Disentangled Contrastive Learning for Knowledge-Aware Recommender System

Shuhua Huang[1], Chenhao Hu[1], Weiyang Kong[1], and Yubao Liu[1,2](✉)

[1] Sun Yat-Sen University,Guangzhou, China
{huangshh33,huchh8,kongwy3}@mail2.sysu.edu.cn, liuyubao@mail.sysu.edu.cn
[2] Guangdong Key Laboratory of Big Data Analysis and Processing, Guangzhou, China

Abstract. Knowledge Graphs (KGs) play an increasingly important role as useful side information in recommender systems. Recently, developing end-to-end models based on graph neural networks (GNNs) becomes the technical trend of knowledge-aware recommendation. However, we argue that prior methods are insufficient to discover multifaceted user preferences based on diverse aspects of item attributes, since they only learn a single representation for each user and item. To alleviate this limitation, we focus on exploring user preferences from multiple aspects of item attributes, and propose a novel disentangled contrastive learning framework for knowledge-aware recommendation (DCLKR). Technically, we first disentangle item knowledge graph into multiple aspects for the knowledge view, and user-item interaction graph for the collaborative view, equipped with attentive neighbor assignment and embedding propagation mechanisms. Then we perform intra-view contrastive learning to encourage differences among disentangled representations in each view, and inter-view contrastive learning to transfer knowledge between the two views. Extensive experiments conducted on three benchmark datasets demonstrate the superior performance of our proposed method over the state-of-the-arts. The implementations are available at: https://github.com/Jill5/DCLKR..

Keywords: Recommender System · Knowledge Graphs · Disentangled Representation Learning · Contrastive Learning · Graph Neural Networks

1 Introduction

Recommender systems are crucial for many online services to discover interested items for users. For developing effective recommendation approaches, learning high-quality user and item representations is of great significance. In recent years, a great deal of research effort is devoted to utilizing knowledge graphs (KGs) to improve the representation learning of recommendation [28,33,45]. A KG is a

T. R. Payne et al. (Eds.): ISWC 2023, LNCS 14265, pp. 140–158, 2023.
https://doi.org/10.1007/978-3-031-47240-4_8

semantic network of real-world entities, and illustrates the relationship between them. The rich entity and relation information can not only reveal various relatedness among items (*e.g.*, co-directed by a director) but also be used to interpret user preference (*e.g.*, attributing a user's choice of a movie to its director).

Early studies on knowledge-aware recommendation focus on bridging different knowledge graph embedding (KGE) models [2,16,39] with recommendation models, by pre-processing KGs with KGE models and feeding the learned entity embeddings into recommendation frameworks. Some follow-on studies [10,36,42] propose to construct multi-hop paths along with multiple relations in KGs from users to items, exploiting the high-order KG connectivity to model user-item relations better. More recently, due to the powerful capabilities of graph neural networks (GNNs) [7,15,25], the information aggregation schemes of GNNs become the mainstream in knowledge-aware recommendation [23,32–34]. Such methods unify user-item interactions and KGs as user-item-entity graphs, then recursively integrate multi-hop neighbors into node representations.

However, we argue that prior methods are insufficient to discover multi-faceted user preferences. The key reason is that each item contains diverse relation and entity information, but prior methods only learn a single representation for each item, which is further used to characterize user preferences. An underlying fact has been ignored that user preferences are multi-faceted based on diverse aspects of item attributes, and a user likes an item doesn't mean he/she likes all the attributes of the item. Taking Fig. 1 as an example, the movie *Batman Begins* has multiple aspects of relation and entity information, user u_1 saw the movie *Batman Begins* because he liked its *genre*, while user u_2 saw this movie for its *director* and *star*. Ignoring the diverse facets behind user preferences limits the performance of recommendation. To solve this limitation, we propose to explore user preferences at a more granular level, by disentangling item knowledge graph and user-item interaction graph under multiple aspects of item attributes, which form the knowledge view and the collaborative view, respectively. The main challenge is how to learn such disentangled representations of users and items in two views, while transferring knowledge between these two views for knowledge-aware recommendation.

Recently, contrastive learning, one of the classical self-supervised learning (SSL) methods, shows excellent performance on learning discriminative representations from unlabeled data, via maximizing the distance between negative samples while minimizing the distance between positive samples [17]. Besides, contrastive learning enables knowledge transferring between views by maximizing the mutual information between those augmented views of the same instance (*i.e.*, user or item) [47].

Motivated by the advantage of contrastive learning in representation learning, we develop a novel **D**isentangled **C**ontrastive **L**earning framework for **K**nowledge-aware **R**ecommendation (DCLKR). More specifically, we first initialize multi-aspect embeddings via multiple gate units, coupling each gate unit with an aspect. We then apply graph disentangling modules in the knowledge view and collaborative view separately, equipped with attentive neighbor assign-

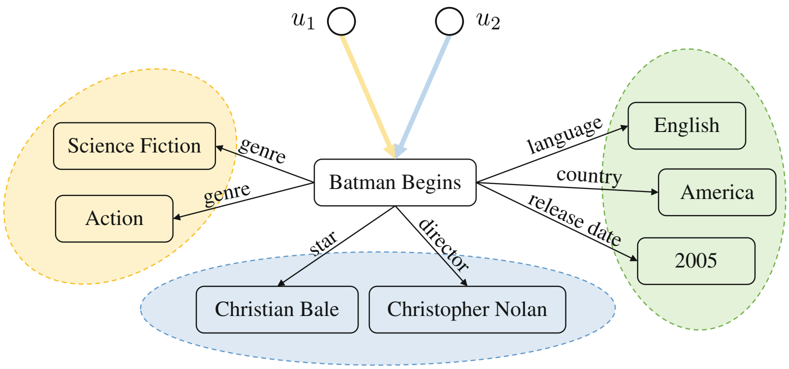

Fig. 1. A toy example of different facets of user preferences. Best viewed in color.

ment and embedding propagation mechanisms. In particular, attentive neighbor assignment exploits node-neighbor affinity to refine the graph in each aspect, highlighting the importance of influential connections, *i.e.*, user-item interactions and KG triplets. In turn, embedding propagation on such graphs updates a node embedding relevant to a certain aspect. By iteratively performing such disentangling operations, we establish a set of disentangled representations under multiple aspects. Simultaneously, a contrastive learning module is introduced, consisting of intra-view contrastive learning and inter-view contrastive learning. The intra-view contrastive learning is performed to encourage differences among disentangled representations in each view. Besides, the inter-view contrastive learning is conducted to align item representations between two views, for transferring item knowledge to the collaborative view as well as collaborative signals to the knowledge view.

Our contributions are summarized as follows:

- This work emphasizes the significance of exploring multi-faceted user preferences based on different aspects of item attributes, and presents the idea of modeling multi-faceted user preferences by disentangled representation learning.
- We propose a novel model DCLKR, which builds a disentangled contrastive learning framework for knowledge-aware recommendation. DCLKR learns disentangled representations of users and items from the knowledge view and the collaborative view. Besides, it performs intra-view and inter-view contrastive learning to enhance representation learning.
- We conduct extensive experiments on three benchmark datasets to demonstrate the advantages of our DCLKR in recommendation, and investigate the effectiveness of each component with ablation studies.

2 Preliminaries

In this section, we introduce main notations used throughout the paper and formulate the knowledge-aware recommendation task.

In a typical recommendation scenario, let $\mathcal{U}$ be a set of users and $\mathcal{I}$ be a set of items, respectively. Let $\mathcal{O}^+ = \{(u,i)|u \in \mathcal{U}, i \in \mathcal{I}\}$ be a set of observed feedback, where each (u,i) pair indicates that user u has engaged item i before.

KGs store plentiful real-world facts associated with items, *e.g.*, item attributes, or external commonsense knowledge, in the form of heterogeneous graphs. Let a KG be a collection of triplets $\mathcal{G} = \{(h,r,t)|h,t \in \mathcal{V}, r \in \mathcal{R}\}$, where each triplet (h,r,t) indicates that a relation r exists from head entity h to tail entity t; $\mathcal{V}$ and $\mathcal{R}$ refer to the sets of entities and relations in $\mathcal{G}$, respectively. Here, $\mathcal{V}$ is comprised of items $\mathcal{I}$ and non-item entities $\mathcal{V}/\ \mathcal{I}$. For example, the triplet (*Batman Begins*, *star*, *Christian Bale*) describes that *Christian Bale* is the *star* of movie *Batman Begins*.

Given the user-item interaction data $\mathcal{O}^+$ and the knowledge graph $\mathcal{G}$, our task of knowledge-aware recommendation is to learn a function that can predict the probability that a user $u \in \mathcal{U}$ would interact with an item $i \in \mathcal{I}$.

3 Methodology

In this section, we present the proposed DCLKR. It aims to incorporate contrastive learning into knowledge-aware recommendation to model disentangled multi-faceted user preferences. The framework of DCLKR is illustrated in Fig. 2, which consists of three key components: (1) **Knowledge Graph Disentangling Module**. It incorporates an attentive neighbor assignment mechanism into a path-aware GNN to encode disentangled knowledge-aware representations of items. (2) **Interaction Graph Disentangling Module**. It applies an attentive light aggregation scheme to encode the interaction graphs, under the guidance of the multi-faceted representations from the knowledge view. (3) **Contrastive Learning Module**. First, it separately performs intra-view contrastive learning in the two views, then conducts inter-view contrastive learning to aligned item representations between the two views. We next present the three components in details.

3.1 Multi-aspect Embeddings Initialization

Before the graph disentangling, we need to initialize embeddings for multiple aspects. Formally, we assume that there are total K aspects. Instead of slicing ID embeddings into K chunks [35], we utilize element-wise self-gating units to control the information flow from ID embeddings to each aspect, as follow:

$$\mathbf{e}_{i,k} = f_{gate}^{k}(\mathbf{e}_i) = \mathbf{e}_i \odot \sigma(\mathbf{W}_k \mathbf{e}_i + \mathbf{b}_k)\,, \tag{1}$$

where $\mathbf{e}_i$ is ID embedding of item i, $\mathbf{e}_{i,k}$ is the initial embedding of item i under the k-th aspect, $\mathbf{W}_k \in \mathbb{R}^{d\times d}$ and $\mathbf{b}_k \in \mathbb{R}^d$ are parameters to be learned, $\odot$ denotes the element-wise product and σ is the sigmoid function. Analogously, $\mathbf{e}_{u,k}$, $\mathbf{e}_{v,k}$, $\mathbf{e}_{r,k}$, are established for user u, entity v and relation r, respectively. The self-gating mechanism effectively learns non-linear gates to modulate ID

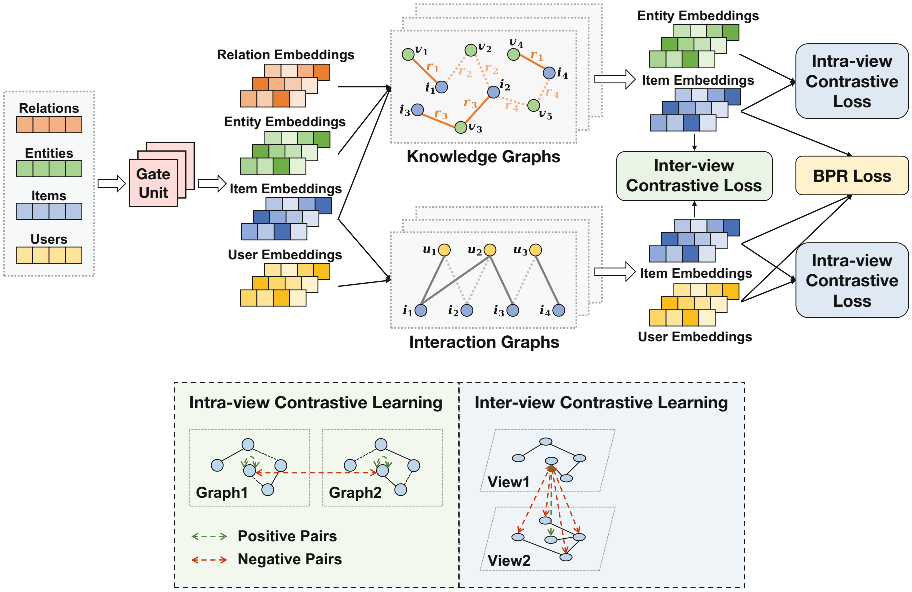

Fig. 2. Illustration of the proposed DCLKR model. The upper part is the model framework of DCLKR, and the lower part is the details of intra-view and inter-view contrastive learning mechanism. Best viewed in color.

embeddings under different aspects at element-wise granularity through dimension re-weighting, which is more adaptive than simply dividing embeddings into multiple chunks.

3.2 Knowledge Graph Disentangling Module

In this component, we aim to learn disentangled knowledge-aware representations to distinguish different aspects of relations and entities. Inspired by [34], we propose a path-aware GNN to encode the relation information in item knowledge graphs. The path-aware GNN aggregates neighboring information for L times, *i.e.*, aggregation depth, meanwhile preserving the path information, *i.e.*, long-range connectivity such as item-relation-entity-relation-item.

However, in the knowledge graph disentanglement task, we should not aggregate all the neighbors when re-constructing node representations in one aspect, as only a subset of neighbors are highly correlated with this aspect. Taking Fig. 1 as an example, movie *Batman Begins*'s neighbors, (*genre*, *Science Fiction*) and (*genre*, *Action*) are strongly relevant to the aspect of *genre*, while (*language*, *English*) and (*release date*, *2005*) are weakly correlated. Thus, in order to better capture the affinity between item i and its neighbors in each aspect, we leverage an attentive neighbor assignment mechanism to infer the importance of each neighbor in aggregation. Here we simply adopt the similarity-based attention based on the hypothesis that the more similar the item i and the neighbor (r, v)

are in the k-th aspect, the better neighbor (r, v) characterizes the feature of item i in terms of the k-th aspect. The attention score of item i's neighbor (r, v) in the k-th aspect is formulated as:

$$\alpha^k_{(i,r,v)} = \frac{\exp(\mathbf{e}^\top_{i,k}(\mathbf{e}_{r,k} \odot \mathbf{e}_{v,k}))}{\sum_{k' \in K} \exp(\mathbf{e}^\top_{i,k'}(\mathbf{e}_{r,k'} \odot \mathbf{e}_{v,k'}))}\,. \tag{2}$$

With the attention score, the l-th layer aggregation in the k-th aspect can be formulated as:

$$\mathbf{e}^{(l+1)}_{i,k} = \frac{1}{|\mathcal{N}^s_i|} \sum_{(r,v)\in\mathcal{N}^s_i} \alpha^k_{(i,r,v)} \mathbf{e}_{r,k} \odot \mathbf{e}^{(l)}_{v,k}\,, \tag{3}$$

where $\mathcal{N}^s_i$ represents a set of item i's neighbors in the knowledge graph, $\mathbf{e}^{(l+1)}_{i,k}$ denotes the k-th representation of item i after $l+1$ layers aggregation, and representation $\mathbf{e}^{(l)}_{v,k}$ of entity v is obtained by l layers aggregation in a similar way.

Then we sum all layers' representations up to obtain the final representations specific to the k-th aspect:

$$\mathbf{x}^s_{i,k} = \mathbf{e}^{(0)}_{i,k} + \cdots + \mathbf{e}^{(L)}_{i,k}\,, \tag{4}$$

where $\mathbf{e}^{(0)}_{i,k}$ is equal to the initial embedding of item i under the k-th aspect.

3.3 Interaction Graph Disentangling Module

The collaborative view lays stress on collaborative signals in user-item interactions, *i.e.*, user-item-user and item-user-item co-occurrences. As a result, collaborative information could be captured by modeling long-range connectivity in the user-item interaction graphs, where an edge between a user and an item indicates that the user has interacted with the item. Thus, we adopt a light aggregation scheme referred to LightGCN [9], which adopts a simple message passing and aggregation mechanism without feature transformation and non-linear activation, effective and computationally efficient.

However, like knowledge graph disentangling, it is unwise to aggregate all the interacted neighbors under one aspect when disentangling interaction graphs, as only a subset of neighbors are strongly correlated with this aspect. Taking Fig. 1 as an example, the relations and entities of the *genre* aspect are the main factors leading to the interaction between user u_1 and movie *Batman Begins*, while those of other aspects are not. Thus, we also leverage an attentive neighbor assignment mechanism to refine the interaction graph by inferring the importance of each interaction under different aspects, which is based on disentangled knowledge-aware representations. In particular, the attention score of an interaction (u, i) under the k-th aspect is formulated as:

$$\alpha^k_{(u,i)} = \frac{\exp(\mathbf{e}^\top_{u,k}\mathbf{x}^s_{i,k})}{\sum_{k' \in K} \exp(\mathbf{e}^\top_{u,k'}\mathbf{x}^s_{i,k'})}\,. \tag{5}$$

At the l-th layer under the k-th aspect, the aggregation can be formulated as:

$$\mathbf{e}_{u,k}^{(l+1)} = \frac{1}{|\mathcal{N}_u^c|} \sum_{i \in \mathcal{N}_u^c} \alpha_{(u,i)}^k \mathbf{e}_{i,k}^{(l)}, \quad \mathbf{e}_{i,k}^{(l+1)} = \frac{1}{|\mathcal{N}_i^c|} \sum_{u \in \mathcal{N}_i^c} \alpha_{(u,i)}^k \mathbf{e}_{u,k}^{(l)}, \tag{6}$$

where $\mathcal{N}_u^c$ and $\mathcal{N}_i^c$ represent sets of user u's neighbors and item i's neighbors in the interaction graph.

Then representations at different layers are summed up as the collaborative representations of the k-th aspect, as follows:

$$\mathbf{x}_{u,k}^c = \mathbf{e}_{u,k}^{(0)} + \cdots + \mathbf{e}_{u,k}^{(L)}, \quad \mathbf{x}_{i,k}^c = \mathbf{e}_{i,k}^{(0)} + \cdots + \mathbf{e}_{i,k}^{(L)}, \tag{7}$$

where $\mathbf{e}_{u,k}^{(0)}$ and $\mathbf{e}_{i,k}^{(0)}$ are equal to the initial embeddings of user u and item i in the k-th aspect.

3.4 Contrastive Learning Module

Intra-view Contrastive Learning. We expect that there should be a weak dependence among disentangled representations from different aspects. In general, disentangled representations with unique information will be able to supply diverse and complementary angles to characterize node features. Otherwise, they might be less informative and not capable of achieving comprehensive disentanglement.

Here we utilize contrastive learning among disentangled knowledge graphs, as well as among disentangled interaction graphs, to guide the independent representation learning. First, we define the positive and negative samples. In particular, for any node in one view, its representations under the same aspect form the positive pairs, and its representations under different aspects form the negative pairs. With the positive and negative samples, we have the following contrastive loss in the knowledge view:

$$\mathcal{L}_{intra}^s = \sum_{v \in \mathcal{V}} \sum_{k \in K} -\log \frac{e^{s(\mathbf{x}_{v,k}^s, \mathbf{x}_{v,k}^s)/\tau}}{\sum_{k' \in K} e^{s(\mathbf{x}_{v,k}^s, \mathbf{x}_{v,k'}^s)/\tau}}, \tag{8}$$

where $s(\cdot)$ denotes the cosine similarity calculating, and τ denotes a temperature parameter. In a similar way, we can obtain the contrastive loss of the collaborative view as follow:

$$\mathcal{L}_{intra}^c = \sum_{n \in \mathcal{U} \cup \mathcal{I}} \sum_{k \in K} -\log \frac{e^{s(\mathbf{x}_{n,k}^c, \mathbf{x}_{n,k}^c)/\tau}}{\sum_{k' \in K} e^{s(\mathbf{x}_{n,k}^c, \mathbf{x}_{n,k'}^c)/\tau}}. \tag{9}$$

The complete intra-view contrastive loss is the sum of the above two losses:

$$\mathcal{L}_{intra} = \mathcal{L}_{intra}^s + \mathcal{L}_{intra}^c. \tag{10}$$

In this way, we successfully learn discriminative disentangled node representations from various perspectives with the guidance of contrastive learning.

Inter-view Contrastive Learning. In order to transfer item knowledge to the collaborative view, as well as collaborative signals to the knowledge view, we conduct inter-view contrastive learning to align item representations between these two views. For any item in one view, the same item embedding learned by the other view forms the positive sample, and the item embeddings except itself in the other view are naturally regarded as negative samples. With the positive and negative samples, we have the following inter-view contrastive loss:

$$\mathcal{L}_{inter} = \sum_{i \in \mathcal{I}} \sum_{k \in K} -\log \frac{e^{s(\mathbf{x}^s_{i,k}, \mathbf{x}^c_{i,k})/\tau}}{\sum_{i' \in \mathcal{I}} e^{s(\mathbf{x}^s_{i',k}, \mathbf{x}^c_{i,k})/\tau} + \sum_{i' \in \mathcal{I}} e^{s(\mathbf{x}^s_{i,k}, \mathbf{x}^c_{i',k})/\tau}} . \tag{11}$$

3.5 Model Prediction

In this module, we first conduct aspect-level prediction and then leverage an attentive scoring mechanism to guide the fusion of results from different aspects. For the aspect-level prediction, we combine embeddings from two views, and predict their matching scores through inner product as follows:

$$\mathbf{z}_{u,k} = \mathbf{x}^c_{u,k} , \quad \mathbf{z}_{i,k} = \mathbf{x}^s_{i,k} + \mathbf{x}^c_{i,k} , \quad \hat{y}^k_{(u,i)} = \mathbf{z}^\top_{u,k} \mathbf{z}_{i,k} . \tag{12}$$

As discussed in Sect. 3.3, each interaction has different correlation with different aspects. Thus, we adopt an attentive fusion of prediction scores from different aspects to get the final results, as follows:

$$\begin{aligned} \beta^k_{(u,i)} &= \frac{\exp((\mathbf{e}_u \odot \mathbf{e}_i)^\top (\mathbf{e}_{u,k} \odot \mathbf{e}_{i,k}))}{\sum_{k' \in K} \exp((\mathbf{e}_u \odot \mathbf{e}_i)^\top (\mathbf{e}_{u,k'} \odot \mathbf{e}_{i,k'}))} , \\ \hat{y}_{(u,i)} &= \sum_{k \in K} \beta^k_{(u,i)} \hat{y}^k_{(u,i)} . \end{aligned} \tag{13}$$

3.6 Multi-task Training

We apply a multi-task learning strategy to jointly train the recommendation loss and the contrastive losses. For the knowledge-aware recommendation task, we employ a pairwise BPR loss [22] as follow:

$$\mathcal{L}_{BPR} = \sum_{(u,i,j) \in \mathcal{O}} -\ln \sigma(\hat{y}_{(u,i)} - \hat{y}_{(u,j)}) , \tag{14}$$

where $\hat{y}_{(u,i)}$ and $\hat{y}_{(u,j)}$ are predicted scores, $\mathcal{O} = \{(u,i,j)|(u,i) \in \mathcal{O}^+, (u,j) \in \mathcal{O}^-\}$ is the training dataset consisting of the observed interactions $\mathcal{O}^+$ and unobserved counterparts $\mathcal{O}^-$; σ is the sigmoid function. By combining the intra-view and inter-view contrastive losses with BPR loss, we minimize the following objective function to learn the model parameters:

$$\mathcal{L}_{DCLKR} = \mathcal{L}_{BPR} + \lambda_1 \mathcal{L}_{intra} + \lambda_2 \mathcal{L}_{inter} + \lambda_3 \|\Theta\|^2_2 , \tag{15}$$

where Θ is the model parameter set, λ_1 and λ_2 are the hyper-parameters to control the weights of the intra-view and inter-view contrastive losses, λ_3 is the hyper-parameter to control L_2 regularization term, respectively.

Table 1. Statistics and hyper-parameter settings for the three datasets.

		Book-Crossing	MovieLens-1M	Last.FM
User-item Interaction	# users	17,860	6,036	1,872
	# items	14,967	2,445	3,846
	# interactions	139,746	753,772	42,346
Knowledge Graph	# entities	77,903	182,011	9,366
	# relations	25	12	60
	# triplets	151,500	1,241,996	15,518
Hyper-parameter Settings	# K	3	3	3
	# L	2	3	2
	# λ_1	0.1	0.01	0.01
	# λ_2	0.1	0.01	0.01

4 Experiment

Extensive experiments are performed on three public datasets, which are widely used in knowledge-aware recommender systems, to evaluate the effectiveness of our proposed DCLKR by answering the following research questions:

- **RQ1:** How does DCLKR perform, compared with the state-of-the-art knowledge-aware recommender models?
- **RQ2:** Are the key components in our DCLKR framework really improving the overall performance?
- **RQ3:** How do different hyper-parameter settings affect DCLKR?

4.1 Experiment Settings

Dataset Description. Three benchmark datasets are Book-Crossing[1], MovieLens-1M[2], and Last.FM[3], which vary in size, interaction sparsity and knowledge graph characteristics, making our experiments more convincing. Table 1 presents the statistical information of our experimented datasets.

We follow RippleNet [28] to transform the explicit ratings into the implicit marks where 1 indicates that the user has rated the item (the threshold of the rating to be viewed as positive is 4 for MovieLens-1M, but no threshold is set for Book-Crossing and Last.FM due to their sparsity). Closely following RippleNet, we use Microsoft Satori[4] to construct the KGs for three datasets. We gather

[1] http://www2.informatik.uni-freiburg.de/~cziegler/BX/.
[2] https://grouplens.org/datasets/movielens/1m/.
[3] https://grouplens.org/datasets/hetrec-2011/.
[4] https://searchengineland.com/library/bing/bing-satori.

Satori IDs of all valid items through their names, and match the IDs with the heads and tails of all KG triplets to extract all well-matched triplets.

Evaluation Metrics. We conduct the evaluation in two experimental scenarios: (1) In click-through rate (CTR) prediction, we adopt two widely used metrics AUC and $F1$. (2) In top-N recommendation, we choose $Recall@N$ to evaluate the recommended lists, where N is set to 5, 10, 20, 50, and 100 for consistency.

Baselines. To comprehensively demonstrate the effectiveness of our proposed DCLKR, we compare it with different types of recommender system methods:

- BPRMF [22]: It is a conventional collaborative filtering method that uses pairwise matrix factorization for implicit feedback optimized by the pairwise ranking loss.
- CKE [45]: This method first encodes items' semantic knowledge, then unifies knowledge embeddings, text embeddings, and image embeddings into recommendation framework.
- RippleNet [28]: This method propagates users' preferences along with paths in KGs to encode user embeddings.
- KGAT [33]: This GNN-based method designs an attentive message passing scheme over the user-item-entity graph for embedding fusion.
- CKAN [38]: This GNN-based method utilizes different neighbor aggregation schemes over the user-item interaction graphs and KGs, respectively.
- KGIN [34]: It is a state-of-the-art GNN-based knowledge-aware method, which performs relational path-based aggregation on the user-intent-item-entity graph to identify latent intention of users.
- KDR [20]: This method utilizes KGs to guide the implicit disentangled representation learning on the user-item interaction graph.
- KGIC [48]: This method constructs local and non-local graphs for users and items in KGs, and conducts layer-wise contrastive learning on these graphs.
- MCCLK [47]: It is a state-of-the-art knowledge-aware method with contrastive learning, which generates three different graph views and performs contrastive learning across three views on both local and global levels.

Parameter Settings. Our proposed DCLKR is implemented with PyTorch. For a fair comparison, we fix the embedding dimensionality as 64 for all models, and the embedding parameters are initialized with the Xavier method [6]. We optimize our method with Adam [14] with the learning rate of $1e^{-3}$ and the batch size of 2048. And λ_3 of L_2 regularization term is set to $1e^{-5}$. Other hyper-parameter settings are provided in Table 1, including the number of disentangled aspects K, aggregation depth L, intra-view contrastive loss weight λ_1, and inter-view contrastive loss weight λ_2. The best settings for hyper-parameters in all comparison methods are researched by either empirical study or following the original papers.

Table 2. The results of AUC and $F1$ in CTR prediction.

Model	Book-Crossing		MovieLens-1M		Last.FM	
	AUC	*F1*	*AUC*	*F1*	*AUC*	*F1*
BPRMF	0.6583	0.6117	0.8920	0.7921	0.7563	0.7010
CKE	0.6759	0.6235	0.9065	0.8024	0.7471	0.6740
RippleNet	0.7211	0.6472	0.9190	0.8422	0.7762	0.7025
KGAT	0.7314	0.6544	0.9140	0.8440	0.8293	0.7424
CKAN	0.7439	0.6676	0.9091	0.8466	0.8421	0.7607
KGIN	0.7225	0.6730	0.9321	0.8601	0.8602	0.7803
KDR	0.7246	0.6528	0.9265	0.8463	0.8550	0.7790
KGIC	<u>0.7573</u>	0.6723	0.9252	0.8560	0.8590	0.7802
MCCLK	0.7508	<u>0.6774</u>	<u>0.9325</u>	<u>0.8603</u>	<u>0.8742</u>	<u>0.7908</u>
DCLKR	**0.7910***	**0.6983***	**0.9445***	**0.8703***	**0.8936***	**0.8105***
%Imp	4.45%	3.09%	1.29%	1.16%	2.22%	2.49%

4.2 Performance Comparison (RQ1)

We report the overall performance evaluation of all methods in Table 2 and Fig. 3, where %Imp. denotes the relative improvements of the best performing method (starred) over the strongest baselines (underlined). By analyzing the results, we summarize the following observations:

- **Our proposed DCLKR achieves the best results.** DCLKR consistently performs better than other baselines in all cases of measures. More specifically, it achieves considerable improvements over the strongest baselines *w.r.t.* AUC by 4.45%, 1.29%, and 2.22% in Book-Crossing, MovieLens-1M and Last.FM datasets, respectively. In top-N recommendation scenario, DCLKR also achieves best performance *w.r.t.* $Recall@N$ (N = 5, 10, 20, 50, 100). We attribute such improvements to the following aspects: (1) By disentangling the user-item interaction graphs and KGs, DCLKR is able to capture users' multi-faceted preferences based on diverse aspects of item attributes. (2) The contrastive mechanism preserves features from both knowledge view and collaborative view, hence prompting the representations to be more informative for DCLKR.
- **Incorporating KGs benefits recommender systems.** We can observe that all the models that incorporate KGs perform better than conventional CF methods. Compared with BPRMF, CKE simply incorporating KG embeddings into matrix factorization elevates the model performance, which clarifies the significance of bringing in KGs as side information.
- **Extracting more informative KG facts boosts the model performance.** KGIN disentangles user-item interactions at the fine granularity of user intents which related to semantic relations in KGs, so that KGIN is the state-of-the-art in GNN-based knowledge-aware methods. The truth inspires

us to explore user preferences on different aspects of item attributes implicit in KGs. KDR also learns disentangled representations from the knowledge and collaborative view, but ignores different importance of the connections in graphs, which degrades its performance.

– **Contrastive learning benefits graph learning.** We can observe that the methods based on contrastive learning paradigm achieve better performance than the GNN-based methods in most cases, which indicates that contrastive learning brings benefits to the graph learning of recommendation.

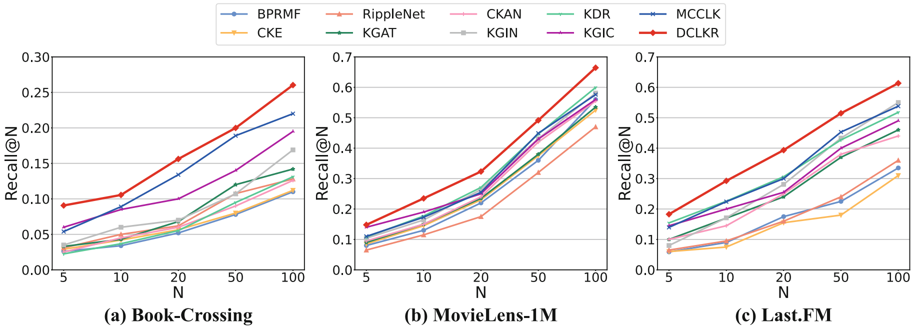

Fig. 3. The results of *Recall@N* in top-*N* recommendation.

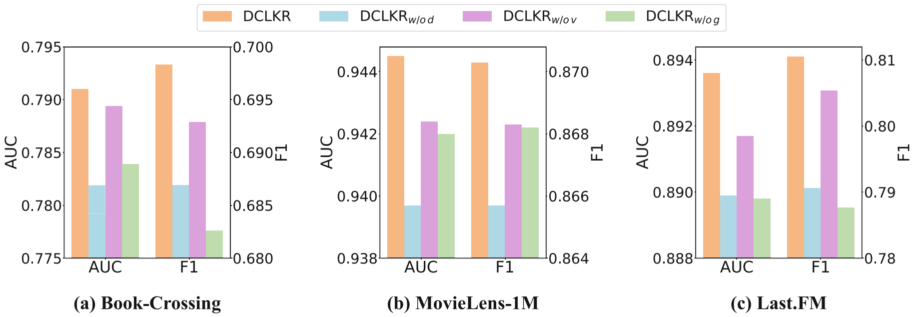

Fig. 4. Effect of ablation study.

4.3 Ablation Studies (RQ2)

We investigate the effect of main components in our model to the final performance by comparing DCLKR with the following three variants:

– DCLKR$_{w/o\,d}$: the variant of DCLKR without disentangled representation learning. Naturally, the intra-view contrastive learning is also removed.
– DCLKR$_{w/o\,v}$: the variant of DCLKR without the inter-view contrastive learning.
– DCLKR$_{w/o\,g}$: the variant of DCLKR which removes the self-gating units and initializes multi-aspect embeddings by simply slicing ID embeddings into multiple chunks.

As shown in Fig. 4, we have the following observations: (1) Without disentangled representation learning, DCLKR$_{w/o\,d}$ leads to a significant performance decrease, which demonstrates that disentangled representation learning is propitious to comprehensive modeling of multi-faceted user preferences. (2) Removing the inter-view contrastive learning degrades the model performance. It makes sense since DCLKR$_{w/o\,v}$ fails to transfer item knowledge to the collaborative view, as well as collaborative signals to the knowledge view, which is beneficial for representation learning. (3) The decreased performance of DCLKR$_{w/o\,g}$ indicates that the self-gating units is superior to the operation of slicing the embeddings, since it can modulate ID embeddings under different aspects at a finer element-wise granularity.

Table 3. Effect of disentangled aspects number K.

	Book-Crossing		MovieLens-1M		Last.FM	
	AUC	*F1*	*AUC*	*F1*	*AUC*	*F1*
$K = 2$	0.7862	0.6893	0.9436	0.8690	0.8925	0.8064
$K = 3$	**0.7910**	**0.6983**	**0.9445**	**0.8703**	**0.8936**	**0.8105**
$K = 4$	0.7906	0.6916	0.9430	0.8696	0.8928	0.8071
$K = 5$	0.7844	0.6903	0.9435	0.8660	0.8919	0.8058

Table 4. Effect of aggregation depth L.

	Book-Crossing		MovieLens-1M		Last.FM	
	AUC	*F1*	*AUC*	*F1*	*AUC*	*F1*
$L = 1$	0.7828	0.6939	0.9332	0.8634	0.8777	0.7937
$L = 2$	**0.7910**	**0.6983**	0.9420	0.8696	**0.8936**	**0.8105**
$L = 3$	0.7908	0.6668	**0.9445**	**0.8703**	0.8892	0.7803
$L = 4$	0.7840	0.6691	0.9424	0.8685	0.8889	0.7782

4.4 Sensitivity Analysis (RQ3)

Effect of Disentangled Aspects Number. To analyze the effect of disentangled aspects number, we vary K in range of {2, 3, 4, 5} and illustrate the performance comparison on Book-Crossing, MovieLens-1M and Last.FM in Table 3. We observe that increasing the number of disentangled aspects enhances the predictive results, as it enables model to capture user preferences from more diverse perspectives. However, excessive number of disentangled aspects impairs model performance, as it is detrimental to the independence among disentangled aspects. DCLKR performs best on all three datasets when $K = 3$.

Effect of Aggregation Depth. To study the influence of graph aggregation depth, we vary L in range of {1, 2, 3, 4} and demonstrate the performance comparison on Book-Crossing, MovieLens-1M and Last.FM in Table 4. We can observe that DCLKR substantially achieves improvements on Book-Crossing, MovieLens-1M and Last.FM when $L = 2, 3, 2$, respectively. But further stacking more layers leads to performance degradation, cause neighbors in too long distance may introduce noise to node representations.

Effect of Contrastive Loss Weights. The trade-off parameters λ_1 and λ_2 control the influence of intra-view and inter-view contrastive losses in final loss, respectively. To study the effect of contrastive loss weights, we vary both λ_1 and λ_2 in {0.001, 0.01, 0.1, 1}. According to the results shown in Fig. 5 and Fig. 6, we can observe that DCLKR performs best when $\lambda_1 = 0.1, 0.01, 0.01$ and $\lambda_2 = 0.1, 0.01, 0.01$ on Book-Crossing, MovieLens-1M and Last.FM, respectively. The intra-view contrastive learning is deployed to encourage the independence among disentangled representations from different aspects, and the inter-view contrastive learning is adopted to transfer knowledge between the knowledge and collaborative views. Tuning the contributions of these two contrastive losses to a proper degree could boost the model performance.

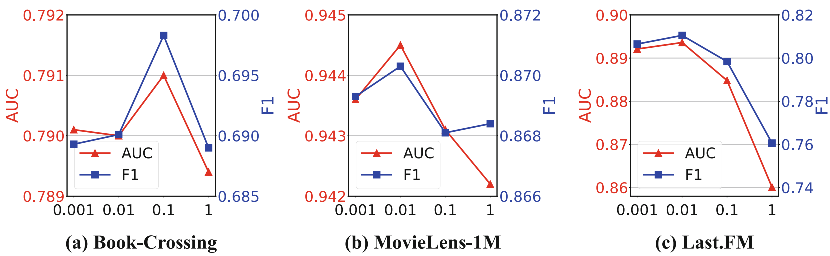

Fig. 5. Effect of intra-view contrastive loss weight λ_1.

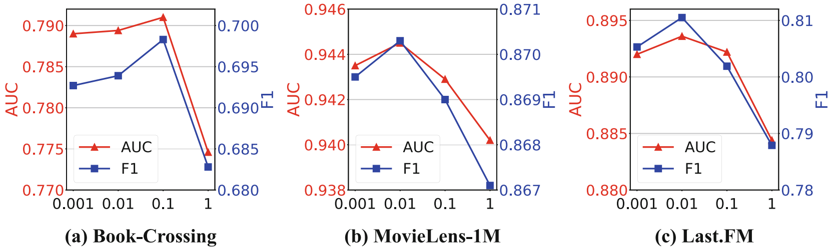

Fig. 6. Effect of inter-view contrastive loss weight λ_2.

5 Related Work

5.1 Knowledge-Aware Recommendation

In recent years, there is a surge of interest in the knowledge-aware recommendation. A typical approach is to pre-train the entity embeddings with knowledge graph embedding (KGE) algorithms [2,16,39], then incorporate them into recommendation frameworks [3,11,27,29,31,45]. CKE [45] adopts TransR [16] to encode items' semantic knowledge, and then combines knowledge embeddings, text embeddings, and image embeddings together for collaborative filtering. Besides, some methods [4,10,19,24,28,36,44] focus on exploring various patterns of connections among items to afford supplementary assistance for recommendation. RippleNet [28] propagates users' historical interacted items along with paths in KGs to explore users' potential long-range preferences via a memory-like neural model. More recently, the information aggregation mechanisms of GNNs [7,15,25] become the technical trend of knowledge-aware recommendation [12,30,32–34,38]. KGAT [33] unifies user-item interactions and KGs as user-item-entity graphs, then utilizes GCN with an attention mechanism to perform aggregation on it. But, CKAN [38] separately applies different neighbor aggregation schemes over the user-item interaction graphs and KGs. KGIN [34] disentangles user-item interactions at the granularity of user intents, and further performs the relational path-aware aggregation for both user-intent-item and KG triplets.

5.2 Disentangled Representation Learning

Disentangled representation learning aims to separate the underlying factors in the data through embedding objects from multiple perspectives [1,18], which has been applied to many fields, such as texts [13], images [5], and knowledge graph embeddings [41]. There are also some effort [20,35,40,46] has been done towards disentangled representation learning on recommendation. DGCF [35] disentangles the intents hidden in the user-item interaction graphs and learns the intent-aware disentangled representations. KDR [20] leverages KGs to guide the disentangled representation learning in recommendation, making the disentangled representations interpretable. MDKE [46] is proposed to disentangle the knowledge-aware recommendation into semantic-level and structural-level subspaces, and then utilize two levels disentangled representations to enhance recommendation. Our work considers the fact that relations and entities in KGs have different correlation with different aspects which is ignored by KDR and MDKE, and emphasizes influential relations and entities by attention mechanisms.

5.3 Contrastive Learning

Contrastive learning methods [8, 21, 26, 37] learn discriminative node representations from unlabeled data by maximizing the distance between negative pairs while minimizing the distance between positive pairs. Recently, there are several efforts [43, 47, 48] that apply contrastive learning on knowledge-aware recommendation. KGCL [43] proposes a KG augmentation schema to suppress KG noise in information aggregation to derive more robust knowledge-aware representations for items, and exploits the KG augmentation to guide cross-view contrastive learning. MCCLK [47] generates three different graph views from collaborative interactions and KGs, then performs contrastive learning across three views on both local and global levels. KGIC [48] constructs local and non-local graphs for users and items in KGs, and conducts layer-wise contrastive learning on these graphs. All the above methods conduct contrastive learning to align node representations among different views, but do not perform contrastive learning among the disentangled graphs, which helps to distinguish user preferences under different semantic aspects.

6 Conclusion

In this work, we focus on exploring user preferences from multiple aspects of item attributes, and propose a novel disentangled contrastive learning framework for knowledge-aware recommendation, DCLKR, which achieves better recommendation performance from two dimensions: (1) It disentangles the knowledge graph and the user-item interaction graph into multiple aspects, and uses attentive neighbor assignment mechanisms to highlight the importance of influential connections. (2) It performs intra-view and inter-view contrastive learning to enhance the disentangled representation learning. The experimental results on three public datasets demonstrate the superior performance of our proposed method over the state-of-the-arts.

Acknowledgments. The authors would like to thank the anonymous reviewers for their helpful comments. This work was supported by the National Nature Science Foundation of China (NSFC 61572537), and the CCF-Huawei Populus Grove Challenge Fund (CCF-HuaweiDBC202305).

References

1. Bengio, Y., Courville, A.C., Vincent, P.: Representation learning: a review and new perspectives. IEEE Trans. Pattern Anal. Mach. Intell. **35**(8), 1798–1828 (2013)

2. Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., Yakhnenko, O.: Translating embeddings for modeling multi-relational data. In: Advances in Neural Information Processing Systems, vol. 26 (2013)
3. Cao, Y., Wang, X., He, X., Hu, Z., Chua, T.S.: Unifying knowledge graph learning and recommendation: towards a better understanding of user preferences. In: The World Wide Web Conference, pp. 151–161 (2019)
4. Catherine, R., Cohen, W.: Personalized recommendations using knowledge graphs: a probabilistic logic programming approach. In: Proceedings of the 10th ACM Conference on Recommender Systems, pp. 325–332 (2016)
5. Chen, X., Duan, Y., Houthooft, R., Schulman, J., Sutskever, I., Abbeel, P.: Infogan: interpretable representation learning by information maximizing generative adversarial nets. In: Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5–10, 2016, Barcelona, Spain, pp. 2172–2180 (2016)
6. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, pp. 249–256 (2010)
7. Hamilton, W., Ying, Z., Leskovec, J.: Inductive representation learning on large graphs. In: Advances in Neural Information Processing Systems, vol. 30 (2017)
8. Hassani, K., Khasahmadi, A.H.: Contrastive multi-view representation learning on graphs. In: International Conference on Machine Learning, pp. 4116–4126 (2020)
9. 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, pp. 639–648 (2020)
10. 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, pp. 1531–1540 (2018)
11. Huang, J., Zhao, W.X., Dou, H., Wen, J.R., Chang, E.Y.: Improving sequential recommendation with knowledge-enhanced memory networks. In: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, pp. 505–514 (2018)
12. Jin, J., et al.: An efficient neighborhood-based interaction model for recommendation on heterogeneous graph. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 75–84 (2020)
13. John, V., Mou, L., Bahuleyan, H., Vechtomova, O.: Disentangled representation learning for non-parallel text style transfer. In: Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers, pp. 424–434 (2019)
14. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. In: 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7–9, 2015, Conference Track Proceedings (2015)
15. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. In: 5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24–26, 2017, Conference Track Proceedings (2017)
16. Lin, Y., Liu, Z., Sun, M., Liu, Y., Zhu, X.: Learning entity and relation embeddings for knowledge graph completion. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 29 (2015)
17. Liu, X., et al.: Self-supervised learning: generative or contrastive. IEEE Trans. Knowl. Data Eng. **35**(1), 857–876 (2023)

18. Locatello, F., et al.: Challenging common assumptions in the unsupervised learning of disentangled representations. In: Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9–15 June 2019, Long Beach, California, USA, pp. 4114–4124 (2019)
19. Ma, W., et al.: Jointly learning explainable rules for recommendation with knowledge graph. In: The World Wide Web Conference, pp. 1210–1221 (2019)
20. Mu, S., Li, Y., Zhao, W.X., Li, S., Wen, J.: Knowledge-guided disentangled representation learning for recommender systems. ACM Trans. Inf. Syst. **40**(1), 6:1–6:26 (2022)
21. Peng, Z., Huang, W., Luo, M., Zheng, Q., Rong, Y., Xu, T., Huang, J.: Graph representation learning via graphical mutual information maximization. In: Proceedings of The Web Conference, vol. 2020, pp. 259–270 (2020)
22. Rendle, S., Freudenthaler, C., Gantner, Z., Schmidt-Thieme, L.: BPR: bayesian personalized ranking from implicit feedback. In: UAI 2009, Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence, Montreal, QC, Canada, June 18–21, vol. 2009, pp. 452–461 (2009)
23. Sha, X., Sun, Z., Zhang, J.: Hierarchical attentive knowledge graph embedding for personalized recommendation. Electron. Commer. Res. Appl. **48**, 101071 (2021)
24. Sun, Z., Yang, J., Zhang, J., Bozzon, A., Huang, L.K., Xu, C.: Recurrent knowledge graph embedding for effective recommendation. In: Proceedings of the 12th ACM Conference on Recommender Systems, pp. 297–305 (2018)
25. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y., et al.: Graph attention networks. Stat **1050**(20), 10–48550 (2017)
26. Velickovic, P., Fedus, W., Hamilton, W.L., Liò, P., Bengio, Y., Hjelm, R.D.: Deep graph infomax. ICLR (Poster) **2**(3), 4 (2019)
27. 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, pp. 592–600 (2018)
28. Wang, H., et al.: RippleNet: propagating user preferences on the knowledge graph for recommender systems. In: Proceedings of the 27th ACM International Conference on Information and Knowledge Management, pp. 417–426 (2018)
29. Wang, H., Zhang, F., Xie, X., Guo, M.: Dkn: deep knowledge-aware network for news recommendation. In: Proceedings of the 2018 World Wide Web Conference, pp. 1835–1844 (2018)
30. Wang, H., et al.: Knowledge-aware graph neural networks with label smoothness regularization for recommender systems. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 968–977 (2019)
31. Wang, H., Zhang, F., Zhao, M., Li, W., Xie, X., Guo, M.: Multi-task feature learning for knowledge graph enhanced recommendation. In: The World Wide Web Conference, pp. 2000–2010 (2019)
32. Wang, H., Zhao, M., Xie, X., Li, W., Guo, M.: Knowledge graph convolutional networks for recommender systems. In: The world Wide Web Conference, pp. 3307–3313 (2019)
33. 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, pp. 950–958 (2019)
34. Wang, X., et al.: Learning intents behind interactions with knowledge graph for recommendation. In: Proceedings of the Web Conference 2021, pp. 878–887 (2021)

35. Wang, X., et al.: Disentangled graph collaborative filtering. In: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 1001–1010 (2020)
36. Wang, X., Wang, D., Xu, C., He, X., Cao, Y., Chua, T.S.: Explainable reasoning over knowledge graphs for recommendation. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp. 5329–5336 (2019)
37. Wang, X., Liu, N., Han, H., Shi, C.: Self-supervised heterogeneous graph neural network with co-contrastive learning. In: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, pp. 1726–1736 (2021)
38. Wang, Z., Lin, G., Tan, H., Chen, Q., Liu, X.: CKAN: collaborative knowledge-aware attentive network for recommender systems. In: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 219–228 (2020)
39. Wang, Z., Zhang, J., Feng, J., Chen, Z.: Knowledge graph embedding by translating on hyperplanes. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 28 (2014)
40. Wu, J., Fan, W., Chen, J., Liu, S., Li, Q., Tang, K.: Disentangled contrastive learning for social recommendation. In: Proceedings of the 31st ACM International Conference on Information & Knowledge Management, Atlanta, GA, USA, October 17–21, 2022, pp. 4570–4574 (2022)
41. Wu, J., et al.: DisenKGAT: knowledge graph embedding with disentangled graph attention network. In: CIKM '21: The 30th ACM International Conference on Information and Knowledge Management, Virtual Event, Queensland, Australia, November 1–5, vol. 2021, pp. 2140–2149 (2021)
42. Xian, Y., Fu, Z., Muthukrishnan, S., De Melo, G., Zhang, Y.: Reinforcement knowledge graph reasoning for explainable recommendation. In: Proceedings of the 42nd international ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 285–294 (2019)
43. Yang, Y., Huang, C., Xia, L., Li, C.: Knowledge graph contrastive learning for recommendation. In: SIGIR '22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, July 11–15, vol. 2022, pp. 1434–1443 (2022)
44. Yu, X., et al.: Personalized entity recommendation: a heterogeneous information network approach. In: Proceedings of the 7th ACM International Conference on Web Search and Data Mining, pp. 283–292 (2014)
45. 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, pp. 353–362 (2016)
46. Zhou, H., Liu, Q., Gao, X., Chen, G.: MDKE: multi-level disentangled knowledge-based embedding for recommender systems. In: Database Systems for Advanced Applications - 27th International Conference, DASFAA 2022, Virtual Event, April 11–14, 2022, Proceedings, Part II, pp. 3–18 (2022)
47. Zou, D., et al.: Multi-level cross-view contrastive learning for knowledge-aware recommender system. In: SIGIR '22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, July 11–15, 2022, pp. 1358–1368 (2022)
48. Zou, D., et al.: Improving knowledge-aware recommendation with multi-level interactive contrastive learning. In: Proceedings of the 31st ACM International Conference on Information & Knowledge Management, pp. 2817–2826 (2022)