

Article

Knowledge-Aware Multispace Embedding Learning for Personalized Recommendation

Meng Jian ¹ , Chenlin Zhang ¹, Xin Fu ², Lifang Wu ^{1,*} and Zhangquan Wang ³

¹ Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China; mjian@bjut.edu.cn (M.J.); zhangchenlin@emails.bjut.edu.cn (C.Z.)

² School of Water Conservancy and Environment, University of Jinan, Jinan 250022, China; stu_fux@ujn.edu.cn

³ Inner Mongolia Aerospace Power Machinery Testing Institute, Huhhot 010076, China; zhangquanw@163.com

* Correspondence: lfwu@bjut.edu.cn

Abstract: Recommender systems help users filter items they may be interested in from massive multimedia content to alleviate information overload. Collaborative filtering-based models perform recommendation relying on users' historical interactions, which meets great difficulty in modeling users' interests with extremely sparse interactions. Fortunately, the rich semantics hidden in items may be promising in helping to describing users' interests. In this work, we explore the semantic correlations between items on modeling users' interests and propose knowledge-aware multispace embedding learning (KMEL) for personalized recommendation. KMEL attempts to model users' interests across semantic structures to leverage valuable knowledge. High-order semantic collaborative signals are extracted in multiple independent semantic spaces and aggregated to describe users' interests in each specific semantic. The semantic embeddings are adaptively integrated with a target-aware attention mechanism to learn cross-space multisemantic embeddings for users and items, which are fed to the subsequent pairwise interaction layer for personalized recommendation. Experiments on real-world datasets demonstrate the effectiveness of the proposed KMEL model.

Keywords: collaborative filtering; user interest; knowledge graph; recommender system



Citation: Jian, M.; Zhang, C.; Fu, X.; Wu, L.; Wang, Z. Knowledge-Aware Multispace Embedding Learning for Personalized Recommendation. *Sensors* **2022**, *22*, 2212. <https://doi.org/10.3390/s22062212>

Academic Editor: Alberto Gotta

Received: 9 February 2022

Accepted: 9 March 2022

Published: 12 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the era of big data, people are surrounded by ubiquitous information. This has created the urgent need to filter the content that users require from the massive amount of available content. Personalized recommender systems help users find candidate content to meet users' requirements, thereby alleviating the problem of information overload. Recommender systems have been widely applied in personalized music radio, e-commerce, multimedia platforms and other fields. As a core strategy, collaborative filtering refers to users' historical interactions to predict users' interests for personalized recommendation. However, it is difficult to capture their personalized preferences relying only on extremely sparse historical interactions. Personalized recommendation meets a significant challenge from sparsity, and cold start issues [1].

Fortunately, abundant semantic correlations exist among items that explain knowledge clues on users' interests. For example, in the online movie application domain, the semantic relationship between both director and actor are crucial for capturing users' interests. It is intuitive to involve semantic relations upon user-item interactions to alleviate the performance limitations from data sparsity and cold start issues. Some current studies [2–5] embed semantic associations in interest mining, which performs effectively for capturing users' interests. Their success inspires us to think about the role of rich semantic correlations on recommendation tasks. As a hot topic, semantic correlation extraction has been actively studied in the field of knowledge graphs [6,7]. A knowledge graph, as an auxiliary structure, provides additional semantic correlations among items. Knowledge graphs also play a role to assist matching pairwise user-items by naturally integrating

semantic correlation between items [8–11] strive to model users' interests with high-order user–item semantic associations on the knowledge graph. Since knowledge graphs include huge-scale entities and diverse semantic relations, high-order mining may get lost in semantics for personalized recommendation; however, the semantic clue associated with the target user–item pair is worth noting.

We argue that the semantics in knowledge graphs make more sense than the large-scale entities on users' interests. In this work, we investigate semantic correlations among items in view of multispace learning and propose knowledge-aware multispace embedding learning (KMEL) for personalized recommendation. We extend the user–item interaction graph with a specific semantic relation among items in a knowledge graph, resulting a semantic-specific user–item–item hybrid graph. With multiple semantics in the knowledge graph, the proposed KMEL derives multiple semantic spaces. In each semantic space, items are connected with the corresponding semantic relation; therefore, the connections vary in different semantic spaces. Due to the independence of semantics, it is intuitive to explore the embedding compensation among the semantic spaces, including the interaction space. KMEL attempts to model users' interests across semantic structures to leverage the valuable knowledge information. It extracts the positive impact of semantics from multiple independent semantic spaces. Specifically, the high-order semantic correlation between items is built respectively in multiple independent semantic spaces, aggregated by propagating semantic correlation to embed users' interests in each semantic space. Then, users' interests in multiple semantic domains are adaptively integrated as a whole to recommend items. The contributions of this work are summarized as follows.

- We leverage diverse semantic correlations among items to compensate users' sparse interaction records to mine users' interests for recommendations;
- We propose a knowledge-aware multispace embedding learning model that respects users' interests on each semantic and learns user embedding in a manner of divide-and-conquer across multiple semantic spaces;
- We demonstrate the effectiveness of the proposed KMEL model by experiments and corresponding analysis on two real-world datasets.

2. Related Work

The proposed KMEL model is related to collaborative filtering, graph-based recommendation, and knowledge-aware recommendation models. We review the literature and highlight the difference to the proposed KMEL model.

2.1. Collaborative Filtering

Collaborative filtering [12–14] as a core strategy has been widely applied in recommendation scenarios due to its simple and practical effectiveness. CF-based recommendation models exploit users' interaction records to mine users' interests and recommend items to users having similar interests to their owners, which do not require additional prior knowledge for personalized recommendation. Matrix factorization (MF) [12] pioneers a learnable CF model that performs recommendation by interacting user and item embeddings. MF is further developed by involving deep architecture in embedding learning [15–18] and interaction function [19]. Despite its effectiveness, CF models meet obstacles in capturing users' interests since the personalized recommendation faces severe data sparsity and cold-start problems. In order to alleviate the problem, related studies [20–22] introduce diverse side information to enrich the clue for interest mining, such as social networks [20], knowledge graphs [2,10], and item content information [22], aiming at enhancing the semantic association when interacting between users and items. This work emphasizes the role of higher-order semantic associations between items and strives to leverage semantic knowledge to promote interest mining and personalized recommendation.

2.2. Graph-Based Recommendation

Subsequent studies figure out that rich collaborative signals remain in users' interactions which naturally forms a heterogeneous graph with user and item nodes. With this hypothesis, graph-based recommendation models [15–17] propagate user and item characters layer-by-layer on the interaction graph and naturally integrate higher-order collaborative signals to model users' interest. GCMC [15] models the effect of interactions on users and items using a graph convolutional layer on the interaction graph to encode their embeddings. PinSage [16] applies multiple graph convolutional layers on the graph to perform image recommendation. DGCF [18] builds the fine-grained user–item relationship concerning users' intents to perform graph convolution for recommendation. NGCF [17] recursively performs information propagation on the user–item interaction graph to extract higher-order collaborative signals hidden in interactions for personalized recommendation. A light graph convolution [23] is proposed on interaction graph to model users' interests by simplifying transformation, nonlinear activation in graph learning. Wu et al. [24] explored a context-aware graph convolution on the user–item interaction graph to digest the collaborative signals among users, items, and contexts into interaction estimation. These models aim to use heterogeneous collaborative signals hidden in users' interactions as much as possible to embed users and items for personalized matching. Inspired by their success, this work strives to further investigate users' interaction graph with an additional knowledge graph to deal with the interaction sparsity issue on capturing users' interests.

2.3. Knowledge-Aware Recommendation

The rich semantics among items have been actively explored to mitigate cold-start and data sparsity issues. Knowledge graph, as a piece of standard auxiliary information, is introduced to bridge the semantic gap among interactions [10,11,25]. Ripplenet [11] propagates users' interests along with the high-order paths on the user–item–entity graph by extending heterogeneous interaction graph with a knowledge graph to model semantic collaborative signals into embeddings of users and items. It employs user-interacted items as seed nodes to propagate users' interests layer-by-layer on the knowledge graph to learn users' embeddings. KGAT [10] utilizes a knowledge-aware neighbor-aggregation mechanism to encode user interests and introduces an attention mechanism to aggregate neighbors adaptively. Zhu et al. [26] built history interest and potential intent respectively from users' clicked sequence and path connectivity in KGs to jointly embed users. It investigates signals from both item space and their connection space of KG. Since knowledge graph takes hybrid semantics, the diverse knowledge hidden in the graph is required to be addressed elaborately. High-order mining is easily hindered by complex semantics for personalized recommendation. The role of semantics rather than high-order user-item semantic associations on the knowledge graph should be taken into account. AIMN [25] performs pairwise user-item matching in multiple knowledge-aware attribute-level and merges the attribute-level interactions to the final score for personalized ranking. Huang et al. [27] explored multityped user–item interactive patterns with coupled graph learning on both social-aware user space and knowledge-aware item space. Users are built by the interacted items with the specific pattern and items are aggregated on all the related interactive patterns. Differently, this work strives to mine users' interests from interaction space and multiple semantic spaces of independent relations in KG for personalized recommendation.

3. Methodology

Though there are redundant semantic correlations among items, involving item semantics shows promise as a way of explaining users' interests. Therefore, we attempt to investigate semantic correlations among items to compensate interactions and model users' interests. Mining users' interests is implemented by embedding users relying on given historical records. As a promising auxiliary reference, an existing knowledge graph provides abundant semantic relations that relate to items' attributes. It provides a valuable clue to

reveal the interests of users who have interacted with them. In this work, we construct interest mining across semantic structures in view of multispace learning and propose knowledge-aware multispace embedding learning (KMEL) for personalized recommendation. It extracts collaborative signals over users' high-order interaction paths in multiple independent semantic spaces and comprehensively models users and items using a siamese attention mechanism. Figure 1 shows the framework of the proposed KMEL model for personalized recommendation. Specifically, the high-order semantic correlation between items is extracted in multiple independent semantic spaces and aggregated by propagating semantic correlation to embed users' interests by graph convolution in each semantic space. Then, users' interests in multiple semantic domains are adaptively integrated as a whole to interact with items for recommendation. This section describes collaborative embedding learning on the user-item interaction graph, semantic embedding learning on the user-item-item graph in multiple independent semantic spaces, cross-space multisemantic fusion based on target-aware attention, and nonlinear interaction and recommendation as follows.

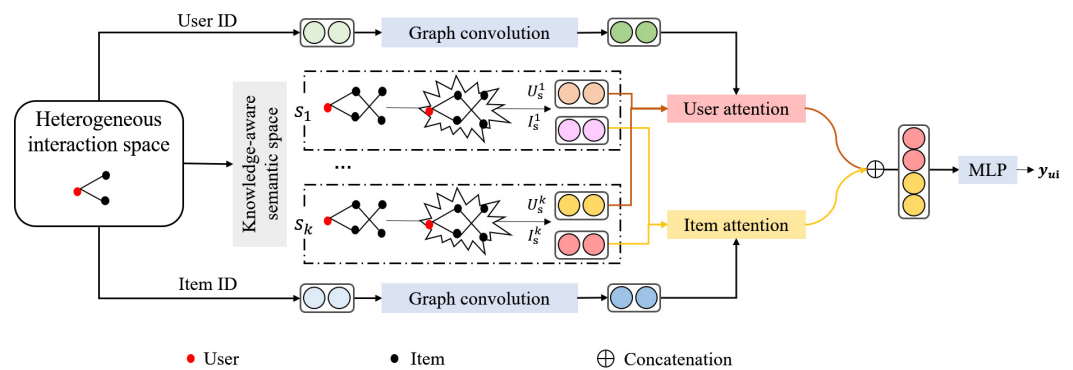


Figure 1. Framework of the proposed knowledge-aware multispace embedding learning (KMEL) for personalized recommendation.

3.1. Collaborative Embedding Learning

The items that a user has interacted with provide evidence of the user's personalized interests, while the users interacting with a specific item reflect the item's audience character. The user-item interactions form a heterogeneous graph that contains rich collaborative signals. It is promising to model users and items by graph propagation for personalized recommendation. As illustrated in Figure 1, We capture the positive impact of high-order collaborative signals using a layer-by-layer interest propagation mechanism on the heterogeneous user-item interaction graph. Considering that users have varying interest degrees for items, it is necessary to take varying contributions of neighbors in building users and items with graph propagation. We adopt an attention-based graph convolutional layer to encode users and items by performing neighbour aggregation.

Following current mainstream recommendation strategies, we initialize user ID and item ID with high-dimensional embeddings $e_u^{(0)}, e_i^{(0)} \in R_d$, where R_d is a d-dimensional interest embedding space. The collaborative interest propagation from interacted item i to user u is defined as

$$m_{u \leftarrow i}^{(1)} = f(e_u^{(0)}, e_i^{(0)}, \alpha_{ui}) \quad (1)$$

where $m_{u \leftarrow i}$ is the collaborative signals propagated from item i to user u , $f()$ represents the interest propagation function, and α_{ui} denotes the learnable contribution parameter of item i to user u . The propagation function $f()$ is defined as

$$m_{u \leftarrow i}^{(1)} = \alpha_{ui} We_i^{(0)}, \quad \alpha_{ui} = \frac{\exp(\sigma(\text{att}^T [We_u^{(0)} \oplus We_i^{(0)}]))}{\sum_{k \in N_u} \exp(\sigma(\text{att}^T [We_u^{(0)} \oplus We_k^{(0)}]))} \quad (2)$$

where α_{ui} is learned by a single-layer attention network with parameter att . The propagation layer performs a softmax normalization to produce the neighbor contribution degree α_{ui} for the following embedding aggregation. W is a learnable linear transformation matrix, $\oplus$ represents concatenation operation, and $\sigma(\cdot)$ is the LeakyReLU activation function. The propagated collaborative signals from interacted items $i \in N_u$ of user u are aggregated to update its embedding as

$$e_u^{(1)} = \sigma(m_{u \leftarrow u}^{(1)} + \sum_{i \in N_u} m_{u \leftarrow i}^{(1)}) \quad (3)$$

where $e_u^{(1)}$ represents the first-order collaborative embedding of user u . We stack multiple propagation layers to model higher-order collaborative embedding for user u . The higher-order collaborative embedding attempts to capture the effect of long-path user-item collaborative signals hidden in users' interactions on user interest modeling. We stack L interest propagation layers to propagate collaborative signals for user u in its L -hop neighbors.

$$e_u^{(l)} = \sigma(m_{u \leftarrow u}^{(l)} + \sum_{i \in N_u} m_{u \leftarrow i}^{(l)}) \quad (4)$$

where $l = 1, 2, \dots, L$ indexes the embedding on layer l . We perform the same operation on item i to derive the higher-order collaborative embedding of item i correspondingly as

$$e_i^{(l)} = \sigma(m_{i \leftarrow i}^{(l)} + \sum_{u \in N_i} m_{i \leftarrow u}^{(l)}) \quad (5)$$

The collaborative embeddings of users and items on the L th layer of the interest propagation network are represented as U_c and I_c . The collaborative embedding reflects the metainterest of users and items, i.e., behavior of users with similar interests and audience of items with similar attributes. We utilize the collaborative embedding U_c and I_c to guide the subsequent modeling of semantic interest embedding of users and items.

3.2. Semantic Embedding Learning

This work investigates semantic correlations among items as a knowledge clue to aid users' interactions and mine users' interests for personalized recommendation. Considering the complex and diverse semantics in the knowledge graph, the proposed KMEL model constructs users' interests across semantic spaces to leverage valuable knowledge information. It extracts semantic positive effects from multiple independent semantic spaces $s_1, s_2, \dots, s_K$. We align the interaction graph to the corresponding knowledge graph forming a hybrid graph and explore the user-item-item semantic structure in each independent semantic space to capture a semantic embedding for users and items.

In each semantic space s_k , we extract semantic collaborative signals in the hybrid graph to model users and items on the specific semantic domain. Since the items located in the same semantic space appear to be relatively compact in terms of semantics, we employ a naive graph convolution to aggregate neighbors in the hybrid graph to learn the semantic embeddings for users and items. On the hybrid graph, embeddings are also initialized with the same ID in Section 3.1, as $h_u^{(0)} = e_u^{(0)}$, $h_i^{(0)} = e_i^{(0)}$. The semantic interest propagation is defined as

$$h_u^{(l+1)} = \sigma(\sum_{i \in N(u)} \frac{1}{c_{iu}} h_i^{(l)} W^{(l)} + b^{(l)}) \quad (6)$$

$$h_i^{(l+1)} = \sigma(\sum_{j \in N(i)} \frac{1}{c_{ji}} h_j^{(l)} W^{(l)} + \sum_{u \in N(i)} \frac{1}{c_{ui}} h_u^{(l)} W^{(l)} + b^{(l)}) \quad (7)$$

where $h_u^{(l+1)}$ and $h_i^{(l+1)}$ are the updated semantic embeddings of user u and item i at layer $l+1$ in semantic space k , respectively, $N(u)$, $N(i)$ represents the neighbor set of user u and item i , respectively, $b^{(l)}$ represents the bias coefficient of the l th layer propagation network, c_{ji} is the product of the square root of node degrees $c_{ji} = \sqrt{|\mathcal{N}(j)|} \sqrt{|\mathcal{N}(i)|}$. On each layer,

interest propagation aggregates neighbor users $u \in \mathcal{N}(i)$ from the interaction graph and neighbor items $j \in \mathcal{N}(i)$ from the knowledge graph to update embedding h_i for item i . For user u , the propagation aggregates the neighbor items $j \in \mathcal{N}(u)$ the user interacts with to update the embedding h_u .

In each semantic space s_k , we capture high-order semantic collaborative signals between user-items by performing multi-layer interest propagation on the hybrid graph. We use U_s^k and I_s^k to represent the semantic embeddings of user u and item i on the L th layer in the semantic space s_k , which reveals the corresponding semantic preference of users and attribute distribution of items. Semantic embeddings from different semantic spaces reflect the diverse semantic preferences of users. With the same semantic interest propagation on the semantic-specific hybrid graph structures, the proposed KMEL derives a set of knowledge-aware semantic embeddings $U_s^1, U_s^2, \dots, U_s^k$ for users and $I_s^1, I_s^2, \dots, I_s^k$ for items.

3.3. Cross-Space Multisemantic Fusion

Till now, the proposed KMEL builds collaborative embeddings U_c and I_c , and semantic embeddings $U_s^1, U_s^2, \dots, U_s^k$ and $I_s^1, I_s^2, \dots, I_s^k$. Multiple semantic embeddings capture users' /items' preferences/attributes in different semantic spaces. As illustrated in Figure 1, we utilize the collaborative embedding U_c and I_c to guide the cross-space embedding fusion to learn an integrated embedding for users and items.

Considering that the various importance of these semantics for modeling users' interests varies from user to user, we introduce a target-aware attention mechanism to learn the contribution of different semantics for modeling users' interests, referred as importance degree.

$$\alpha_k^u = \text{softmax}(\text{ReLU}(W_{att}(U_c \oplus U_s^k) + b_{att})) \quad (8)$$

$$\alpha_k^i = \text{softmax}(\text{ReLU}(W_{att}(I_c \oplus I_s^k) + b_{att})) \quad (9)$$

where α_k^u represents the importance of user u 's semantic embedding U_s^k in semantic space s_k on modeling his/her complete interests. Similarly, α_k^i represents the importance of item i 's semantic embedding I_s^k in semantic space s_k to model its entire attribute. W_{att} and b_{att} are the learnable weights and bias coefficients of a single-layer attention network. We perform a *softmax* operation to normalize the importance degree of semantics. The collaborative embeddings U_c and I_c of users/items participate in the importance estimation of embeddings in other semantic spaces with their semantic-specific embeddings $U_s^1, U_s^2, \dots, U_s^k$ and $I_s^1, I_s^2, \dots, I_s^k$, resulting in varying importance degrees on semantics. Then, we aggregate the semantic embeddings across multiple spaces to construct an integrated embedding for users and items.

$$h_u = \sum_{U_s^k \in \{U_c, U_s^1, U_s^2, \dots, U_s^m\}} \alpha_k^u U_s^k \quad (10)$$

$$h_i = \sum_{I_s^k \in \{I_c, I_s^1, I_s^2, \dots, I_s^m\}} \alpha_k^i I_s^k \quad (11)$$

where h_u and h_i represent the integrated embeddings of user u and item i across multiple semantic spaces.

3.4. Nonlinear Interaction and Recommendation

For the pairwise matching of user u and item i , we concatenate their embeddings h_u and h_i as the pairwise interaction feature. The interaction feature contains matching information between user preferences and item attributes. We leverage a typical MLP

network to filter further the nonlinear interaction correlation between user u and item i and predict the interest degree of user u on item i .

$$\hat{y}_{ui} = MLP(h_u \oplus h_i) \quad (12)$$

where $\hat{y}_{ui}$ represents the predicted interest degree of user u on item i , $MLP()$ is a standard nonlinear interaction function. The proposed KMEL ranks the items i with the highest interest degree and recommends top-K items for user u .

For the model optimization on user/item initialization and model parameters, we employ conventional log loss [28,29] and Adam optimizer to train the model as follows.

$$L = \frac{-1}{|R^+| + |R^-|} \left[\sum_{(u,i) \in R^+} \log(\hat{y}_{ui}) + \sum_{(u,i) \in R^-} \log(1 - \hat{y}_{ui}) \right] + \lambda \|\theta\|^2 \quad (13)$$

where R^+ and R^- represent all positive and negative samples in the training set, respectively, $|R^+|$ and $|R^-|$ are the number of positive and negative samples. During optimization, we employ L_2 regularization and dropout strategy to prevent model overfitting.

4. Experiments

We conduct experiments to verify the effectiveness of the proposed KMEL model for personalized recommendation. With the experiments, We aim to answer the following research questions.

- **RQ1** Compared to advanced recommendation models, how does the proposed KMEL perform?
- **RQ2** How does knowledge signals extracted from multiple semantic spaces affect the performance?
- **RQ3** How does the model hyperparameters work in KMEL?

4.1. Experimental Settings

4.1.1. Dataset Description

Experiments are performed on Amazon-Book [30] and Yelp2018 [10] datasets for personalized recommendation, which have varying sparsity and domain knowledge. Table 1 summarizes the statistics of the experimental datasets.

- **Amazon-Book.** We selected Amazon-book from the widely used product dataset Amazon-review, which has a relatively high sparsity. We kept the users and items with at least 10 interactions to guarantee the reliability of the dataset. Considering the possible significance to a specific domain, three relations *Subjects*, *Author*, *Character* were selected from those given by the dataset to construct semantic spaces. With the three relations, the entity size aligned from the knowledge graph to items is large enough to mine semantic signals for modeling users' interests.
- **Yelp2018.** Yelp2018 is a dataset sampled from the field of music applications. Similarly, we kept the users and items with at least 10 interactions for experiments. Relations *Categories* and *Music* were employed to construct semantic spaces. The relations also provided enough entity to items for modeling users' interests.

In addition to user-item interactions, the proposed KMEL model builds multiple independent semantic spaces on the hybrid interaction-knowledge graph. The interaction-knowledge graph includes a large number of user-item-item triples, each of which is composed of an interacted user-item pair and an item-item pair taking the same semantic relation in the given knowledge graph. We built item-item connections with specific knowledge associations in the datasets. On Amazon-book and Yelp2018 datasets, two items are connected in a specific semantic space when they take the same tail entity in the knowledge graph with the semantic relation. We randomly selected 80% of the user-interacted items as positive samples in the training set, and the remaining 20% as

the test set [10]. In the training set, we randomly matched a negative sample for each positive sample to participate in the model optimization [10]. In the test set, we randomly selected 100 negative samples for each positive sample of users to test the recommendation performance [28].

Table 1. Statistics of the datasets in experiments.

		Amazon-Book	Yelp2018
User-Item Interactions	#Users	70,679	45,919
	#Items	24,915	45,538
	#Interactions	847,733	1,185,068
Knowledge Graph	#Entities	88,572	90,961
	#Relations	39	42
	#Triplets	2,557,746	1,853,704

4.1.2. Evaluation Metrics

The performances were evaluated with the commonly used Normalized Discounted Cumulative Gain at rank K (NDCG@K) and Recall@K on Top-K recommendation lists. We set K to be 10 without specification, i.e., we mainly evaluated the recommendation performance on the Top-10 items in the recommendation lists of users. For the datasets, we show the average recommendation performance based on both metrics on all the users in the test set.

4.1.3. Baselines

We compare the proposed KMEL model with ID-based (NCF [28]), graph-based (NGCF [17], GCMC [15]), and knowledge-based (CKE [2], RippleNet [11], KGAT [10]) models for personalized recommendation.

- **NCF [28]** constructs a multilayer deep network to perform nonlinear user–item interactions, aiming to capture hidden nonlinear collaborative signals between users and items for recommendation. It represents a simplified KMEL model with only ID and nonlinear interaction function;
- **GCMC [15]** utilizes a graph convolutional encoder to embed users and items and feeds them into a nonlinear decoder to predict users’ interests in items. It performs the same as the collaborative embedding channel in KMEL;
- **NGCF [17]** recursively performs interest propagation on users’ interaction graph to extract higher-order collaboration signals for embedding users and items. It additionally encodes relations to GCMC, however, only the interaction relation is investigated;
- **CKE [2]** investigates knowledge base of items to enrich latent embeddings of items and performs interaction between the enhanced item embedding with the naive latent embedding of users for pairwise matching. Knowledge enriches only item representation in CKE, but both users and items in KMEL;
- **RippleNet [11]** treats the items interacted by users as seeds and aggregates high-order semantic signals through path propagation of the seeds over a knowledge graph to learn embeddings. It involves neighbors in KG to help propagate interests while KMEL investigate both neighbors and the corresponding various relations;
- **KGAT [10]** designs knowledge-aware attention on neighbors by graph convolution to involve semantic collaborative signals of the knowledge graph into embeddings. Compared to the fine-grained attention of KGAT, the proposed KMEL applies only semantic attention to each space.

4.1.4. Parameter Settings

We implemented the proposed KMEL model with the deep learning framework pytorch. The embedding size of users and items was fixed to 64, and the batch size was set

to 1024. All model parameters were initialized with Gaussian distribution. We adjusted the learning rate in the range $[0.0001, 0.001, 0.01, 0.1]$ and searched L_2 regularization strength in the range $[10^{-6}, 10^{-5}, \dots, 10^{-1}, 1]$ to prevent overfitting. The interest propagation depth of graph convolution in the proposed KMEL model were tuned in $\{1, 2, 3, 4\}$. Without specification, all the comparison models adopted the same hyperparameter settings to compare their recommendation performance for fairness.

4.2. Performance Comparison (RQ1)

The proposed KMEL is compared with NCF, GCMC, NGCF, CKE, RippleNet and KGAT on the Amazon-Book and Yelp2018 datasets by NDCG@10 and Recall@10. Figure 2 shows the performance comparison of the Top-K recommendation lists of the proposed KMEL and its comparisons. Table 2 provides a detailed comparison among them by NDCG@10 and Recall@10, which can be observed from the performances:

- The performance of KMEL w.r.t NDCG and Recall consistently outperformed its comparisons on the Amazon-Book and Yelp2018 datasets. By NDCG@10, the proposed KMEL achieved 2.79% and 0.94% improvement over the strongest baseline on the Amazon-Book and Yelp2018 datasets, respectively. Such performance improvement proves the effectiveness of the proposed KMEL in modeling users' interests. The proposed KMEL is capable of finding semantic correlations among items to aid collaborative embedding learning and alleviating the data sparsity issue for personalized recommendation.
- The poor recommendation performance of NCF on the two datasets compared to other models proves the effectiveness of graph-based interest propagation and even knowledge propagation for mining users' interests. In detail, the performance improvement of GCMC and NGCF over NCF demonstrates the significance of collaborative signals hidden in the interaction graph on revealing interests. The performances of CKE and RippleNet further illustrate the valuable role of knowledge to enrich users' representations and comprehensively model users' interests.
- NGCF performed better than GCMC on both datasets, demonstrating the positive role of high-order collaborative signals in modeling users' interests. Both NGCF and KGAT explore higher-order collaborative signals to embed users, with a difference of knowledge extension in KGAT. KGAT introduced a knowledge graph to aggregate higher-order knowledge structures. Compared with NGCF, KGAT achieved better performance, which verifies the positive impact of semantics in knowledge graph to mine users' interests.
- CKE, RippleNet, and KGAT showed an improved recommendation performance compared to NCF, GCMC, NGCF, due to the valuable knowledge structure information from the knowledge graph in building users' interests. RippleNet outperformed CKE on both datasets, which indicates that introducing multihop neighbor items in a path-propagation manner is relatively effective for learning users' interests, while the regularization-based method may not fully utilize the rich semantics of items. Compared with RippleNet, the performance of KGAT on both datasets shows that the embedding propagation method can utilize the rich semantics of items more effectively than the path-based and regularization-based models. Further, the attention mechanism in KGAT not only improves the interpretability of the recommendation results and further improves the recommendation performance.
- Compared with GCMC and NGCF, the excellent performance of the proposed KMEL on both datasets demonstrates the capability of semantic correlations among items to promote learning users' interests. Meanwhile, KMEL outperformed RippleNet and KGAT, proving the effectiveness of the cross-space multisemantic structures in modeling users' interests for personalized recommendation.

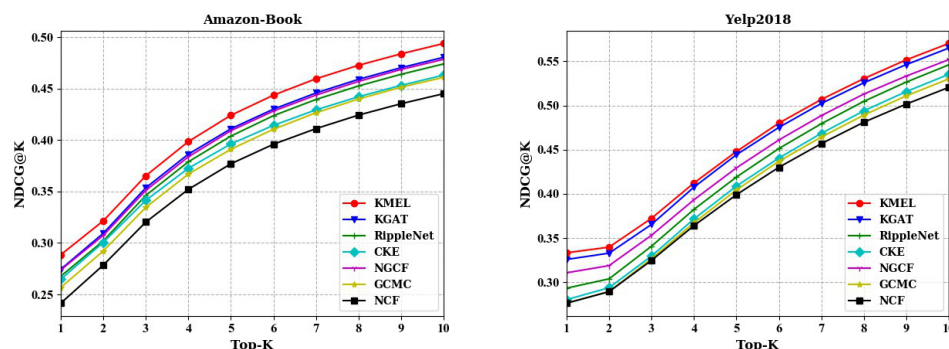


Figure 2. Performance comparison of Top-K recommendations by NDCG@K on Amazon-Book and Yelp2018 datasets, $K = 1, 2, \dots, 10$.

Table 2. Performance comparison by NDCG@10 and Recall@10.

	Amazon-Book		Yelp2018	
	NDCG@10	Recall@10	NDCG@10	Recall@10
NCF	0.4451	0.5105	0.5204	0.4163
GCMC	0.4609	0.5274	0.5302	0.4354
NGCF	0.4785	0.5419	0.5517	0.4401
CKE	0.4631	0.5206	0.5351	0.4408
RippleNet	0.4739	0.5391	0.5458	0.4502
KGAT	0.4804	0.5493	0.5647	0.4532
KMEL	0.4938	0.5581	0.5709	0.4593
%Improve	2.79%	1.6%	0.94%	1.36%

4.3. Knowledge-Aware Semantics (RQ2)

4.3.1. Impact of Independent Semantic Structures

Here, we attempt to evaluate the role of multiple independent semantic structures on modeling users' interests for recommendation. MLP, as the base model, performs interaction on IDs of users and items, which is employed subsequently to explore the increment of each independent semantic structure on interest learning. Specifically, we first conducted recommendation by MLP, i.e., the castrated KMEL, removing the graph convolution layer and all semantic structures. Then, we introduced the aforementioned collaborative embedding learning with graph convolution (+Gconv) and semantic embedding learning on semantic structures (+Character, +Author, +Subjects, +Categories, +Music), respectively. Figure 3 shows the recommendation performance of the relevant submodels (+Gconv, +Character, +Author, +Subjects, +Categories, +Music) compared to that of MLP (+None) by NDCG@10. On both datasets, it can be observed that the recommendation performance gradually improved by involving more semantics, and the proposed KMEL attained the best performance on recommendation when all the semantic structures were introduced. This demonstrates the complementary role of collaborative and knowledge embeddings on mining users' interests. Additionally, the performance also illustrates the varying impact of semantics on recommendation. This is reasonable, since the importance of semantics varies in revealing users' interests. Some semantics derive a relatively apparent improvement on performance, while others bring a relatively small improvement. Taking the Amazon-Book dataset as an example, the semantics of book author greatly improves the performance compared with other semantics. This coincides with the reality that readers tend to prefer books written by their favorite authors.

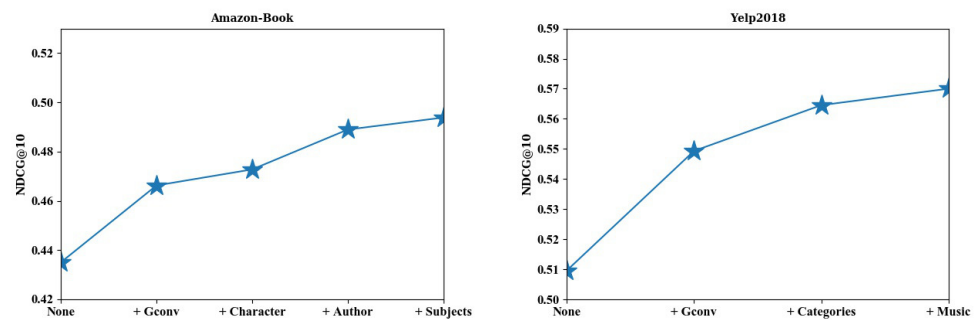


Figure 3. Recommendation performance of relevant submodels (+Gconv,+Character, +Author, +Subjects, +Categories, +Music) compared to that of MLP (+None) on Amazon-Book and Yelp2018 datasets by NDCG@10.

4.3.2. Impact of Target-Aware Multispace Fusion

Experiments were performed to measure the impact of the multispace fusion mechanism on mining users' interests. We compared the recommendation performance of KMEL with linear average aggregation (ave) to that with target-aware attentive aggregation (att). Figure 4a shows the experimental performance of the proposed KMEL with average and target-aware aggregation mechanisms on Amazon-Book and Yelp2018 datasets by NDCG@10. The results show that the target-aware attention-based aggregation outperformed linear aggregation on both datasets, demonstrating the advantage of the target-aware attention mechanism to aggregate multiple spaces on embedding learning adaptively. Since users' interests are inherently complex, global linear aggregation cannot fit users' varying personalized interest distribution. The target-aware attention mechanism can model the importance of semantics on revealing users' interests, which helps effectively improve the recommendation performance.

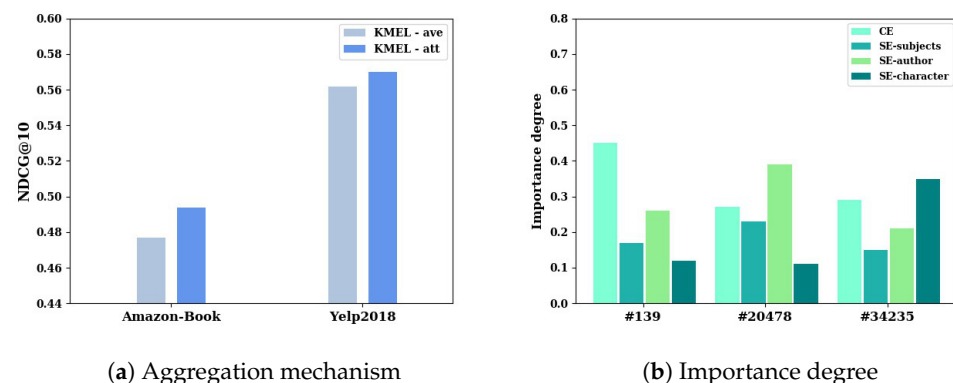


Figure 4. Performance comparison on aggregation by NDCG@10 with (a) linear (ave) and target-aware (att) aggregation mechanisms, and (b) importance degree of collaborative embeddings (CE) and semantic embeddings of semantics (SE-character, SE-author, SE-subjects) for users #139, #20478 and #34235 from Amazon-Book dataset.

The importance degrees of semantics are evaluated by the attention mechanism. Additionally, in order to further track the effect of attention mechanism on user embedding learning, we randomly select users #139, #20478 and #34235 from Amazon-Book dataset to show the importance degree of spaces in Figure 4b. We observe the variation among these users on semantics. For user #139, collaborative embedding (CE) contributes more to modeling the user's interests than other semantics. The other two users have relatively strong preferences on different semantics, respectively. The importance degree varies among semantics even varies among users. The varying importance degrees support the

conclusion that semantics contribute different to a specific user, and the proposed KMEL is capable of building adaptive aggregation with the target-aware attention mechanism.

4.4. Study of KMEL (RQ3)

To measure the impact of hyperparameters on model recommendation performance, we perform experiments on interaction manner and propagation depth on Amazon-Book and Yelp2018 datasets.

4.4.1. Effect of Nonlinear Interaction

User interests and item attributes are diverse and complex, making their interaction hard to predict. To measure the role of interaction manner on recommendation performance, we performed KMEL with linear and nonlinear interaction functions on Amazon-Book and Yelp2018 datasets. The inner product of user-item pair conducted standard linear interaction and the MLP layer performed typical nonlinear interaction. Figure 5 shows the recommendation performance of KMEL with linear and nonlinear interaction functions by NDCG@10. It can be seen that the nonlinear interaction function performed better than the linear interaction. This proves that the nonlinear correlation inherently exists between users and items in interaction. The proposed KMEL leverages the advantage of MLP on nonlinear mapping to perform complex user-item interactions for personalized recommendation.

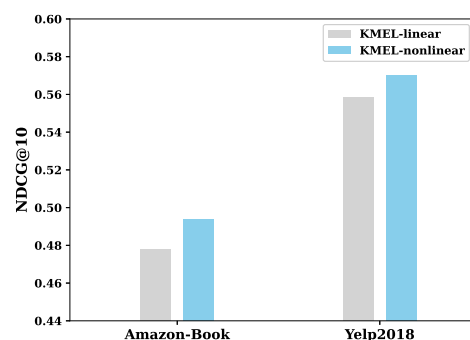


Figure 5. Performance of the proposed KMEL with linear and nonlinear interaction on Amazon-Book and Yelp2018 datasets.

4.4.2. Effect of Propagation Depth

We adjusted the propagation depth of graph convolution on both collaborative signals and semantic signals in the range of $\{1, 2, 3, 4\}$ to explore the effect of propagation depth of the embedding learning layer. We used KMEL-1 to characterize the model with single-layer propagation and KMEL-2,3,4 to characterize models with more depths. Table 3 summarizes the performance of the proposed KMEL with varying propagation depth by NDCG@10 and Recall@10. It shows that as the propagation depth increases, the recommendation performance gradually increases until it reaches the optimal. The optimal depths of the proposed KMEL on Amazon-Book and Yelp2018 datasets are at 3-layers for KMEL-3 and 2-layers for KMEL-2, respectively. This difference is attributed to the varying characters of the datasets, especially the difference in data sparsity. Deep propagation is required on the relatively sparse Amazon-Book dataset to capture relatively more collaborative signals on embedding users' interests. Meanwhile, we observe that KMEL-1 outperformed other baselines in most cases, further demonstrating the positive effect of semantic knowledge on modeling users' interests. Considering effectiveness and efficiency, the proposed KMEL adopts 2-layer propagation with graph convolution to learn embeddings for recommendation.

Table 3. Performance of the proposed KMEL with varying propagation depth by NDCG@10 and Recall@10.

	Amazon-Book		Yelp2018	
	NDCG@10	Recall@10	NDCG@10	Recall@10
KMEL-1	0.4751	0.5244	0.5515	0.4344
KMEL-2	0.4938	0.5581	0.5709	0.4594
KMEL-3	0.4970	0.5584	0.5701	0.4580
KMEL-4	0.4941	0.5564	0.5694	0.4583

5. Conclusions

We have proposed a knowledge-aware multispace embedding learning model for personalized recommendation. The proposed KMEL extracts collaborative signals from multiple independent semantic structures and adaptively integrates collaborative and semantic signals to predict users' interests with a target-aware aggregation. The proposed KMEL uses the semantic correlation among items to learn users' interests, which coincides with the reality that knowledge systems exist and impact users' interests. Extensive experiments on two real-world datasets demonstrate the effectiveness of the proposed KMEL on modeling users' interests with multiple semantic knowledge. Personalized recommendation tasks have always faced heavy data sparsity and cold-start issues. Existing knowledge acts to explain various relations in the view of causal or codependent relationships. For example, a user likes a movie due to the famous actor. This movie-actor knowledge provides a significant clue to explain users' interests. Valuable knowledge helps a comprehensive understanding of the world, which would be a promising way to mine and build users' interests as a future direction.

Author Contributions: Methodology, M.J. and C.Z.; Writing—original draft, X.F. and L.W.; Writing—review & editing, Z.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China under Grant NO. 62176011, NO. 61702022, NO. 61802011, and NO. 61976010, Beijing Municipal Education Committee Science Foundation under Grant NO. KM201910005024, Inner Mongolia Autonomous Region Science and Technology Foundation under Grant NO. 2021GG0333, and Beijing Postdoctoral Research Foundation under Grant NO. Q6042001202101.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sun, Z.; Guo, Q.; Yang, J.; Fang, H.; Guo, G.; Zhang, J.; Burke, R. Research commentary on recommendations with side information: A survey and research directions. *Electron. Commer. Res. Appl.* **2019**, *37*, 100879. [[CrossRef](#)]
2. Zhang, F.; Yuan, N.J.; Lian, D.; Xie, X.; Ma, W.Y. Collaborative knowledge base embedding for recommender systems. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 353–362.
3. Huang, J.; Zhao, W.X.; Dou, H.; Wen, J.R.; Chang, E.Y. Improving sequential recommendation with knowledge-enhanced memory networks. In Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, Ann Arbor, MI, USA, 8–12 July 2018; pp. 505–514.
4. Wang, H.; Zhang, F.; Zhao, M.; Li, W.; Xie, X.; Guo, M. Multi-task feature learning for knowledge graph enhanced recommendation. In Proceedings of the World Wide Web Conference, San Francisco, CA, USA, 13–17 May 2019; pp. 2000–2010.
5. Fan, W.; Ma, Y.; Li, Q.; He, Y.; Zhao, Y.E.; Tang, J.; Yin, D. Graph Neural Networks for Social Recommendation. In Proceedings of the World Wide Web Conference, WWW 2019, San Francisco, CA, USA, 13–17 May 2019; ACM: New York, NY, USA, 2019; pp. 417–426.
6. Ehrlinger, L.; Wöß, W. Towards a definition of knowledge graphs. In Proceedings of the SEMANTICS (Posters Demos SuCCESS), Leipzig, Germany, 13–14 September 2016; Volume 48, p. 2.

7. Gomez-Perez, J.M.; Pan, J.Z.; Vetere, G.; Wu, H. Enterprise knowledge graph: An introduction. In *Exploiting Linked Data and Knowledge Graphs in Large Organisations*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 1–14.
8. Li, C.; Hu, L.; Shi, C.; Song, G.; Lu, Y. Sequence-aware Heterogeneous Graph Neural Collaborative Filtering. In Proceedings of the 2021 SIAM International Conference on Data Mining (SDM), Virtual Event, 29 April–1 May 2021; SIAM: Philadelphia, PA, USA, 2021; pp. 64–72.
9. Hu, B.; Shi, C.; Zhao, W.X.; Yu, P.S. Leveraging Meta-path based Context for Top- N Recommendation with A Neural Co-Attention Model. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2018, London, UK, 19–23 August 2018; ACM: New York, NY, USA, 2018; pp. 1531–1540.
10. Wang, X.; He, X.; Cao, Y.; Liu, M.; Chua, T.S. Kgat: Knowledge graph attention network for recommendation. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 4–8 August 2019; pp. 950–958.
11. Wang, H.; Zhang, F.; Wang, J.; Zhao, M.; Li, W.; Xie, X.; Guo, M. Ripplenet: Propagating user preferences on the knowledge graph for recommender systems. In Proceedings of the 27th ACM International Conference on Information and Knowledge Management, Torino, Italy, 22–26 October 2018; pp. 417–426.
12. Koren, Y.; Bell, R.; Volinsky, C. Matrix factorization techniques for recommender systems. *Computer* **2009**, *42*, 30–37. [\[CrossRef\]](#)
13. Hu, Y.; Koren, Y.; Volinsky, C. Collaborative filtering for implicit feedback datasets. In Proceedings of the 2008 Eighth IEEE International Conference on Data Mining, Pisa, Italy, 15–19 December 2008; pp. 263–272.
14. Wang, C.; Zhu, H.; Zhu, C.; Qin, C.; Xiong, H. Setrank: A setwise bayesian approach for collaborative ranking from implicit feedback. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 6127–6136.
15. Van den Berg, R.; Kipf, T.N.; Welling, M. Graph Convolutional Matrix Completion. *arXiv* **2017**, arXiv:1706.02263.
16. Ying, R.; He, R.; Chen, K.; Eksombatchai, P.; Hamilton, W.L.; Leskovec, J. Graph convolutional neural networks for web-scale recommender systems. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, UK, 19–23 August 2018; pp. 974–983.
17. Wang, X.; He, X.; Wang, M.; Feng, F.; Chua, T.S. Neural graph collaborative filtering. In Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, Paris, France, 21–25 July 2019; pp. 165–174.
18. Wang, X.; Jin, H.; Zhang, A.; He, X.; Xu, T.; Chua, T.S. Disentangled graph collaborative filtering. In Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, Xi'an, China, 25–30 July 2020; pp. 1001–1010.
19. He, X.; Du, X.; Xiang, W.; Feng, T.; Chua, T.S. Outer Product-based Neural Collaborative Filtering. In Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence IJCAI-18, Stockholm, Sweden, 13–19 July 2018.
20. Jamali, M.; Ester, M. A matrix factorization technique with trust propagation for recommendation in social networks. In Proceedings of the Fourth ACM Conference on Recommender Systems, Barcelona, Spain, 26–30 September 2010; pp. 135–142.
21. Wang, H.; Zhang, F.; Hou, M.; Xie, X.; Guo, M.; Liu, Q. Shine: Signed heterogeneous information network embedding for sentiment link prediction. In Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, Marina Del Rey, CA, USA, 5–9 February 2018; pp. 592–600.
22. Sun, Y.; Yuan, N.J.; Xie, X.; McDonald, K.; Zhang, R. Collaborative intent prediction with real-time contextual data. *ACM Trans. Inf. Syst.* **2017**, *35*, 1–33. [\[CrossRef\]](#)
23. He, X.; Deng, K.; Wang, X.; Li, Y.; Zhang, Y.; Wang, M. Lightgcn: Simplifying and powering graph convolution network for recommendation. In Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, Xi'an, China, 25–30 July 2020; pp. 639–648.
24. Wu, J.; He, X.; Wang, X.; Wang, Q.; Chen, W.; Lian, J.; Xie, X. Graph convolution machine for context-aware recommender system. *Front. Comput. Sci.* **2022**, *16*, 1–12. [\[CrossRef\]](#)
25. Yang, R.; Jian, M.; Shi, G.; Wu, L.; Xiang, Y. Attribute-Level Interest Matching Network for Personalized Recommendation. In *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 486–497.
26. Qiannan, Z.; Zhou, X.; Wu, J.; Tan, J.; Guo, L. A Knowledge-Aware Attentional Reasoning Network for Recommendation. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 6999–7006. [\[CrossRef\]](#)
27. Huang, C.; Xu, H.; Xu, Y.; Dai, P.; Xiao, L.; Lu, M.; Bo, L.; Xing, H.; Lai, X.; Ye, Y. Knowledge-aware coupled graph neural network for social recommendation. In Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI), Virtual Event, 2–9 February 2021.
28. He, X.; Liao, L.; Zhang, H.; Nie, L.; Hu, X.; Chua, T.S. Neural collaborative filtering. In Proceedings of the 26th International Conference on World Wide Web, Perth, Australia, 3–7 April 2017; pp. 173–182.
29. Xue, F.; He, X.; Wang, X.; Xu, J.; Liu, K.; Hong, R. Deep item-based collaborative filtering for top-n recommendation. *ACM Trans. Inf. Syst.* **2019**, *37*, 1–25. [\[CrossRef\]](#)
30. He, R.; McAuley, J.J. Ups and Downs: Modeling the Visual Evolution of Fashion Trends with One-Class Collaborative Filtering. In Proceedings of the 25th International Conference on World Wide Web, WWW 2016, Montreal, QC, Canada, 11–15 April 2016; ACM: New York, NY, USA, 2016; pp. 507–517.