Negative-Aware Collaborative Filtering

Sheng-Chieh Lin
Academia Sinica, Taiwan
jacklin_64@citi.sinica.edu.tw

Yu-Neng Chuang
National Chengchi University, Taiwan
107753011@nccu.edu.tw

Sheng-Fang Yang
National Chengchi University, Taiwan
106753011@nccu.edu.tw

Ming-Feng Tsai
National Chengchi University, Taiwan
mftsai@nccu.edu.tw

Chuan-Ju Wang
Academia Sinica, Taiwan
cjwang@citi.sinica.edu.tw

ABSTRACT

Most traditional recommender systems regard unseen user-item associations as negative user preferences and optimize recommendation models mainly based on observed associations and some *negative instances* sampled from unseen associations. However, such unseen user-item associations may contain potential positive user preferences on items and are not uniformly distributed in terms of the possibility of being negative (or positive) user preference; therefore, it is essential to quantify such associations for model training. Along this line, in this paper, in contrast to existing recommendation models, which equally treat all unseen associations as negative samples, we present a negative-aware recommendation approach that explicitly models the likelihood of each unseen association being a potentially positive preference. Empirical results on real-world datasets in different fields show that our approach consistently improves recommendation performance.

KEYWORDS

recommendation, collaborative filtering, unseen associations, asymmetric user similarity

ACM RecSys 2019 Late-breaking Results, 16th-20th September 2019, Copenhagen, Denmark

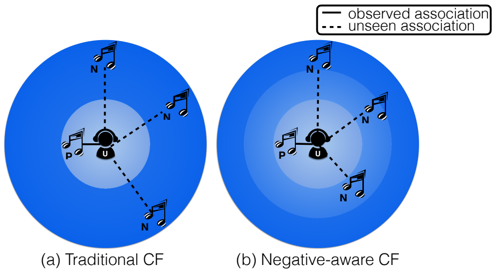

Figure 1: An illustrative example for negative-aware collaborative filtering

INTRODUCTION

With the rapid development of online services over the last decade, recommender systems have gained much importance, finding use in areas such as music, news, movies, books, and products in general. For such an important problem, collaborative filtering (CF) is a common yet powerful approach that generates user recommendations using only user-item interaction data [3]. Some CF-based algorithms have been shown to yield reasonable performance among diverse situations and have been used in many real-world applications [7].

One major challenge for CF-based recommendation algorithms is the sparsity of interaction data; that is, most users provide feedback on only a few items. This challenge is attributed to the fact that in most recommendation scenarios, there is an extremely large pool of items; thus it is unfeasible to expect user feedback for most items. As shown in Figure 1(a), traditional model-based collaborative filtering takes this into account by treating observed interactions as positive associations and treating the majority of unseen interactions as negative ones. However, this approach introduces noise into the modeling process as unseen interactions are not necessarily to be negative instances. In the literature, a few studies attempt to implicitly address this problem. For example, weighted regularized matrix factorization (WRMF) [2] treats unseen associations as a kind of uncertainty instead of negative feedback and uses case weights to reduce the impact of negative examples. In addition, Bayesian personalized ranking (BPR) [6] deals with such uncertainty problem by modeling relative user preference on items. Despite that, these studies consider that unseen associations are uniformly distributed in terms of the possibility of being negative user preferences and do not granulate these associations by quantifying the degree of uncertainty.

In this paper, in contrast to existing recommendation models, which equally treat all unseen associations as negative user preferences, we propose quantifying the degree of uncertainty for unseen associations by leveraging user preference similarity, and explicitly model the likelihood of each unseen association being a potentially positive user preference (illustrated in Figure 1(b)). Note that the proposed quantification of unseen associations can be applicable to other recommendation algorithms with the use of *negative sampling*. Empirical results on two real-world datasets show that our approach improves recommendation performance.

METHODOLOGY

In collaborative filtering(CF), an interaction matrix, denoted as $\mathcal{A} = (a_{u,i}) \in \mathbb{R}^{|U| \times |I|}$, represents the user-item associations, where U and I denote the sets of users and items, respectively; $a_{u,i} = 1$ if there exits an observed association between user $u \in U$ and item $i \in I$, and otherwise, $a_{u,i} = 0$.

We first introduce the negative-aware matrix $\mathcal{N} \in \mathbb{R}^{|U| \times |I|}$ to quantify the uncertainty of unseen user-item associations for later recommendation model training, each element $n_{ui} \in \mathcal{N}$ of which is

(a) Adjacency matrix (symmetric) (b) Shared item matrix (symmetric) (c) Asymmetric relations (d) User similarity matrix (asymmetric)

Figure 2: Asymmetric user preference similarity

Dataset	Movielens	CiteULike
Users ($\|U\|$)	938	3,527
Items ($\|I\|$)	950	6.339
Feedback	54,413	77,546
Density	6.100%	0.347%

Table 1: Datasets

calculated as

$$n_{ui} = \begin{cases} 0 & \text{if } \sum_{k=1}^{|U|} \hat{s}_{uk} a_{ki} = 0, \\ \left(\sum_{k=1}^{|U|} \hat{s}_{uk} a_{ki}\right) / \left(\sum_{k=1}^{|U|} 1_{\{\hat{s}_{uk}>0\}} a_{ki}\right) & \text{otherwise.} \end{cases}$$

where $\hat{\mathcal{S}} = (\hat{s}_{jk}) \in \mathbb{R}^{|U|\times|U|}$ denotes the user preference similarity.

Inspired by [4] [5], which show that asymmetric similarity can better represent the intrinsic characteristic of user preference similarity, each element in user preference similarity matrix $\hat{\mathcal{S}}$ is formulated as $\hat{s}_{jk} = \frac{|I_j \cap I_k|}{|I_k|}$, where I_j and I_k denote the sets of items liked by user j and k, respectively. The intuition of asymmetric similarity between user preference is shown in Figure 2(c): user j likely has a preference for the items liked by user k, but the reverse may not be true as user j has a much wider preference than user k. Furthermore, the denominator of each element n_{ui} in $\mathcal{N}$ is for normalization and denotes the number of users who have positive feedback for item i and at the same time have shared items with user u. Note that $0 \leq n_{ui} \leq 1$. For an unseen association between user u and item i, the intuition behind this design is that if the users that have given positive feedback on item i are on average very similar to user u, item i is likely to match user u's preference. Thus, n_{ui} (or $1 - n_{ui}$) can be interpreted as the likelihood of the association between user u and item i being a positive (or negative, respectively) preference.

We then tailor the negative-aware matrix using pointwise and pairwise approaches to account for implicit user feedback for recommendation. Both approaches attempt to estimate the latent factors of the following two sets: $\boldsymbol{\theta}_U, \boldsymbol{\theta}_I \subseteq \Theta$, where $\boldsymbol{\theta}_U \in \mathbb{R}^{|U|\times d}$ for users, $\boldsymbol{\theta}_I \in \mathbb{R}^{|I|\times d}$ for items, d is the dimension of the low-rank latent factor space, and Θ is a superset of $\boldsymbol{\theta}_I$ and $\boldsymbol{\theta}_U$ consisting of all the parameters in the model. Let $\theta_u(\theta_i)$ denote the row vector for user u (item i, respectively) from $\boldsymbol{\theta}_U$ ($\boldsymbol{\theta}_I$, respectively).

Of the pointwise approaches, the most representative method is matrix factorization (MF). To incorporate the designed negative-aware matrix into the optimization, we modify the objective function of MF for implicit feedback proposed by [2] as

$$\mathcal{L}_{\mathrm{MF}}^{\mathcal{N}} = \sum_{u,i} a_{ui} \left(1 - \theta_u^\top \theta_i\right)^2 + \left(1 - a_{ui}\right) \left(n_{ui} - \theta_u^\top \theta_i\right)^2 + \lambda \left\|\Theta\right\|_2^2, \tag{1}$$

where $u \in U$, $i \in I$, and λ is a hyperparameter preventing overfitting to the observations. Note that Eq. (1) can be seen as a variant of WRMF, the case weight of which is however a hyperparameter requiring to be exogenously determined; in contrast, n_{ui} plays a similar role of the case weight and is shaped by the observed user-item associations.

Dataset	Movielens				CiteULike			
	P@5	MAP@5	P@10	MAP@10	P@5	MAP@5	P@10	MAP@10
MF	0.237	0.169	0.199	0.123	0.060	0.058	0.045	0.048
$\text{MF}_{\text{n-aware}}$	*__0.241__	*__0.173__	*__0.202__	*__0.125__	**0.062**	*__0.061__	*__0.048__	*__0.052__
BPR	0.257	0.189	0.211	0.136	0.064	0.064	0.049	0.054
$\text{BPR}_{\text{n-aware}}$	*__0.262__	***__0.195__	**__0.214__	**__0.140__	*__0.066__	**0.066**	*__0.050__	**0.055**

Table 2: Top-N recommedation

For the pairwise approaches, we integrate the negative awareness of unseen associations into Bayesian Personalized Ranking (BPR)[6], a popular ranking-based model:

$$\mathcal{L}_{\text{BPR}}^{N} = -\sum_{u,(i,i')} \log\left(\frac{1}{n_{ui'}}\right) \log \sigma(\theta_u^{\intercal}\theta_i - \theta_u^{\intercal}\theta_{i'}) + \lambda \,\|\Theta\|_2^2 \,. \tag{2}$$

Note that as the proposed negative-aware matrix is independent of the recommendation models, it can be seen as a generic device applicable to other recommendation algorithms with the use of negative sampling.

EXPERIMENTS

We conduct experiments on two publicly available real-world datasets in different fields: MovieLens-100K,[1] and CiteULike.[2] To demonstrate the effectiveness of our negative-aware approach, we implement both pointwise and pairwise recommendation models based on the loss functions in Eqs. (1) and (2), and compare the models with and without incorporating the proposed negative-aware matrix. For all approaches, we set the dimension of latent factors d to 64 and the number of negative samples to 5. The dot product is used as the scoring function for an association given the latent factors of the corresponding user and item. To evaluate the model capability for the task of top-N item recommendation, we use the two commonly adopted evaluation metrics: precision@N and MAP@N [1]. For each dataset, we randomly divide the observed user-item associations into 80% and 20% as training and testing sets, respectively, and obtain the averaged results by randomly dividing the data 5 times in this manner.

Table 2 tabulates the performance of our negative-aware approach on pointwise and pairwise recommendation models. In the table, *, **, and *** indicate significance levels of $p < 0.05$, $p < 0.005$, and $p < 0.0005$ based on the paired t-test with respect to its counterpart; the reported numbers are averaged over the five test results.

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

[2] http://www.wanghao.in/data/ctrsr_datasets.rar

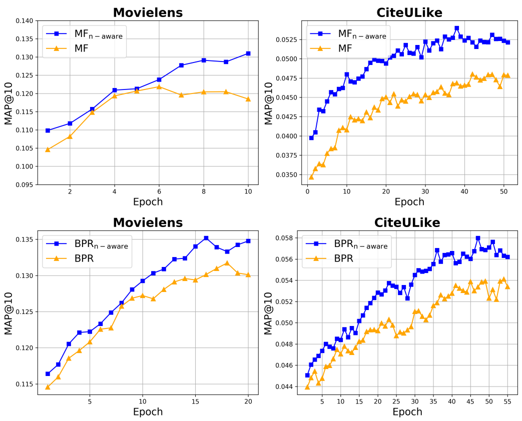

Figure 3: Performance (MAP@10) at each training stage

We first evaluate our models, denoted as MF$_{\text{n-aware}}$ and BPR$_{\text{n-aware}}$, on the original datasets with observed positive feedback, the results of which are listed in the top panel of Table 2. Observe that the proposed models in most cases outperform or yield performance comparable to their counterparts.

Figure 3 shows the model performance at different training epochs, where each point in the figure denotes the performance in terms of MAP@10. As shown in the figure, our negative-aware models (blue curves) are generally capable of maintaining better performance than the traditional models at each training epoch. This clearly demonstrates that our approach boosts the recommendation quality of both pointwise and pairwise recommendation CF models.

CONCLUSIONS

In this work, we present a negative-aware recommendation approach that explicitly addresses the uncertainty of unseen user-item associations This approach is shown to be applicable to recommendation algorithms with the use of *negative sampling*. Empirical results on two real-world datasets show that our approach improves the performance of pointwise and pairwise recommendation models.

REFERENCES

[1] Jonathan L. Herlocker, Joseph A. Konstan, Loren G. Terveen, and John T. Riedl. 2004. Evaluating Collaborative Filtering Recommender Systems. *ACM Transactions on Information Systems* 22, 1 (2004), 5–53.

[2] Yifan Hu, Yehuda Koren, and Chris Volinsky. 2008. Collaborative Filtering for Implicit Feedback Datasets. In *Proceedings of IEEE International Conference on Data Mining (ICDM '08)*, Vol. 8. 263–272.

[3] Dietmar Jannach, Paul Resnick, Alexander Tuzhilin, and Markus Zanker. 2016. Recommender Systems — Beyond Matrix Completion. *Communications of the ACM* 59, 11 (2016), 94–102.

[4] Marta Millan, Maria Trujillo, and Edward Ortiz. 2007. A Collaborative Recommender System Based on Asymmetric User Similarity. In *Intelligent Data Engineering and Automated Learning - IDEAL 2007*, Hujun Yin, Peter Tino, Emilio Corchado, Will Byrne, and Xin Yao (Eds.). 663–672.

[5] Parivash Pirasteh, Dosam Hwang, and Jason J. Jung. 2015. Exploiting matrix factorization to asymmetric user similarities in recommendation systems. *Knowl.-Based Syst.* 83 (2015), 51–57.

[6] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian Personalized Ranking from Implicit Feedback. In *Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence (UAI '09)*. 452–461.

[7] Xiaoyuan Su and Taghi M. Khoshgoftaar. 2009. A Survey of Collaborative Filtering Techniques. *Advances in Artificial Intelligence* 2009 (2009), Article ID 421425.