Computer Science and Information Systems 20(3):933–948 https://doi.org/10.2298/CSIS221031041K

Read between the Interactions: Understanding Non-interacted Items for Accurate Multimedia Recommendation

Jiyeon Kim*, Taeri Kim*, and Sang-Wook Kim†

Department of Computer Science, Hanyang University
222 Wangsimni-ro, Seongdong-gu, Seoul, Korea
{jiyeon7, taerik, wook}@hanyang.ac.kr

Abstract. This paper addresses the problem of multimedia recommendation that additionally utilizes multimedia data, such as visual and textual modalities of items along with the user-item interaction information. Existing multimedia recommender systems assume that *all the non-interacted items of a user have the same degree of negativity*, thus regarding them as candidates for negative samples when training the model. However, this paper claims that a user's non-interacted items do not have the same degree of negativity. We classify these non-interacted items of a user into two kinds of items with different characteristics: *unknown and uninteresting items*. Then, we propose a novel negative sampling technique that only considers the uninteresting items (*i.e.*, rather than the unknown items) as candidates for negative samples. In addition, we show that using the multiple Bayesian personalized ranking (BPR) losses with both unknown and uninteresting items (*i.e.*, all the non-interacted items) in existing multimedia recommendation methods is very effective in improving recommendation accuracy. By conducting extensive experiments with three real-world datasets, we show the superiority of our ideas. Our ideas can be easily and orthogonally applied to any multimedia recommender systems.

Keywords: recommender systems, multimedia recommendation, uninteresting items.

1. Introduction

Due to the abrupt increase in the number and variety of items around us, the problem of information overload is becoming a big issue in many applicantions. Recommender systems are a vital technique to solve this problem and thus are widely used in various domains, such as movie recommendations and music recommendations. *Collaborative filtering* (CF) is one of the most widely used approaches in recommender systems; intuitively, for a target user, it finds the items commonly preferred by the users with the tastes similar to hers (*i.e.*, neighbors) based on her interaction information (*e.g.*, purchase history and click logs) [6, 9, 11–16, 19, 21, 22, 25, 28, 31–34, 39]. Despite the simplicity and robustness of CF in recommendation, the sparse nature of the interaction information brings CF the limitation of not being able to accurately capture the users' preferences on items [3].

* The first two authors have equally contributed to this work.
† Corresponding author.

To alleviate this limitation of CF, various methods have been proposed [4, 8, 11, 18, 19, 21, 23, 36, 37, 43, 44]. They can be classified into two categories: i) additional utilization of non-interacted items and ii) additional utilization of external data. The methods in category i) first divide a user's non-interacted items into her unknown items and uninteresting items [11, 18, 19, 21] where the unknown items are the items that the user did not interact with because she did not know their existence and the uninteresting items are the items that the user did not interact with even though she knew their existence but did not want to interact with the item. Then, the methods mitigate the data sparsity problem by selecting her uninteresting items amongst non-interacted items and imputing low values for the uninteresting items selected [11, 18, 19, 21]. The methods in category ii) use additional multimedia data (*e.g.*, visual data such as the item's image and textual data such as the item's specifications) along with the user-item interaction information. The recommender systems of this category are referred to as *multimedia recommender system*s [4, 8, 23, 36, 37, 43, 44].

Most multimedia recommender systems use deep learning models such as convolutional neural networks (CNNs) [1, 17, 26, 42] and recurrent neural networks (RNNs) [7, 10, 38] to extract multimodal features from the items' multimedia data. They utilize these multimodal features to represent the item embeddings; they conduct a dot product between an item embedding and a user embedding to predict the user's preference on the item. They use the *Bayesian personalized ranking* (BPR) loss [30], a representative pairwise loss to learn the ranking difference between a user's positive and negative items, to train their models. In model training, positive items are sampled from the user's *interacted items* and the negative items are randomly sampled from the user's *non-interacted items*. In other words, they simply use all the non-interacted items as the candidates for negative items based on the assumption that *all the non-interacted items for a user have the same degree of negativity*.

However, we claim that this assumption does not hold in the real-world data; *i.e.*, non-interacted items could have different degrees of negativity. Then, we propose the methods that utilize the two categories of a user's non-interacted items for accurate multimedia recommendation. Note that the proposed methods can be easily applied (*i.e.*, orthogonally applicable) to existing multimedia recommender systems. To this end, we first classify a user's non-interacted items into two categories of unknown and uninteresting items for her based on the degrees of her negativity, obtained by using the user's interacted items. Then, we propose a novel negative sampling technique that uses only the uninteresting items (rather than unknown items) as candidates for negative samples. Furthermore, we propose to use multiple BPR losses which utilize both unknown and uninteresting items (*i.e.*, all non-interacted items), in multimedia recommender systems. To demonstrate the effectiveness of our proposed methods, we employ three well-known multimedia recommender systems (spec. VBPR [8], MMGCN [37], and LATTICE [43]) and three real-world Amazon datasets.[1] To show the superiority of our negative sampling method, we compare the following three methods: i) using those randomly sampled from non-interacted items as negative samples (*i.e.*, original negative sampling); ii) using unknown items as negative samples; iii) using uninteresting items as negative samples (*i.e.*, our negative sampling). Our experimental result shows that our proposed method (*i.e.*, method iii)) provides the best recommendation accuracy. The result also confirms that

[1] http://jmcauley.ucsd.edu/data/amazon/links.html

using multiple BPR losses is more effective in multimedia recommender systems: specifically, applying the multiple BPR losses leads to a gain of up to 20.13% and 4.12%, in terms of Recall@20, compared to the state-of-the-art multimedia recommender systems, MMGCN [37] and LATTICE [43], respectively.

The main contributions of our work are summarized as follows:

- We point out the problem of the assumption employed in existing multimedia recommender systems.
  - All the non-interacted items for a user have the same degree of negativity.
- We propose two methods that improve the recommendation accuracy by exploiting the different degrees of negativity in non-interacted items.
  - We propose a novel sampling technique that uses only the uninteresting items as negative samples.
  - We use multiple BPR losses to learn the rank differences between positive, unknown, and uninteresting items.
- We validate our proposed methods by conducting extensive experiments using three real-world datasets.

The rest of this paper is organized as follows. In Section 2, we briefly review the related work to multimedia recommender systems. In Section 3, we describe our proposed methods in detail. In Section 4, we conduct experiments to verify the effectiveness of our methods. Finally, in Section 5, we summarize and conclude our paper.

2. Related Work

In this section, we briefly introduce the research on multimedia recommender systems. Early multimedia recommender systems utilized only one modality amongst the items' multimedia data (*e.g.*, visual, textual, and acoustic modality) along with the user-item interaction information [2, 5, 8, 12, 24, 35, 40, 41]. VBPR [8], the most popular model among them, captures the features of the visual modality of items and builds an additional embedding that reflects each user's preference for the visual modality of items. Then, it uses the well-known BPR loss in training VBPR. However, the early multimedia recommender systems have a limitation that they use only one of various modalities of items to represent the items' characteristics.

To alleviate this limitation, recent multimedia recommender systems have tried to utilize various modalities of items [4, 14, 23, 36, 37, 43, 44]. Specifically, JRL [44] and MAML [23] use deep learning models to capture the features of the various modalities of the item (*e.g.*, visual, textual, and numerical (*i.e.*, rating) modalities for JRL and visual and textual modalities for MAML). Then, they aggregate the captured features to enrich the embedding of an item and the user who interacted with the corresponding item. MMGCN [37] constructs the user-item interaction graphs for visual, textual, and acoustic modalities of items, and uses graph convolutional networks (GCNs) to capture the collaborative signals between the users and the items. Then, MMGCN aggregates the collaborative signals captured by each modality and enriches the embeddings of users and items.

Since the advent of MMGCN, various GCN-based multimedia recommender systems have emerged such as GRCN [36] and LATTICE [43]. They commonly use not only

GCNs but also the attention mechanism to distinguish the degrees of influence of users on different modalities of items. GRCN [36] is based on MMGCN and considers the degrees of influence on different modalities at an individual user level. On the other hand, LATTICE [43] captures the latent item-item structure for each modality using visual and textual modalities of items and then applies GCN to obtain enriched item embeddings. Then, it considers the degrees of influence on different modalities at all user levels (*i.e.*, globally for all users).

The aforementioned methods utilize the BPR loss, which selects positive items among the interacted items and negative items among the non-interacted items and widens the rank discrepancy between positive and negative items, in order to learn their models [8, 36, 37, 43, 44]. However, since there are many items in recommendation domain data, it is difficult for users to know the existence of all items. Therefore, a user's non-interacted items can be categorized into unknown and uninteresting items as follows:

– **Unknown item**: item that a user did not interact with because she did not know its existence.
– **Uninteresting item**: item that a user did not interact with because she did not want to interact with it, even though she knew its existence.

In other words, if the BPR loss is simply employed in a learning process as in existing multimedia recommender systems, some non-interacted items that the user may prefer can be considered as her negative items. Therefore, we argue that, in order to correctly train the model by using the BPR loss, negative items should be sampled not from her non-interacted items, but from her uninteresting items. In addition, we argue that we should train the model so that they will be able to learn all the rank discrepancies among positive, unknown, and uninteresting items.

3. Proposed Methods

In this section, we propose two methods that can be orthogonally applied to existing multimedia recommender systems, exploiting the notions of unknown and uninteresting items for accurate multimedia recommendation. Specifically, in Section 3.1, we describe our novel negative sampling method that uses only uninteresting items as negative samples. Then, in Section 3.2, we describe how to use multiple BPR losses, for interesting, unknown, and uninteresting items in model training.

3.1. Negative Sampling Method

The overall procedure of our negative sampling is shown in Figure 1. As mentioned, a user's non-interacted items are categorized into unknown items and uninteresting items. Our negative sampling method samples only the uninteresting items of a user as negative samples for BPR training.

For this, we first compute the pre-use preferences of each user on her non-interacted items by analyzing users' interaction information. A user's pre-use preference is the preference that the user has when deciding whether to interact with an item or not [11, 18, 19, 21]. Thus, we can say that, for the user's interacted items, she holds a high pre-use preference. On the other hand, for those items that she has not interacted, her pre-use

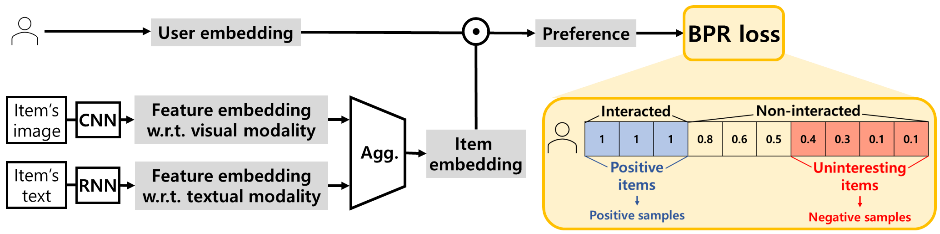

Fig. 1. Overview of our negative sampling method. The blue-colored items indicate the user's interacted items (*i.e.*, positive items), and the red-colored items indicate the user's uninteresting items (*i.e.*, negative items) among her non-interacted items

preferences would be lower than those of the interacted items. Among non-interacted items, the user's pre-use preferences on unknown items are unknown, however, those on uninteresting items should be low since she was not interested in them.

In order to obtain pre-use preference scores for non-interacted items of a user, in this paper, we employ WRMF [27], a widely adopted model in one-class setting, following [11, 18, 21].[2] Specifically, given the pre-use preference matrix $\mathbf{P} \in R^{\#\text{ of users}\times\#\text{ of items}}$ ($p_{u,i} = 1$, if user u has interacted with item i), WRMF predicts users' pre-use preferences for all non-interacted items. To this end, we first initialize users' pre-use preferences for non-interacted items as 0 in $\mathbf{P}$ and assign weights to quantify the relative contribution of each user-item interaction [27]. Then, WRMF repeats the process of decomposing the pre-use preference matrix $\mathbf{P}$ into two low-rank matrices $\mathbf{U} \in R^{\#\text{ of users}\times d}$ and $\mathbf{V} \in R^{\#\text{ of items}\times d}$ where d indicates the dimensionality of each latent feature vector and multiplying these two decomposed matrices to recover the original pre-use preference matrix $\mathbf{P}$. The loss function of WRMF is as follows:

$$\mathcal{L}(\mathbf{U},\mathbf{V}) = \sum_{u,i} w_{u,i}(p_{u,i} - \mathbf{U}_u\mathbf{V}_i^T)^2 + \lambda(\sum_u \|\mathbf{U}_u\|_F^2 + \sum_i \|\mathbf{V}_i\|_F^2), \tag{1}$$

where $p_{u,i}$ denotes user u's pre-use preference of item i; $w_{u,i}$ denotes the weight for $p_{u,i}$ and $\mathbf{U}_u$ and $\mathbf{V}_i$ represent the latent feature vectors of user u and item i, respectively; $\|\cdot\|_F$ denotes the *Frobenius norm* and λ denotes the regularization parameter.

Lastly, we obtain the predicted pre-use preference matrix $\hat{\mathbf{P}}$ using the learned vectors $\mathbf{U}$ and $\mathbf{V}$ as follows:

$$\hat{\mathbf{P}} = \mathbf{U}\mathbf{V}^T. \tag{2}$$

Then, we use $(1 - \hat{p}_{u,i})$ as the final weight for non-interacted item i to be sampled as a negative sample. By doing this, we allow the negative samples to be selected only from the uninteresting items, rather than from all non-interacted items. This is because the pre-use preference $\hat{p}_{u,i}$ of an uninteresting item will be low, thus making the weight (*i.e.*, $(1 - \hat{p}_{u,i})$) high. Finally, we use the negative samples selected from the uninteresting items in the BPR loss of the existing multimedia recommender systems.

[2] Note that, if there is a better model available other than WRMF, it could improve more the recommendation accuracies with our proposed method.

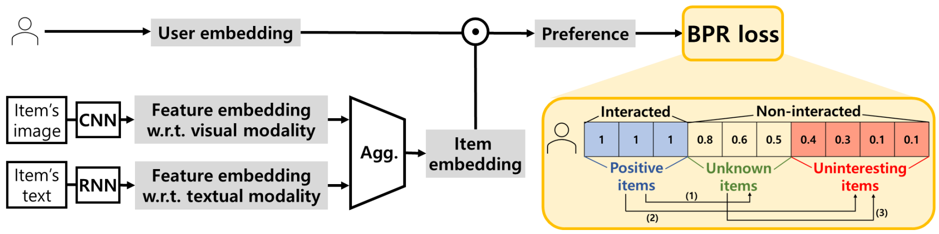

Fig. 2. Overview of the method using the multiple BPR losses. The blue-colored items indicate the user's interacted items, the gray-colored items indicate the user's unknown items among her non-interacted items, and the red-colored items indicate the user's uninteresting items among her non-interacted items. (1)-(3) indicates the rank discrepancies the method using the multiple BPR losses considers

3.2. Multiple BPR losses

Note that the BPR loss employed in existing multimedia recommender systems is only used to correctly learn the rank discrepancy between the predicted preferences of positive and negative items. However, the entire items can be divided into positive, unknown, and uninteresting items; so we can better learn the rank of non-interacted items if we fully exploit the rank discrepancy among all pairs of the above three types. Toward this end, we propose to use multiple BPR losses[3] in existing multimedia recommender systems, which enables to learn the rank discrepancies amongst the predicted preferences of the above three types of items (*i.e.*, not only uninteresting items but also unknown items can be used).

Our method using the multiple BPR losses is shown in Figure 2. With the items categorized into three types (*i.e.*, positive, unknown, and uninteresting items), we train the model by using the following three rank discrepancies: (1) between positive and unknown items, (2) between positive and uninteresting items, and (3) between unknown and uninteresting items. We use the three weights (*i.e.*, α for (1), β for (2), and γ for (3)) to control the importance of the three rank discrepancies. The multiple BPR losses are formulated as follows:

$$\mathcal{L} = -\sum_{u} \sigma(\alpha(\hat{r}_{pos} - \hat{r}_{unk})) + \sigma(\beta(\hat{r}_{pos} - \hat{r}_{unint})) + \sigma(\gamma(\hat{r}_{unk} - \hat{r}_{unint})) + R(\theta), \quad (3)$$

where $\hat{r}_{pos}, \hat{r}_{unk}$, and $\hat{r}_{unint}$ denote the predicted preferences of positive, unknown, and uninteresting items, respectively; $\sigma(\cdot)$ indicates the sigmoid function and $R(\theta)$ does the regularization term for model parameters θ.

To utilize our multiple BPR losses in multimedia recommender systems, we need a user's predicted pre-use preference scores of non-interacted items, as stated in Section 3.1. Then, with those scores, we regard the bottom $\mu\%$ of non-interacted items as uninteresting items, and the rest of them as unknown items. Lastly, we apply the multiple BPR losses in Eq. (3) to multimedia recommender systems.

The proposed methods in Sections 3.1 and 3.2, are easily and orthogonally applicable to existing multimedia recommender systems, helping to provide more accurate multimedia recommendations.

[3] A similar idea proposed in non-multimedia recommendation [20].

Table 1. Statistics of datasets. The sparsity calculated by $\frac{\text{\# of users} \cdot \text{\# of items} - \text{\# of interactions}}{\text{\# of users} \cdot \text{\# of items}} \times 100\ (\%)$.

Dataset	# of users	# of items	# of interactions	Sparsity
Amazon Baby	19,445	7,050	160,792	99.88%
Amazon Men Clothing	4,955	5,028	32,363	99.87%
Amazon Office	4,874	2,406	52,957	99.55%

4. Evaluation

In this section, we evaluate our proposed methods via experiments; the experiments are designed aiming to answer the following key evaluation questions:

- **EQ1**: Do the notions of unknown and uninteresting items help improve the recommendation accuracy of multimedia recommender systems?
- **EQ2**: Is the idea of selecting the uninteresting items as negative samples most effective for improving the recommendation accuracy?
- **EQ3**: How sensitive is the recommendation accuracy of the multiple BPR losses to different hyperparameter values?

4.1. Experimental Settings

Datasets and competitors For evaluation, we adopt three real-world Amazon datasets widely used in multimedia recommender systems research [4, 8, 23, 43, 44]: Amazon Baby, Amazon Men Clothing, and Amazon Office[4]. As done in [37], we kept only the users and items with more than five interactions. Table 1 reports their detailed statistics. These datasets contain visual and textual modality information of items as well as the user-item interaction. Then, we extracted 4,096-dimensional visual feature embeddings using the deep CNN [17] and 1,024-dimensional textual feature embeddings using sentence transformers [29], following [43].

To evaluate the effectiveness of our proposed methods, we use the following three baselines:

- VBPR [8]: A multimedia recommender system based on matrix factorization (MF) trained with a BPR loss.
- MMGCN [37]: A multimedia recommender system based on graph convolutional networks (GCNs) using non-linear propagation trained with a BPR loss.
- LATTICE [43]: A multimedia recommender system based on graph convolutional networks (GCNs) using linear propagation trained with a BPR loss.

Evaluation protocol and metrics We repeated all our experiments five times. For each experiment, we randomly split interactions per user into 8:1:1, each for train, validation,

[4] All the datasets are publicly available at http://jmcauley.ucsd.edu/data/amazon/links.html.

and test set in the same way as in [37, 43]. We assess the accuracy of top-N recommendation by using the following three widely used metrics: Precision (Prec, in short), Recall, and normalized discounted cumulative gain (NDCG). Prec and Recall are traditional accuracy metrics. They are used to validate whether the ground-truth items is in top-N recommendation list and computed as follows:

$$Prec@N = \frac{|Rel_u \cap N_u|}{N} \tag{4}$$

$$Recall@N = \frac{|Rel_u \cap N_u|}{Rel_u} \tag{5}$$

where Rel_u indicates the relevant observed items of user u and N_u indicates the top-N items of user u.

NDCG is a rank-sensitive metric which considers the position of the ground-truth item in the top-N recommendation list and is computed as follows:

$$NDCG@N = \frac{DCG@N}{IDCG@N}. \tag{6}$$

Additionally, $DCG@N$ in Eq. (6) is computed as follows:

$$DCG@N = \sum_{k=1}^{N} \frac{2^{y_k} - 1}{\log_2(k+1)}, \tag{7}$$

where y_k indicates the binary variable for the k-th item i_k in N_u and, if $i_k \in Rel_u$, y_k is set as 1, otherwise, y_k is set as 0. And, $IDCG@N$ in Eq. (6) stands for ***ideal DCG*** at N where, for every item i_k in N_u, y_k is set as 1. We set N to 10 and 20 for all aforementioned metrics.

Hyperparameter Settings For a fair comparison, we fine-tuned the hyperparameters of competitors and our proposed methods via grid search using the validation set. More specifically, we set the learning rate in the range {0.0001, 0.0005, 0.001, 0.005, 0.01} and the regularization weight in the range {0, 0.00001, 0.0001, 0.001, 0.01}. Also, for MMGCN [37] and LATTICE [43], we set the number of GCN-layers in the range {1, 2, 3, 4}; for LATTICE [43], we set the dropout ratio in the range {0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8}.

4.2. Experimental Results

EQ1: Do the notions of unknown and uninteresting items help improve the recommendation accuracy of multimedia recommender systems? To show the effectiveness of our proposed methods, we compared the three (original) competitors (*i.e.*, VBPR [8], MMGCN [37], LATTICE [43]) and their six variations equipped with our methods, on three datasets. Table 2 reports all the accuracy results in top-10/20 recommendation. Here, 'neg' refers to our method employing our negative sampling idea and 'mbpr' refers to our method employing multiple BPR losses. The best and the second-best recommendation accuracies on each dataset and the (original) competitor are shown in bold and underlined, respectively.

Table 2. Recommendation accuracies (%) of three state-of-the-art multimedia recommender systems and six variants, where each of our proposed methods (*i.e.*, a novel negative sampling method, neg, and multiple BPR losses, mbpr) orthogonally applied at each original method, respectively. 'gain' denotes the gains in accuracy of variants over the corresponding original method

	Amazon Baby											
	Prec@10	gain	Recall@10	gain	NDCG@10	gain	Prec@20	gain	Recall@20	gain	NDCG@20	gain
LATTICE	0.537	-	5.112	-	2.846	-	0.420	-	7.975	-	3.601	-
LATTICE-neg	0.543	1.12	5.170	1.13	**2.897**	1.79	0.426	1.43	8.098	1.54	3.669	1.89
LATTICE-mbpr	**0.554**	3.17	**5.281**	3.31	2.892	1.62	**0.433**	3.10	**8.231**	3.21	**3.670**	1.92
MMGCN	0.384	-	3.638	-	1.906	-	0.321	-	6.071	-	2.548	-
MMGCN-neg	0.384	0.00	3.653	0.41	1.931	1.31	0.316	-1.56	6.007	-1.05	2.551	0.12
MMGCN-mbpr	**0.456**	18.75	**4.334**	19.13	**2.329**	22.19	**0.364**	13.40	**6.917**	13.94	**3.011**	18.17
VBPR	0.222	-	2.102	-	1.083	-	0.177	-	3.336	-	1.469	-
VBPR-neg	0.226	1.80	2.139	1.76	1.165	7.57	0.180	1.69	3.393	1.71	1.501	2.18
VBPR-mbpr	**0.318**	43.24	**2.993**	42.39	**1.649**	52.26	**0.251**	41.81	**4.723**	41.58	**2.112**	43.77

	Amazon Men Clothing											
	Prec@10	gain	Recall@10	gain	NDCG@10	gain	Prec@20	gain	Recall@20	gain	NDCG@20	gain
LATTICE	0.415	-	4.136	-	2.194	-	0.309	-	6.160	-	2.705	-
LATTICE-neg	0.417	0.37	4.157	0.52	**2.224**	1.38	0.317	2.59	6.332	2.79	2.765	2.22
LATTICE-mbpr	**0.418**	0.67	**4.168**	0.76	**2.224**	1.38	**0.321**	4.04	**6.414**	4.12	**2.794**	3.29
MMGCN	0.270	-	2.694	-	1.328	-	0.223	-	4.447	-	1.769	-
MMGCN-neg	0.272	0.74	2.712	0.67	1.337	0.68	0.234	4.93	4.659	4.77	1.826	3.22
MMGCN-mbpr	**0.329**	21.85	**3.283**	21.86	**1.665**	25.38	**0.268**	20.18	**5.342**	20.13	**2.183**	23.40
VBPR	0.304	-	3.028	-	1.590	-	0.245	-	4.894	-	2.061	-
VBPR-neg	0.307	0.99	3.050	0.73	1.569	-1.32	0.246	0.41	4.885	-0.18	2.024	-1.80
VBPR-mbpr	**0.380**	25.00	**3.791**	25.20	**1.917**	20.57	**0.300**	22.45	**5.980**	22.19	**2.469**	19.80

	Amazon Office											
	Prec@10	gain	Recall@10	gain	NDCG@10	gain	Prec@20	gain	Recall@20	gain	NDCG@20	gain
LATTICE	1.109	-	9.213	-	5.776	-	0.839	-	13.719	-	7.137	-
LATTICE-neg	1.108	-0.13	9.215	0.02	5.783	0.12	0.842	0.36	13.739	0.15	7.158	0.29
LATTICE-mbpr	**1.134**	2.25	**9.451**	2.58	**5.854**	1.35	**0.859**	2.38	**14.061**	2.49	**7.253**	1.63
MMGCN	0.616	-	5.077	-	2.963	-	0.532	-	8.726	-	4.223	-
MMGCN-neg	0.637	3.41	5.143	1.30	3.007	1.48	0.545	2.44	8.814	1.01	4.106	-2.77
MMGCN-mbpr	**0.833**	35.23	**6.765**	33.25	**4.075**	37.53	**0.674**	26.69	**10.923**	25.18	**5.333**	26.28
VBPR	0.699	-	5.717	-	3.524	-	0.558	-	9.072	-	4.544	-
VBPR-neg	0.694	-0.72	5.655	-1.08	3.430	-2.67	0.558	0.00	9.088	0.18	4.465	-1.74
VBPR-mbpr	**0.792**	13.30	**6.366**	11.35	**3.932**	11.58	**0.623**	11.65	**9.874**	8.84	**5.029**	10.67

In Table 2, we can see that the variations with the multiple BPR losses show the superiority over the original ones and those with our negative sampling, in all datasets and all models. Specifically, in the case of LATTICE [43], as it is the most recent and best performing method, its variation applied with our negative sampling outperforms the original method by up to 2.79% (see Amazon Men Clothing) and the variation applied with the multiple BPR losses outperforms the original method by up to 4.12% (see Amazon Men Clothing), both in terms of Recall@20. In the case of MMGCN [37], the variation applied

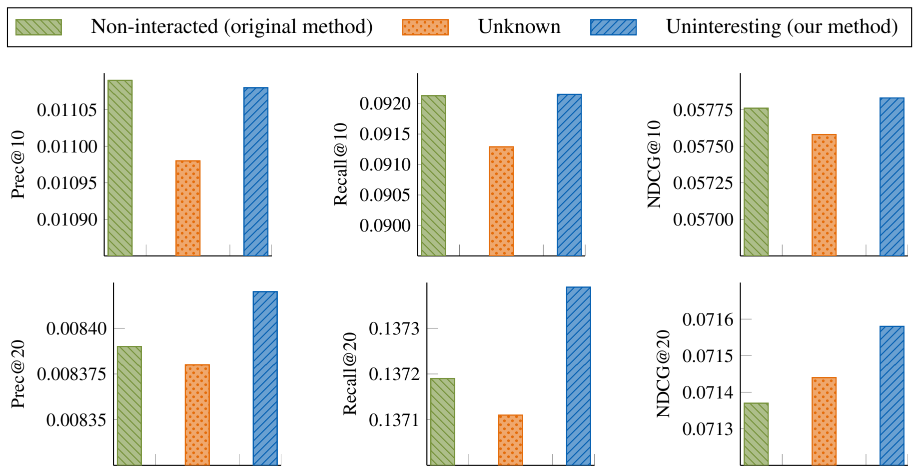

Fig. 3. Recommendation accuracies of a best state-of-the-art multimedia recommender system (*i.e.*, LATTICE [43]) and two variants equipped with two different cases of negative sampling methods, respectively. 'Random' indicates the method of using randomly chosen non-interacted items as negative samples (*i.e.*, the original negative sampling method), 'Unknown' indicates the method of using unknown items as negative samples, and 'Uninteresting' indicates the method of using uninteresting items as negative samples (*i.e.*, our proposed negative sampling method)

with our negative sampling outperforms the original method by up to 4.93% in terms of Prec@20 (see Amazon Men Clothing) and the variation applied with our multiple BPR losses outperforms the original method by up to 37.53% in terms of NDCG@10 (see Amazon Office). Lastly, in the case of VBPR [8], the variation applied with our negative sampling outperforms the original method by up to 7.53% in terms of NDCG@10 (see Amazon Baby) and the variation applied with the multiple BPR losses outperforms the original method by up to 52.26% in terms of NDCG@10 (see Amazon Baby).

Based on the results above, we have confirmed that i) employing the notions of unknown and uninteresting items (instead of the non-interacted items) in training the model is effective in terms of recommendation accuracy and ii) employing multiple BPR losses over interesting, unknown, and uninteresting items is more effective than a single BPR loss over interesting and non-interacted items in terms of recommendation accuracy in training the model.

EQ2: Is the idea of selecting the uninteresting items as negative samples most effective for improving the recommendation accuracy? To verify the effectiveness of our negative sampling method, we compare the recommendation accuracy of the following three cases: sampling negative items randomly i) from the non-interacted items (*i.e.*, original method), ii) from unknown items, and iii) from uninteresting items (*i.e.*, our proposed method). Figure 3 shows the recommendation accuracy of the three negative sam-

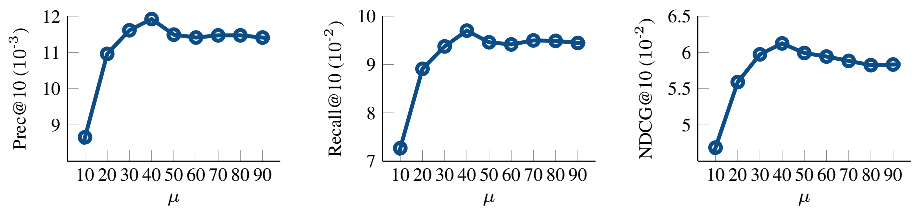

Fig. 4. The effect of μ on recommendation accuracies

pling methods on the Amazon Office dataset with LATTICE [43].[5] Here, 'Non-interacted' refers to case i), 'Unknown' refers to case ii), and 'Uninteresting' refers to case iii).

In Figure 3, we see that our method (*i.e.*, only using uninteresting items) outperforms the original negative sampling method. The method of using unknown items as negative samples shows very poor recommendation accuracy. This indicates that, as unknown items could be the items that the users' preferences are high, they should not be used in the training process as negative samples. When randomly sampling the users' non-interacted items as negative samples, unknown items might be included as negative samples, thus likely to confuse the model in training. Therefore, this result validates that selecting negative samples from uninteresting items helps improve the accuracy in multimedia recommendation.

EQ3: How sensitive is the recommendation accuracy of the multiple BPR losses to different hyperparameter values? For our multiple BPR losses, we consider two types of hyperparameters. First, μ is to determine the ratio of uninteresting items to non-interacted items. Second, the weights α, β, and γ for different BPR losses to indicate the importance in training. Regarding the hyperparameters, we conducted experiments to answer the following two sub-questions:

- **EQ3-1**: How sensitive is the accuracy from employing the multiple BPR losses to the ratios of uninteresting and unknown items out of non-interacted items?
- **EQ3-2**: How sensitive is the accuracy from employing the multiple BPR losses to the weight for each BPR loss?

EQ3-1: Sensitiveness of hyperparameter μ. We analyze how the recommendation accuracy changes with different values of $\mu \in \{10, 20, 30, 40, 50, 60, 70, 80, 90\}$ on the Amazon Office dataset with LATTICE [43]. Figure 4 shows the recommendation accuracy with different values of μ. As shown in Figure 4, we observe that the recommendation accuracy increases until μ increases to 40 and then decreases. The result shows that, if μ is set as too small (resp. large), some uninteresting (resp. unknown) items might be misclassified as unknown (resp. uninteresting) items, which causes the model to be confused in the training process. Therefore, the proper setting of μ allows the model to be better learned and provides a more-effective recommendation result. Based on this observation, we set μ as 40% for our proposed multiple BPR losses.

[5] For EQ2 and EQ3, the tendencies of recommendation accuracy on other datasets with other competitors are all similar; so, we only include the results on Amazon Office with LATTICE, the latest and most powerful method.

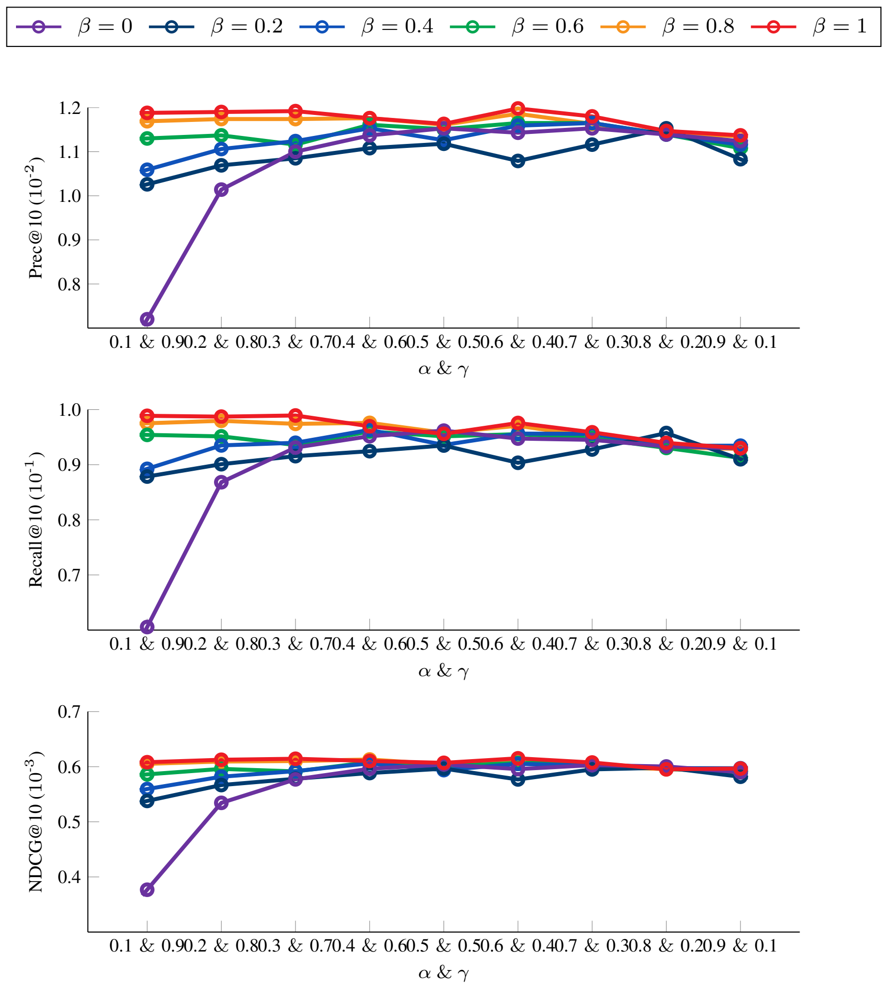

Fig. 5. The effect of α, β, and γ on recommendation accuracies

EQ3-2: Sensitiveness of hyperparameters α, β, and γ. We analyze the change of recommendation accuracy with varying the values of α, β, and $\gamma \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$ on the Amazon Office dataset with LATTICE [43]. Figure 5 illustrates the recommendation accuracy with different values of α, β, and γ. The recommendation accuracy becomes highest when $\alpha = 0.4, \beta = 1.0,$, and $\gamma = 0.6$ in terms of Prec@10. The result shows that bigger the value of β, higher the recommendation accuracy regardless of the values of α and γ. Also, the result shows that the recommendation accuracy overall shows robustness regardless of the values of α and γ. Therefore, based on this result, we set $\alpha = 0.4, \beta = 1.0$, and $\gamma = 0.6$, in the previous experiments.

The experimental results can be summarized as follows: i) applying concept(s) of unknown and uninteresting items helps to improve the recommendation accuracy of multimedia recommender systems; ii) selecting the uninteresting items as negative samples is more effective in improving the recommendation accuracy than selecting random (original negative sampling method) or unknown items; iii) utilizing both unknown and uninteresting items (*i.e.*, all non-interacted items) in multimedia recommender systems significantly improves most of their original recommendation accuracies, also this method (*i.e.*, our method) is easily and orthogonally applicable to any multimedia recommender systems.

5. Conclusions

In this paper, we have pointed out the limitation of existing multimedia recommender systems that they do not fully exploit the characteristics of non-interacted items for users. Then, we proposed two methods to alleviate the limitation, thereby enabling existing systems to exploit the non-interacted items of users appropriately by classifying the non-interacted items into unknown and uninteresting items. Specifically, our first idea is to allow only the items highly likely not to be preferred by a user as negative items during training the recommender model. Further, our second idea is to use the multiple BPR losses, which makes possible rank discrepancies among positive, unknown, and uninteresting items learned correctly in the training process. Extensive experiments on three real-world Amazon datasets validate that our proposed methods outperform three state-of-the-art multimedia recommender systems and that our ideas are all effective in improving recommendation accuracy.

Acknowledgments. This research was supported by Culture, Sports and Tourism R&D Program through the Korea Creative Content Agency grant funded by the Ministry of Culture, Sports and Tourism in 2022 (Project Name: Development of Intelligent Personalized Rehabilitation Service Technology, Project Number: SR202104001, Contribution Rate: 100%).

References

1. Albawi, S., Mohammed, T.A., Al-Zawi, S.: Understanding of a convolutional neural network. In: International Conference on Engineering and Technology (ICET). pp. 1–6 (2017)
2. Bao, Y., Fang, H., Zhang, J.: Topicmf: Simultaneously exploiting ratings and reviews for recommendation. In: AAAI. pp. 2–8 (2014)

3. Chae, D., Kim, J., Chau, D.H., Kim, S.: AR-CF: Augmenting virtual users and items in collaborative filtering for addressing cold-start problems. In: ACM SIGIR. pp. 1251–1260 (2020)
4. Chen, X., Chen, H., Xu, H., Zhang, Y., Cao, Y., Qin, Z., Zha, H.: Personalized fashion recommendation with visual explanations based on multimodal attention network: Towards visually explainable recommendation. In: ACM SIGIR. pp. 765–774 (2019)
5. Cohen, R., Shalom, O.S., Jannach, D., Amir, A.: A black-box attack model for visually-aware recommender systems. In: ACM WSDM. pp. 94–102 (2021)
6. Goldberg, D., Nichols, D.A., Oki, B.M., Terry, D.B.: Using collaborative filtering to weave an information tapestry. Commun. ACM 35(12), 61–70 (1992)
7. Graves, A., Mohamed, A., Hinton, G.E.: Speech recognition with deep recurrent neural networks. In: IEEE ICASSP. pp. 6645–6649 (2013)
8. He, R., McAuley, J.: VBPR: visual bayesian personalized ranking from implicit feedback. In: AAAI. pp. 144–150 (2016)
9. He, X., Du, X., Wang, X., Tian, F., Tang, J., Chua, T.: Outer product-based neural collaborative filtering. In: IJCAI. pp. 2227–2233 (2018)
10. Hidasi, B., Karatzoglou, A., Baltrunas, L., Tikk, D.: Session-based recommendations with recurrent neural networks. CoRR abs/1511.06939 (2015)
11. Hwang, W., Parc, J., Kim, S., Lee, J., Lee, D.: "Told you I didn't like it": Exploiting uninteresting items for effective collaborative filtering. In: IEEE ICDE. pp. 349–360 (2016)
12. Kim, D.H., Park, C., Oh, J., Lee, S., Yu, H.: Convolutional matrix factorization for document context-aware recommendation. In: ACM RecSys. pp. 233–240 (2016)
13. Kim, T., Kim, Y., Lee, Y.C., Shin, W., Kim, S.: Is it enough just looking at the title?: Leveraging body text to encrich title words towards accurate news recommendation. In: ACM CIKM. pp. 4138–4142 (2022)
14. Kim, T., Lee, Y.C., Shin, K., Kim, S.: Mario:Modality-Aware attention and modality-preserving decoders for multimedia recommendation. In: ACM CIKM. pp. 993–1002 (2022)
15. Ko, Y., Yu, J.S., Bae, H.K., Park, Y., Lee, D., Kim, S.W.: Mascot: A quantization framework for efficient matrix factorization in recommender systems. In: IEEE ICDM. pp. 290–299 (2021)
16. Kong, T., Kim, T., Jeon, J., Choi, J., Lee, Y.C., Park, N., Kim, S.W.: Linear, or non-linear, that is the question! In: ACM WSDM. pp. 517–525 (2022)
17. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. Commun. ACM 60(6), 84–90 (2017)
18. Lee, J., Hwang, W., Parc, J., Lee, Y., Kim, S., Lee, D.: l-injection: Toward effective collaborative filtering using uninteresting items. IEEE TKDE 31(1), 3–16 (2019)
19. Lee, Y., Kim, S., Lee, D.: gOCCF: Graph-theoretic one-class collaborative filtering based on uninteresting items. In: AAAI. pp. 3448–3456 (2018)
20. Lee, Y.C., Kim, T., Choi, J., He, X., Kim, S.: M-bpr: A novel approach to improving bpr for recommendation with multi-type pair-wise preferences. Information Sciences 547, 255–270 (2021)
21. Lee, Y., Kim, S., Park, S., Xie, X.: How to impute missing ratings?: Claims, solution, and its application to collaborative filtering. In: WWW. pp. 783–792 (2018)
22. Lim, H., Lee, Y.C., Lee, J.S., Han, S., Kim, S., Jeong, Y., Kim, C., Kim, J., Han, S., Choi, S., Ko, H., Lee, D., Choi, J., Kim, Y., Bae, H.K., Kim, T., Ahn, J., You, H.S., Kim, S.W.: Airs: A large-scale recommender system at naver news. In: IEEE ICDE. pp. 3386–3398 (2022)
23. Liu, F., Cheng, Z., Sun, C., Wang, Y., Nie, L., Kankanhalli, M.: User diverse preference modeling by multimodal attentive metric learning. In: ACM MM. pp. 1526–1534 (2019)
24. Liu, Q., Wu, S., Wang, L.: Deepstyle: Learning user preferences for visual recommendation. In: ACM SIGIR. pp. 841–844 (2017)
25. Niu, W., Caverlee, J., Lu, H.: Neural personalized ranking for image recommendation. In: ACM WSDM. pp. 423–431 (2018)
26. O'Shea, K., Nash, R.: An introduction to convolutional neural networks. arXiv preprint arXiv:1511.08458 (2015)

27. Pan, R., Zhou, Y., Cao, B., Liu, N.N., Lukose, R.M., Scholz, M., Yang, Q.: One-class collaborative filtering. In: IEEE ICDM. pp. 502–511 (2008)
28. Park, S.J., Chae, D.K., Bae, H.K., Park, S., Kim, S.W.: Reinforcement learning over sentiment-augmented knowledge graphs towards accurate and explainable recommendation. In: ACM WSDM. pp. 784–793 (2022)
29. Reimers, N., Gurevych, I.: Sentence-BERT: Sentence embeddings using siamese bert-networks. arXiv preprint arXiv:1908.10084 (2019)
30. Rendle, S., Freudenthaler, C., Gantner, Z., Schmidt-Thieme, L.: BPR: Bayesian personalized ranking from implicit feedback. In: UAI. pp. 452–461 (2009)
31. Ricci, F., Rokach, L., Shapira, B. (eds.): Recommender Systems Handbook. Springer (2015)
32. Sarwar, B.M., Karypis, G., Konstan, J.A., Riedl, J.: Item-based collaborative filtering recommendation algorithms. pp. 285–295 (2001)
33. Sedhain, S., Menon, A.K., Sanner, S., Xie, L.: Autorec: Autoencoders meet collaborative filtering. In: WWW. pp. 111–112 (2015)
34. Tang, J., Wang, K.: Personalized top-N sequential recommendation via convolutional sequence embedding. In: ACM WSDM. pp. 565–573 (2018)
35. Wang, S., Wang, Y., Tang, J., Shu, K., Ranganath, S., Liu, H.: What your images reveal: Exploiting visual contents for point-of-interest recommendation. In: WWW. pp. 391–400 (2017)
36. Wei, Y., Wang, X., Nie, L., He, X., Chua, T.: Graph-refined convolutional network for multimedia recommendation with implicit feedback. In: ACM MM. pp. 3541–3549 (2020)
37. Wei, Y., Wang, X., Nie, L., He, X., Hong, R., Chua, T.: MMGCN: Multi-modal graph convolution network for personalized recommendation of micro-video. In: ACM MM. pp. 1437–1445 (2019)
38. Wu, C., Ahmed, A., Beutel, A., Smola, A.J., Jing, H.: Recurrent recommender networks. In: ACM WSDM. pp. 495–503 (2017)
39. Xue, H., Dai, X., Zhang, J., Huang, S., Chen, J.: Deep matrix factorization models for recommender systems. In: IJCAI. pp. 3203–3209 (2017)
40. Yao, W., He, J., Wang, H., Zhang, Y., Cao, J.: Collaborative topic ranking: Leveraging item meta-data for sparsity reduction. In: AAAI. pp. 374–380 (2015)
41. Ying, H., Chen, L., Xiong, Y., Wu, J.: Collaborative deep ranking: A hybrid pair-wise recommendation algorithm with implicit feedback. In: PAKDD. pp. 555–567 (2016)
42. Zagoruyko, S., Komodakis, N.: Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer. arXiv preprint arXiv:1612.03928 (2016)
43. Zhang, J., Zhu, Y., Liu, O., Wu, S., Wang, S., Wang, L.: Mining latent structures for multimedia recommendation. In: ACM MM. pp. 3872–3880 (2021)
44. Zhang, Y., Ai, Q., Chen, X., Croft, W.B.: Joint representation learning for top-n recommendation with heterogeneous information sources. In: ACM CIKM. pp. 1449–1458 (2017)

Jiyeon Kim received her bachelor's degree in International studies (B.A.) and Computer Science (B.S.) from Hanyang university, Seoul, Korea in 2020. She received her master's degree in computer science from Hanyang University, Seoul, Korea, in 2022. Her research interests include data mining, recommender systems and machine learning.

Taeri Kim received a B.S. degree in Multimedia Engineering from Hansung University, Seoul, Korea, in 2015. Since 2017, she has been pursuing a Ph.D. degree in Computer Science at Hanyang University. Her current research interests include data mining, graph neural network analysis, and recommendation.

Sang-Wook Kim received the B.S. degree in computer engineering from Seoul National University, in 1989, and the M.S. and Ph.D. degrees in computer science from the Korea

Advanced Institute of Science and Technology (KAIST), in 1991 and 1994, respectively. From 1995 to 2003, he served as an associate professor with Kangwon National University. In 2003, he joined Hanyang University, Seoul, Korea, where he currently is a professor in the Department of Computer Science and the director of the BrainKorea-21-FOUR research program. He is also leading a National Research Lab (NRL) Project funded by the National Research Foundation since 2015. From 2009 to 2010, he visited the Computer Science Department, Carnegie Mellon University, as a visiting professor. From 1999 to 2000, he worked with the IBM T. J. Watson Research Center, USA, as a postdoc. He also visited the Computer Science Department of Stanford University as a visiting researcher in 1991. He is an author of more than 200 papers in refereed international journals and international conference proceedings. His research interests include databases, data mining, multimedia information retrieval, social network analysis, recommendation, and web data analysis. He is a member of the ACM and the IEEE.

Received: October 31, 2022; Accepted: December 29, 2022.