

*Corresponding author

WWW '20 Companion, April 20–24, 2020, Taipei, Taiwan

ACM ISBN 978-1-4503-7024-0/20/04.
https://doi.org/10.1145/3366424.3382719

2 THE PROPOSED FRAMEWORK

Supposing a user-item bipartite network $G = (U, I, L)$, where U and I denote the set of users and items respectively and $(u, i) \in L$ represents a link. If the user u provides feedback for item i, the entry $r_{ui} = 1$; otherwise $r_{ui} = 0$. The links L in graph G are divided into training set L^T and probe set L^P, $L^T \cup L^P = L$ and $L^T \cap L^P = \emptyset$. s_{ui} is the preference score of a user u towards an item i, obtained by an arbitrary recommendation method. We usually implicitly assume that a user-item link with higher score is more likely to be a potential future link, which could be rephrased as

$$p(r_{ui} = 1|s_{ui}) > p(r_{u'i'} = 1|s_{u'i'}), if \quad s_{ui} > s_{u'i'}, \tag{1}$$

where s_{ui} and $s_{u'i'}$ represent the preference scores of two non-existing links in the training set L^T respectively.

According to the bayesian estimation,

$$p(r_{ui} = 1|s_{ui}) = \frac{p(r_{ui} = 1) \cdot p(s_{ui}|r_{ui} = 1)}{p(s_{ui})}, \tag{2}$$

where $p(s_{ui})$ is the probability that a random chosen (existing or non-existing) link in the training set has score s_{ui}. It is obvious that $p(s_{ui})$ is mainly determined by the score distribution of non-existing links. That is to say, $p(s_{ui}) \approx p_n(s_{ui})$, where $p_n(s_{ui})$ is the probability that a random chosen non-existing link in the training set has score s_{ui}. Thus, we arrive at

$$p(r_{ui} = 1|s_{ui}) = \frac{p(r_{ui} = 1) \cdot p(s_{ui}|r_{ui} = 1)}{p_n(s_{ui})} = c_1 \cdot \frac{p(s_{ui}|r_{ui} = 1)}{p_n(s_{ui})}, \tag{3}$$

where $c_1 = p(r_{ui} = 1) = |L^T|/(|U| \times |I|)$ is a constant representing the probability of an existing link between a random pairwise user item in the training set. Neglecting the constant term, we propose **R**ecall-to-**N**oise **R**atio (RNR),

$$RNR(s) = \frac{p(s|r_{ui} = 1)}{p_n(s)} = \frac{p_r(s)}{p_n(s)}, \tag{4}$$

where $p_r(s) = p(s|r_{ui} = 1)$ is the probability that a random chosen existing link in the training set has score s. Thus, our framework use $s' = RNR(s)$, an one-variable transferring function, to transfer original scores into s' and then use s' the recommend items.

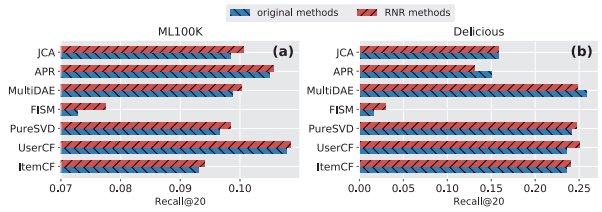

Figure 1: Comparisons of *Recall@20* results of the RNR recommended links with that of the corresponding original methods on two datasets. (a) ML100K. (b) Delicious.

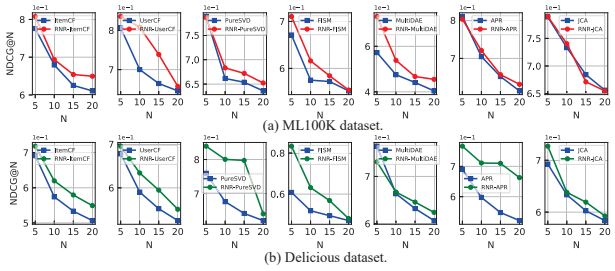

Figure 2: The *NDCG@N* as a function of *N* by seven state-of-the-art methods and the corresponding RNR methods on two datasets. (a) ML100K. (b) Delicious.

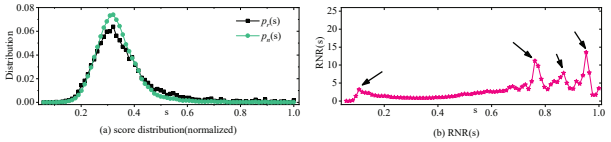

Figure 3: The score distributions of $p_r(s)$ and $p_n(s)$, and the corresponding $RNR(s)$ of Delicious dataset, where preference scores are obtained by FISM method.

3 EXPERIMENTS

3.1 Experimental Setup

Datasets. We conduct experiments on two datasets: (1) **ML100K** dataset has 943 users, 1682 items and 100000 user-item interactions (links). (b) **Delicious** dataset has 1867 users, 69227 items with 104799 user-item interactions (links). We randomly divide each dataset to 90% training set and 10% probe set.

Baselines. We compare our method with these state-of-the-art recommendation methods: ItemCF [6], UserCF [6], PureSVD [1], FISM [4], MultiDAE [5], APR [3] and JCA [7]. We used the original authors's preferred hyper-parameter setting.

Evaluation metrics. The accuracy is mainly measured by two metrics: *Recall@N* and *NDCG@N* (**N**ormalized **D**iscounted **C**umulative **G**ain), where *NDCG@N* considers the position of the correctly recommended items while *Recall@N* does not.

3.2 Experimental Results

3.2.1 Empirical RNR(s) Analysis. We first present an example of the score distributions $p_r(s)$, $p_n(s)$ and the $RNR(s)$ in Fig. 3. In Fig. 3(a), scores of most links are around 0.32 while $s > 0.6$ and $s < 0.1$, $p_n(s) \to 0$. In Fig. 3(b), apart from the high score $s \approx 0.97$, some low scores ($s \approx 0.78$ or 0.09) also have high RNR(s), demonstrating recognizing low-score links and the effectiveness of our proposed scheme as well as the failure of the classical assumption in eq. 1.

3.2.2 Accuracy Performance. **Recall performance:** Figure 1 reports the recall performance. In Fig. 1, our proposed RNR framework significantly outperforms the original methods in most cases, except the MultiDAE and APR on Delicious dataset. Detailed experiments show that in MultiDAE and APR methods, when we calculate $RNR(s) = p_r(s)/p_n(s)$, $p_n(s) \to 0$ introduces much fluctuation for $RNR(s)$. The fluctuation of $RNR(s)$ largely influences the recall accuracy for MultiDAE and APR methods on Delicious dataset.

NDCG performance: Figure 2 shows the *NDCG@N* performance as a function of *N*. In Fig.2, RNR-based methods outperform the state-of-the-art methods in most cases. In practical scenarios, the forward positions usually mean a high click probability and high attraction. We can conclude that our method is inclined to push correct items to the forward positions.

4 CONCLUSIONS

In this paper, we propose a generic framework that utilize a metric, named RNR, to redefine the preference scores between users and items. The experiment results validate that RNR-based methods have higher recommendation accuracy than the original methods. Besides, the RNR function could be efficiently calculated only by counting the score difference of existing and non-existing links, which does not improve the time complexity of classical methods.

ACKNOWLEDGMENTS

Rong-qin Xu thanks to the singer Jay Chou. This work is jointly supported by Science and Technology Innovation Commission of Shenzhen (Nos. JCYJ20160520162743717, JCYJ201506251015-24056, JCYJ20150731160834611), Shenzhen Science and Technology Foundation (Nos. JCYJ20150529164656096, JCYJ2017030215395-5969), P-ostgraduate Innovation Development Fund Project of Shenzhen University (No. PIDFP-ZR2019016) and Tencent Open Research Fund.

REFERENCES

[1] Paolo Cremonesi, Yehuda Koren, and Roberto Turrin. 2010. Performance of recommender algorithms on top-n recommendation tasks. In *Proceedings of the fourth ACM conference on Recommender systems*. ACM, 39–46.
[2] Jesse Davis and Mark Goadrich. 2006. The relationship between Precision-Recall and ROC curves. In *Proceedings of the 23rd international conference on Machine learning*. ACM, 233–240.
[3] Xiangnan He, Zhankui He, Xiaoyu Du, and Tat-Seng Chua. 2018. Adversarial personalized ranking for recommendation. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. ACM, 355–364.
[4] Santosh Kabbur, Xia Ning, and George Karypis. 2013. Fism: factored item similarity models for top-n recommender systems. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 659–667.
[5] Dawen Liang, Rahul G Krishnan, Matthew D Hoffman, and Tony Jebara. 2018. Variational autoencoders for collaborative filtering. In *Proceedings of the 2018 World Wide Web Conference*. International World Wide Web Conferences Steering Committee, 689–698.
[6] Jie Lu, Dianshuang Wu, Mingsong Mao, Wei Wang, and Guangquan Zhang. 2015. Recommender system application developments: a survey. *Decision Support Systems* 74 (2015), 12–32.
[7] Ziwei Zhu, Jianling Wang, and James Caverlee. 2019. Improving Top-K Recommendation via Joint Collaborative Autoencoders. In *The World Wide Web Conference*. ACM, 3483–3482.