Check for updates

# A Dual Perspective Framework of Knowledge-correlation for Cross-domain Recommendation

YUHAN WANG and QING XIE, Wuhan University of Technology, Wuhan, China, Engineering Research Center of Digital Publishing Intelligent Service Technology, Ministry of Education, Wuhan, China, and WHUT Chongqing Research Institute, Chongqing, China

MENGZI TANG, LIN LI, and JINGLING YUAN, Wuhan University of Technology, Wuhan, China and Engineering Research Center of Digital Publishing Intelligent Service Technology, Ministry of Education, Wuhan, China

YONGJIAN LIU, Wuhan University of Technology, Wuhan, China, Engineering Research Center of Digital Publishing Intelligent Service Technology, Ministry of Education, Wuhan, China, and WHUT Chongqing Research Institute, Chongqing, China

Recommender System provides users with online services in a personalized way. The performance of traditional recommender systems may deteriorate because of problems such as cold-start and data sparsity. Cross-domain Recommendation System utilizes the richer information from auxiliary domains to guide the task in the target domain. However, direct knowledge transfer may lead to a negative impact due to data heterogeneity and feature mismatch between domains. In this paper, we innovatively explore the cross-domain correlation from the perspectives of content semanticity and structural connectivity to fully exploit the information of Knowledge Graph. First, we adopt domain adaptation that automatically extracts transferable features to capture cross-domain semantic relations. Second, we devise a knowledge-aware graph neural network to explicitly model the high-order connectivity across domains. Third, we develop feature fusion strategies to combine the advantages of semantic and structural information. By simulating the cold-start scenario on two real-world datasets, the experimental results show that our proposed method has superior performance in accuracy and diversity compared with the SOTA methods. It demonstrates that our method can accurately predict users' expressed preferences while exploring their potential diverse interests.

CCS Concepts: • **Information systems** → **Recommender systems**; • **Computing methodologies** → *Neural networks.*

Additional Key Words and Phrases: Cross-domain recommendation, Knowledge graph, Cold-start, Graph neural network, Domain adaptation

Authors' addresses: Yuhan Wang, wyh0520@whut.edu.cn; Qing Xie, Wuhan University of Technology, Wuhan, China, 430070 and Engineering Research Center of Digital Publishing Intelligent Service Technology, Ministry of Education, Wuhan, China, 430070 and WHUT Chongqing Research Institute, Chongqing, China, 400000, felixxq@whut.edu.cn; Mengzi Tang, tangmz@whut.edu.cn; Lin Li, cathylilin@whut.edu.cn; Jingling Yuan, liuyj@whut.edu.cn, Wuhan University of Technology, Wuhan, China, 430070 and Engineering Research Center of Digital Publishing Intelligent Service Technology, Ministry of Education, Wuhan, China, 430070, yjl@whut.edu.cn; Yongjian Liu, Wuhan University of Technology, Wuhan, China, 430070 and Engineering Research Center of Digital Publishing Intelligent Service Technology, Ministry of Education, Wuhan, China, 430070 and WHUT Chongqing Research Institute, Chongqing, China, 400000, liuyj@whut.edu.cn.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM 1556-4681/2024/3-ART

https://doi.org/10.1145/3652520

Just Accepted

## 1 INTRODUCTION

As a powerful solution to *information overload*, traditional recommender systems mainly focus on analyzing users' historical behaviors in a specific domain (e.g., movies, books, or music). Recently, there has been growing concern in Cross-Domain Recommendation Systems (CDRS), which link or transfer knowledge (e.g., user interests or item attributes) from the source domain to enrich the data in the target domain. CDRS aims to improve recommendation performance by sharing and complementing information from different domains, giving it more opportunities to explore users' broad interests and improving comprehensive recommendation performance (e.g., accuracy, diversity, and serendipity).

Since different domains usually have inconsistent data structures and heterogeneous knowledge content, it is difficult to generalize the source data directly to the target domain. Therefore, the mining of inter-domain correlations is the hinge of cross-domain recommendation methods. One solution is knowledge linking-based CDRS, which uses common knowledge to establish feature relationships between domains[5]. Such common knowledge used in the methods can be tags[32], association rules[1], and semantic networks[6, 19, 26]. However, despite the encouraging results from these methods, most are only suitable for specific scenarios due to the need for manual pre-setting.

Knowledge Graph (KG) is built based on massive amounts of real-world facts. Encyclopedic KG provides multi-relational data correlated with diverse domains, which can effectively deal with cross-domain knowledge heterogeneity. In addition, the ontological nature and standard structured format of KG data allow CDRS to explore inter-domain relationships at the semantic level without extensive manual processing. There are several KG-based cross-domain methods [7, 31] that compute semantic similarity in an implicit manner (e.g., joint matrix factorization) to enhance preference analysis. However, these methods ignore the well-organized structural information of KG, making it difficult for them to capture comprehensive item characteristics and user preferences. Therefore, for effective cross-domain correlation and knowledge augmentation, we introduce KG and propose to leverage the knowledge from two perspectives, i.e., content semanticity and structural connectivity. Specifically, we map the recommendation items to the corresponding entities in KG and then query the entity-related information using Semantic Web standards (i.e., RDF[1] and SPARQL[2]). For example, we query DBpedia[3] for information describing the item in book domain *Harry Potter*, and Fig. 1 shows part of the queried information. We can see that there is a lot of semantic content related to the item: the subject includes "Magic" and "Curse", and its theme involves "Fantasy literature". In addition, KG describes the relevance through the edges between nodes in the graph. This structured format makes it possible to link items in other domains (e.g., the movie *Strike*) beyond the same domain (e.g., the book *The Wonderful Wizard of Oz*), thus discovering the cross-domain correlation. This example demonstrates that KG not only provides rich semantic information but also reveals structural information across domains. This enlightens us to design the framework from the perspective of content semanticity and structural connectivity, and we give two formal definitions:

*Definition 1.1 (Content semanticity).* Content semanticity focuses on descriptions (i.e. properties, classes, categories) that reflect the similarity of entities. Entities whose descriptions are more semantically similar are more closely associated. For example, semantic information can associate a "comedy" movie with a "humorous" book.

*Definition 1.2 (Structural connectivity).* Structural connectivity focuses on graph links that reflect the relatedness of entities. Entities with more links in the graph are more closely associated. For example, KG structured the information that *J.K. Rowling* is not only the author of the book *Harry Potter* but also the screenwriter of the movie *Fantastic Beasts*, showing that the two cross-domain items are related.

[1]https://www.w3.org/TR/rdf-primer
[2]http://www.w3.org/TR/rdf-sparql-query
[3]http://dbpedia.org

Just Accepted

Due to the inherent properties of knowledge obtained from different perspectives, the effects of different types of knowledge in exploring users' domain-specific preferences tend to be distinctive. Our previous work[43] investigates the effect of different types of knowledge in discovering cross-domain user interests. It suggests that the semantic information performs well in the accuracy of user preference prediction, while the structural information can bring more diverse recommendations. Motivated by this, we further design the feature fusion strategy that combines the advantages of both types of information to improve the recommendation performance.

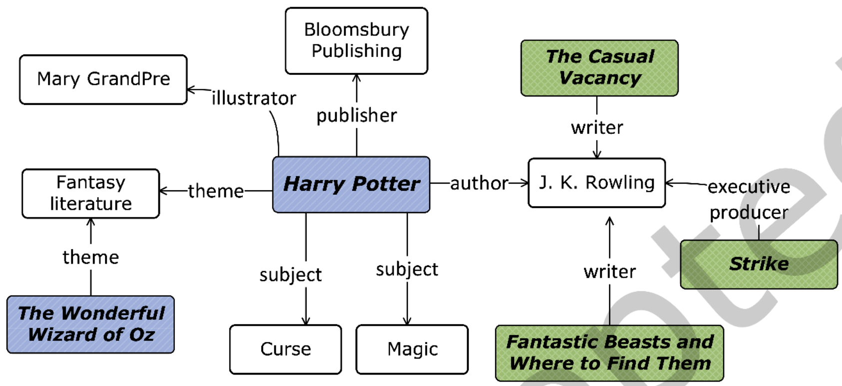



Fig. 1. Partial information display of the example of correlated knowledge based on KG. Blue twill nodes represent book items and green mesh nodes represent movie items.

Just Accepted

Based on the above pre-research, we propose a knowledge-aware cross-domain recommendation method to fully explore the information in KG. The main contributions of this work are as follows:

- By highlighting the advantages of leveraging KG to enhance CDRS, we innovatively propose to mine knowledge correlations from the perspectives of content semanticity and structural connectivity.
- From the perspective of content semanticity, we develop a novel model named Domain Adaptation-based Semantic Feature Extraction (DASFE). Considering the differences in data distribution across domains, this model unifies cross-domain semantic information into the same feature space to avoid negative knowledge transfer. Under the idea of adversarial training, DASFE enables the automatic organization and reasonable correlation of semantic concepts across domains in an unsupervised manner. Besides, we specifically consider the large-scale and multi-label characteristics of KG data.
- From the perspective of structural connectivity, we further develop a Cross-Domain Knowledge Graph Attention Network (CD-KGAT). It jointly models cross-domain high-order relations in our designed knowledge-aware graph structure to relate domains with heterogeneous data. Moreover, we specially devise feature fusion strategies to combine the advantages of semantic and structural information, which aims at enhancing the representation learning ability of our method and thus improving the recommendation performance.
- By conducting extensive experiments on two real-world datasets, we demonstrate the comprehensive superiority of our method. We also perform a series of detailed analyses to validate the effectiveness of our design.

The remaining part of the paper is organized as follows: Section 2 reviews related work. Section 3 presents our proposed method in 3 parts: the overview and problem formation, the semantic extraction model, and our proposed cross-domain recommendation method. Next, section 4 reports and compares the empirical results achieved by the proposed method in a cold-start situation, and we further experimentally verify the design effectiveness of the proposed method. Finally, we end with conclusions and future research directions in Section 5.

## 2 RELATED WORK

In this section, we review three CDRS subfields relevant to our research.

### 2.1 Collaborative filtering-based cross-domain recommendation methods

Collaborative filtering (CF) provides recommendations for target users by analyzing the historical behaviors of similar users[21]. Some early approaches[4, 16] directly merge rating matrices of different domains, treating the task as single-domain recommendations. While easy to implement, this idea requires consistent rating patterns, and ignoring cross-domain differences makes it unreliable in some scenarios. Wang et al.[40] propose a solution that considers domain-specific rating patterns. The model constructs triple-bridge transfer based on latent factors, rating patterns, and adjacency graphs, thereby promoting positive transfer across domains. Singh et al.[33] propose the idea of joint matrix factorization, which adopts stochastic approximations to share the parameters of the rating matrix in different domains. In addition, CDTF[15] uses the "user-item-domain" triadic relation to capture the user preferences in different domains. It factorizes the tensors constructed based on genetic algorithms to obtain domain-specific features. However, most of the aforementioned works are only applicable to scenarios with fully overlapping users. Zhang et al.[47] use partially overlapping items as the bridge for knowledge transfer and adopt diffusion kernel completion to solve cross-domain feature heterogeneity.

The advantages of CF-based methods are flexibility and simplicity, which make full use of the common features across domains. However, these methods are usually susceptible to the degree of information overlapping and the data density in the target domain. In contrast, our method is less limited by introducing auxiliary information KG to mine domain correlations from two perspectives that can complement each other, thus being able to provide effective recommendations for cold-start users in the target domain. Besides, our method employs a GNN for information propagation and aggregation on the cross-domain graph, which is capable of capturing higher-order collaborative information.

### 2.2 Semantic mapping-based cross-domain recommendation methods

Yang et al.[45] explore the semantic relations contained in tags, and introduce social influence and external knowledge base to address the isolation and sparsity between tags in different domains. Its proposed local tag propagation algorithm leverages optimized explicit semantic analysis to measure the correlation of cross-domain tags. However, this tag-based approach can not fit domains with unannotated or unstructured information. Therefore, Kumar et al.[22] use Latent Direchlet Allocation, a topic modeling technique applicable to structured and unstructured textual, to deduce cross-domain semantic correlation. In addition, some approaches[6, 19, 26] build semantic networks for cross-domain recommendations. For example, Fernández-Tobías et al.[6] mine item-relevant information from external knowledge repositories to build a semantic network. It utilizes a graph-based weight spreading mechanism to perform the concept linking of two domains, hence recommending music artists according to users' Point of Interest (POI).

Semantic mapping-based methods have low dependence on ratings and can handle data sparsity and cross-domain data heterogeneity well. However, such methods require laborious human efforts, such as performing manual presetting or filtering reliable correlation medium. In comparison, our method satisfies the generality by introducing an encyclopedic knowledge graph containing facts from multiple domains.



Just Accepted

## 2.3 Deep learning-based cross-domain recommendation methods

Due to the superior representation learning ability, more recent attention has focused on the deep learning-based cross-domain recommendation. Existing deep cross-domain methods fall into three main categories: feature mapping-based methods, adversarial learning-based methods, and network-based methods.

Feature mapping-based methods[9, 44, 50] unify information from different domains by aligning user and item features in the same space. For example, Xu et al.[44] propose a multi-layer perceptron-based mapping method, which transfers the user's sentiment-aware features from the source domain to the target domain. The second category of methods[18, 46] introduces adversarial learning for cross-domain recommendation. For example, Yuan et al.[46] propose a cross-domain recommendation method based on domain adaptation. It interleaves the features of the source and target domains extracted by an autoencoder and then feeds them into an adversarial structure to obtain the rating pattern of the target domain. The third category is network-based deep methods, such as the multi-view-based model MVDNN[11] that jointly learns features from different domains, and TMH[14] that builds a memory network and transfer network for modeling and transfer, respectively. In recent years, there has been attention to exploring cross-domain recommendations based on Graph Neural Network (GNN). PPGN[48] construct a unified graph of user-item interactions in multiple domains. By propagating the users' preference on the graph, it aims to maintain high-order structural information and explicitly model cross-domain interactions. Despite good performance, PPGN does not exploit possible auxiliary information.

Compared to traditional methods, these deep methods have fewer constraints on input data and do not require heavy manual processing. Deep computing, especially GNN, has demonstrated powerful capabilities in CDRS, where learned informative features can enhance the recommendation system. However, existing GNN-based methods focus on capturing structural information in graphs but rarely decompose semantic correlations, making it difficult for them to capture potential content-based similarities between items or users. In comparison, our method proposes to learn domain correlation from both semantic and structural perspectives, aiming to capture comprehensive item characteristics and user preferences.

# 3 METHODOLOGY

This section first introduces our cross-domain recommendation method a clear description of the problem setting and an overview of the procedure. Then we elaborate on the main components and principles of our method in detail.

## 3.1 Overview and problem formulation

In this study, we use $D_S$ and $D_T$ to denote the source domain and the target domain, respectively. The user set and item set in $D_S$ are represented as $U_S = \left\{u_1^S, u_2^S, \dots, u_{M_S}^S\right\}$ and $I_S = \left\{i_1^S, i_2^S, \dots, i_{N_S}^S\right\}$, respectively. The user set and item set in $D_T$ are represented as $U_T = \left\{u_1^T, u_2^T, \dots, u_{M_T}^T\right\}$ and $I_T = \left\{i_1^T, i_2^T, \dots, i_{N_T}^T\right\}$, where $U_T \cap U_S \neq \emptyset$ and $I_T \cap I_S = \emptyset$. $M_S$ and $N_S$ represent the number of items and users in $D_S$. $M_T$ and $N_T$ represent the number of users and items in $D_T$. The goal of our research is to link domains by learning inter-domain correlations. For a user $u$, we analyze his historical interactions $I(u) = \left\{i_1^S, \dots, i_n^S\right\}$ in $D_S$ to predict his preference for unobserved items in $I_T$. In addition, we formally define KG as follows:

*Definition 3.1 (Knowledge graph).* Knowledge graph is a heterogeneous graph[10], where nodes represent real-world entities (e.g. a book, author, publisher) and edges represent the relations. Given $E$ as the entities set and $R$ as the relations set. Formally define the KG as $G_{kg} = \{(h, r, t) \mid h, t \in E, r \in R\}$, where the triple $(h, r, t)$ represents an instance of "entity-relation-entity" in KG, such as "*Harry Potter*, author, *J.K. Rowling*".



Just Accepted

The overall procedure of our method is shown in Fig. 2, with an emphasis on the Latent Factors part. The KG-based Latent Factors learning consists of three components: 1). Semantic feature extraction model, which extracts transferable semantic features based on domain adaptation technique. 2). Connectivity feature extraction, which exploits the high-order structural relations of a cross-domain graph based on Graph Attention Network. 3). The fusion strategy combines semantic information and structural information to complete item features. The first function is implemented by DASFE and the next two functions are implemented by CD-KGAT. Finally, our method predicts recommendations in the target domain based on the learned user and item features.

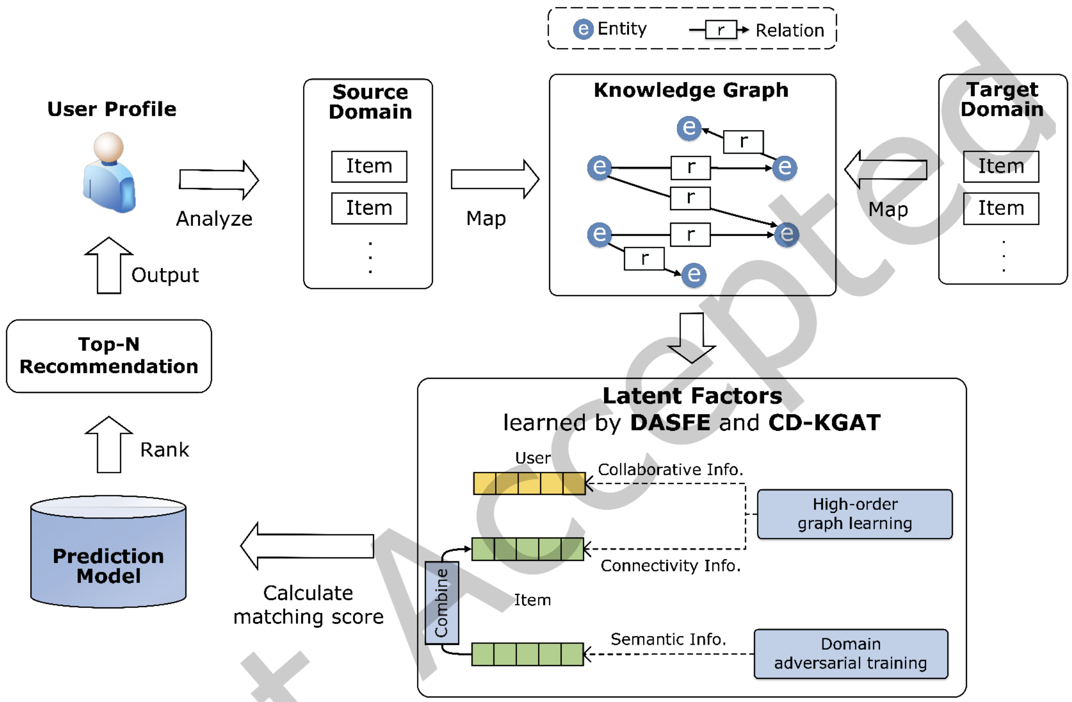



Fig. 2. Overview of Knowledge-correlated Cross-domain Recommendation.

Just Accepted

## 3.2 Domain adaptation-based semantic feature extraction model

Domain adaptation is able to learn a mapping in the presence of a shift between the training and test distributions. Since cross-domain data usually have inconsistent distributions, we employ domain adaptation techniques for unification to achieve effective knowledge transfer.

*Definition 3.2 (Domain adaptation).* Given $\mathbb{X}$ as the input feature space and $\mathbb{Y}$ as the set of labels. There are two different distributions on $\mathbb{X} \times \mathbb{Y}$, denoted as $P_S$ for the source domain and $P_T$ for the target domain. An unsupervised domain adaptation training is based on the *labeled source sample data* $\mathrm{X_S}$ from $P_S$, and the *unlabeled target sample data* $\mathrm{X_T}$ from $P_T^{\mathbb{X}}$, where $P_T^{\mathbb{X}}$ represents the marginal distribution of $P_T$ over $\mathbb{X}$.

$$\mathrm{X_S} = \left\{\left(\mathrm{x_i^S}, \mathrm{y_i^S}\right)\right\}_{\mathrm{i=1}}^{\mathrm{N_S}} \sim \mathrm{P_S}; \quad \mathrm{X_T} = \left\{\left(\mathrm{x_i^T}\right)\right\}_{\mathrm{i=1}}^{\mathrm{N_T}} \sim \mathrm{P_T^{\mathbb{X}}} \tag{1}$$

where $x_i \in \mathbb{X}$ represents the content description of item, and $y_i^S \in \mathbb{Y}$ represents the category label of the item (e.g., "Action", "Comedy"). $N_S$ and $N_T$ represent the amount of data in $\mathrm{X_S}$ and $\mathrm{X_T}$. The goal of domain adaptation is to make the distribution $P_S$ and $P_T$ the same or very similar so that it is feasible to learn a classifier or predictor applicable to the unlabeled target domain, which is not hindered by shifts between domains.

To fully explore the correlation of semantic knowledge between domains and avoid negative transfer, inspired by the domain adaptation approach DANN[8], we propose the semantic feature extraction model DASFE for the recommendation scenario. Specifically, following adversarial training-style convention, we use a paradigm of encode-decode to map heterogeneous cross-domain knowledge into the same feature space. The framework of DASFE is shown in Fig. 3, which includes three modules: feature extractor, label predictor, and domain classifier.

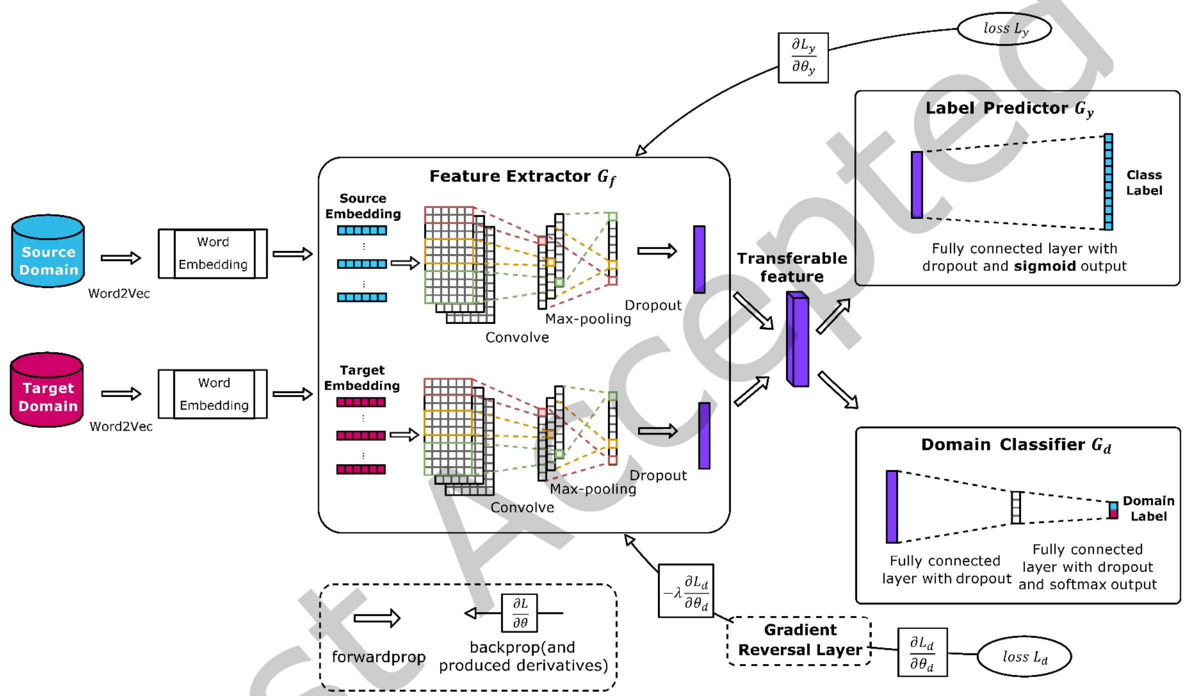



Fig. 3. The framework of DASFE.

Just Accepted

*3.2.1 Feature extractor based on contextual representation learning.* The feature extractor employs contextual representation learning to extract semantic information from the content descriptions of items. Given an item $v$, we obtain the item embedding from its content information through the word embedding technique Word2Vec[28], as follows:

$$e_{1:n} = \{e_1, e_2 \dots e_n\} \tag{2}$$

where $n$ indicates the number of words in the content information, and $e_i \in \mathbb{R}^k$ denotes the $k$-dimensional word vector of the $i$-th word.

Text Convolutional Neural Network (Text-CNN)[20], as an automatic feature extractor, utilizes multi-layer nonlinear transformations to extract features with strong knowledge representation ability. Therefore, we input the obtained item embedding into Text-CNN to further explore the semantic information. Specifically, the convolution kernel $w \in R^{hk}$ is applied to a text region to generate feature $c_i$:

$$c_i = Relu\left(w^T e_{i:i+h-1} + b\right) \tag{3}$$

where $e_{i:i+h-1}$ represents the text region containing $h$ words. $Relu(\cdot)$ is a nonlinear activation function, and $b \in \mathbb{R}$ represents the bias term. Then we extend to all possible text regions in the content information to construct the feature map:

$$c = \{c_i, c_{i+1}, \ldots, c_{n-h+1}\} \tag{4}$$

The next pooling operation performs downsampling to obtain important features of the original samples. Assuming that $t$ kinds of feature maps $c^{(1)}, c^{(2)}, \ldots, c^{(t)}$ are generated through $t$ convolution kernels, the item's semantic embedding is:

$$e^{semantic} = Dropout\left\{P\left(c^{(1)}\right), P\left(c^{(2)}\right), \ldots, P\left(c^{(t)}\right)\right\} \tag{5}$$

where $P(\cdot)$ represents max-pooling. $Dropout$[12] regularizes the pooled vectors to avoid the overfitting.

*3.2.2 Label predictor for multi-label classification with class imbalance.* We perform in an encoding-decoding paradigm to learn cross-domain item semantic features from KG. The feature extractor acts as an encoder to generate embedding containing content semanticity; the label predictor acts as a decoder for restoring the original information, which trains a classifier capable of predicting the category (i.e., genre, subject) of an item based on the generated embedding.

Text data are commonly annotated with more than one label[35]. For example, the subjects of the movie *Toy Story* include "Fantasy", "Comedy", and "Adventure". Since KG is a large-scale knowledge base, the label predictor needs to select a subset from the set with hundreds or thousands of labels when predicting relevant categories of an item. Therefore, our label predictor faces the multi-label classification task. In addition, the set of category labels usually has a long-tailed distribution, which means that popular labels appear much more frequently than unpopular ones. For example, in a movie dataset, the number of "Action" items far exceeds that of "Environmental-friendly" movies. If we directly feed class-imbalanced data into the model, the output will be biased towards more popular labels resulting in poor training results.

We consider improving the model from the loss function for the multi-label classification with class imbalance issues. Different from the previous work[23], our DASFE adopts Focal Loss (FL)[24] as the loss function of the label predictor. FL specifically designs the weights based on the "hardness" of labeling the sample by the model, aiming to balance the hard-to-labeling samples and easy-to-labeling samples. The definition of FL is:

$$L_{FL} = \alpha^k \left(1 - p_i^k\right)^{\gamma} \log\left(p_i^k\right) \tag{6}$$

where $p_i^k$ is the model's estimated probability that the training sample $i$ belongs to class $k$. The focusing parameter $\gamma$ adjusts the rate at easy examples that are downweighted. $\alpha^k$ balances the importance of different classes of samples, and its value is between 0 and 1.

*3.2.3 Feature alignment based on domain adversarial training.* We employ domain adaptation to extract features automatically while aligning the embeddings of semantically similar cross-domain items, aiming to correlate heterogeneous knowledge across domains. Next, we describe the optimization objectives and training process of DASFE.

As shown in Fig. 3, DASFE consists of three modules: the feature extractor extracts item semantic features from KG, denoted as $G_f\left(x; \theta_f\right)$; the label predictor performs multi-label classification on the semantic features extracted by $G_f$ and outputs the probability that an item contains the label $y$, denoted as $G_y\left(G_f\left(x\right); \theta_y\right)$; the domain classifier executes a binary classification task to distinguish the domains of samples, denoted as $G_d\left(G_f\left(x\right); \theta_d\right)$.



Just Accepted

In addition, to enable DASFE to learn semantic-level features with **discriminativeness** (for the main learning task on the source domain) and **domain-invariance** (for the domain identification task), Gradient Reversal Layer (GRL)[8] is set between $G_f$ and $G_d$. GRL reduces the domain sensitivity of the model during representation learning in an adversarial training manner.

The label predictor $G_y$ ensures the learned semantic features have **discriminativeness** (i.e., accurately classify items). Given a source domain sample $j = \left(x_j, y_j\right)$, the loss function of $G_y$ based on Eq. (6) is as follows:

$$\begin{aligned} L_y^j\left(\theta_f, \theta_y\right) = & -\alpha^{y_j}\left(1 - G_y\left(G_f(x); \theta_y\right)_{y_j}\right)^{\gamma} \\ & \cdot \log G_y\left(G_f(x); \theta_y\right)_{y_j} \end{aligned} \tag{7}$$

where $L_y^j\left(\theta_f, \theta_y\right) = L_y\left(G_y\left(G_f\left(x_j; \theta_f\right); \theta_y\right), y_j\right)$, and the conditional probability of the item category label is $G_y\left(G_f(x); \theta_y\right) = G_y\left(G_f\left(x_j; \theta_f\right); \theta_y\right)$.

For the semantic features output by the feature extractor, the domain classifier $G_d$ constrains them to have **domain-invariance** (i.e., cannot distinguish the domain of information), thus aligning the cross-domain feature distribution. The loss function of $G_d$ is as follows:

$$\begin{aligned} L_d^i\left(\theta_f, \theta_d\right) = & d_i \log \frac{1}{G_d\left(G_f\left(x_i; \theta_f\right); \theta_d\right)} \\ & + (1 - d_i) \log \frac{1}{1 - G_d\left(G_f\left(x_i; \theta_f\right); \theta_d\right)} \end{aligned} \tag{8}$$

where $L_d^i\left(\theta_f, \theta_d\right) = L_d\left(G_d\left(G_f\left(x_i; \theta_f\right); \theta_d\right), d_i\right)$. $d_i$ represents the domain label of sample $i$. $d_i = 0$ means sample $i$ belongs to the source domain, otherwise $d_i = 1$.

DASFE jointly optimizes the label predictor loss $L_y\left(\theta_f, \theta_y\right)$ and the domain classifier loss $L_d\left(\theta_f, \theta_d\right)$:

$$\begin{aligned} \mathbb{E}\left(\theta_f, \theta_y, \theta_d\right) = & \frac{1}{N_s} \sum_{i=1}^{N_s} L_y^i\left(\theta_f, \theta_y\right) \\ & - \lambda\left(\frac{1}{N_s + N_T} \sum_{i=1}^{N_s+N_T} L_d^i\left(\theta_f, \theta_d\right)\right) \end{aligned} \tag{9}$$

where the domain adaptation parameter $\lambda$ is used to weigh the importance of the two modules in the objective function. The saddle point is determined by:

$$\hat{\theta}_f, \hat{\theta}_y = \underset{\theta_f, \theta_y}{\operatorname{argmin}} \mathbb{E}\left(\theta_f, \theta_y, \hat{\theta}_d\right) \tag{10}$$

$$\hat{\theta}_d = \underset{\theta_d}{\operatorname{argmax}} \mathbb{E}\left(\hat{\theta}_f, \hat{\theta}_y, \theta_d\right) \tag{11}$$

GRL enables end-to-end training for DASFE that contains multiple modules. We formalize GRL as a "pseudo-function" $R_\lambda(x)$ with adaptation parameter $\lambda$, where GRL performs identity transformation during the forward propagation and automatically changes the gradient direction during backpropagation:

$$R_\lambda(x) = x, \quad \frac{dR_\lambda}{dx} = -\lambda I \tag{12}$$



Just Accepted

The stochastic gradient descent algorithm guides the parameter update of DASFE, as follows:

$$\begin{aligned}
\theta_f &\leftarrow \theta_f - \mu \left( \frac{\partial L_y^i}{\partial \theta_f} - \lambda \frac{\partial L_d^i}{\partial \theta_f} \right) \\
\theta_y &\leftarrow \theta_y - \mu \frac{\partial L_y^i}{\partial \theta_y} \\
\theta_d &\leftarrow \theta_d - \mu \lambda \frac{\partial L_d^i}{\partial \theta_d}
\end{aligned} \tag{13}$$

where $\mu$ is the learning rate.

In addition, we adopt the dynamic parameter $\lambda$ according to the iterative process[8], to enhance the model adaptability and avoid the noise in the early stage of training. The learning process is summarized in Algorithm 1[4].

**Algorithm 1:** Domain adaptation-based semantic feature extraction

**Input:** Source samples $X_S = \left\{ \left( x_i^S, y_i^S \right) \right\}_{i=1}^{N_S}$, Target samples $X_T = \left\{ \left( x_i^T \right) \right\}_{i=1}^{N_T}$
**Output:** Model parameters $\theta_f, \theta_y, \theta_d$

Initialize $x_i^S, \theta_y, \theta_d$;
**while** *stopping criterion is not met* **do**
  **for** *each* $x_i \in \mathrm{X_S} \cup \mathrm{X_T}$ **do**
    Initialize word embedding of $x_i$ using Word2Vec via Eq. (2);
    $e_{1:n} \leftarrow Word2Vec(x_i)$;
    Construct semantic features $e_i^{semantic}$ in feature extractor $G_f$ with $\theta_f$ via Eq. (3) - (5) and the Dropout operation;
    # Analyze the categories of items in label predictor $G_y$;
    $G_y \left( G_f \left( x_i \right); \theta_y \right) \leftarrow \mathrm{sigm} \left( fc \left( e_i^{\mathrm{semantic}} \right) \right)$;
    # Adapt domain regularizer in domain classifier $G_d$;
    $G_d \left( G_f \left( x_i \right); \theta_d \right) \leftarrow \mathrm{softmax} \left( fc_2 \left( fc_1 \left( e_i^{\mathrm{semantic}} \right) \right) \right)$ ;
  **end**
  Compute the total loss via Eq. (9) – (8);
  Update $\theta_f, \theta_y, \theta_d$ via Eq. (13);
**end**
**return** $\theta_f, \theta_y, \theta_d$;

### 3.3 Cross-domain knowledge graph attention network

KG not only contains massive semantic information but also stores the data in a structured format. The previous model DASFE focuses on learning semantic information. Next, we concentrate on explicit structural connectivity across domains to take full advantage of the data available in KG.

We propose the Cross-Domain Knowledge Graph Attention Network (CD-KGAT), which aims to capture the high-order connectivity in a knowledge-aware graph neural network. CD-KGAT jointly models the user-item interactions and KG information, to understand user interests and item relations. In addition, we incorporate

[4]In the pseudo-code, $fc(y)$ refers to the full connection layer. $fc_1$ and $fc_2$ are the same as above.

Just Accepted

semantic information in the representation learning of graph data, aiming to complete item profiles. As illustrated in Fig. 4, CD-KGAT consists of four modules: cross-domain collaborative knowledge graph embedding layer, attentive neighbor aggregation layer, feature fusion layer, and prediction layer.

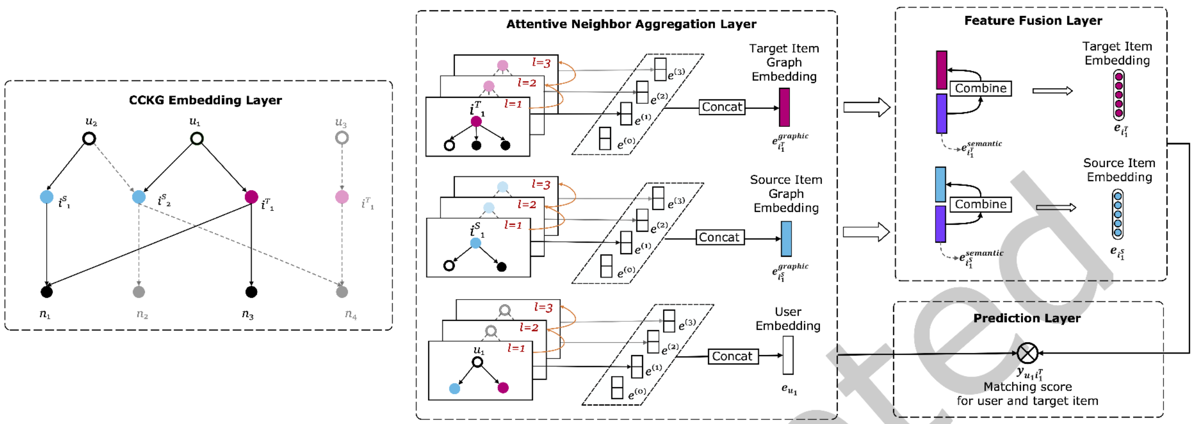



Fig. 4. The framework of CD-KGAT.

*3.3.1 Cross-domain collaborative knowledge graph.* First, we build a unified graph, Cross-domain Collaborative Knowledge Graph (CCKG), which is based on the idea of collaborative filtering with KG as auxiliary information. CCKG links users and items in different domains, aiming to break down the independence between user interactions and enable the correlation of cross-domain knowledge. We define the user-item bipartite graph and CCKG as follows.

*Definition 3.3 (User-item bipartite graph).* Given $U$ and $I$ as the user and item sets, respectively. Define the user-item bipartite graph $G_{ui} = \{(u, y_{ui}, i) \mid u \in U, i \in I\}$, where the triple $(u, y_{ui}, i)$ represents the interaction between user $u$ and item $i$. And $y_{ui} \in \{Interact\}$, where $Interact = 1$ indicates that user $u$ has interacted with (purchased, clicked, or watched, etc.) item $i$, otherwise $Interact = 0$.

*Definition 3.4 (Cross-domain Collaborative Knowledge Graph).* As shown in Fig. 5, the user-item bipartite graph and KG form a unified CCKG. Based on Definition 3.1 and Definition 3.3, We formally define CCKG as $G = \{(h, r, t) \mid h, t \in E', r \in R'\}$, where $E' = E \cup U, R' = R \cup \{Interact\}$. The item nodes in $G_{ui}$ are aligned (mapped) to the corresponding entity nodes in $G_{kg}$, and these nodes act as bridge points connecting $G_{ui}$ with $G_{kg}$ to form $G$.

CCKG combines user behavior and external knowledge. If we only consider the user-item bipartite graph containing collaborative information, for user $u_1$ in Fig. 5, it can only recommend the book *The Snow Queen* ($i_1^T$) based on the path $p_1 = u_1 \xrightarrow{r_1} i_2^S \xrightarrow{-r_1} u_2 \xrightarrow{r_1} i_1^T$. This approach requires a certain amount of overlapping users across domains. In contrast, our proposed CCKG provides possibilities for discovering users' multifaceted preferences. For example, the long-distance path $p = u_1 \xrightarrow{r_1} i_2^S \xrightarrow{r_5} n_4 \xrightarrow{-r_4} i_2^T$ indicates the scriptwriter *J.K. Rowling* ($e_1$) of the movie $i_2^S$ that $u_1$ likes is also the author of the book $i_2^T$, so $u_1$ may be more interested in this book. CCKG fully explores high-order connectivity similar to the path $p$ and links different domains while considering collaborative information.

Knowledge Graph Embedding parameterizes entities and relations into low-dimensional dense vectors in a specific space, while preserving the original structure of the graph[41]. We adopt TransR[25] to represent



Just Accepted

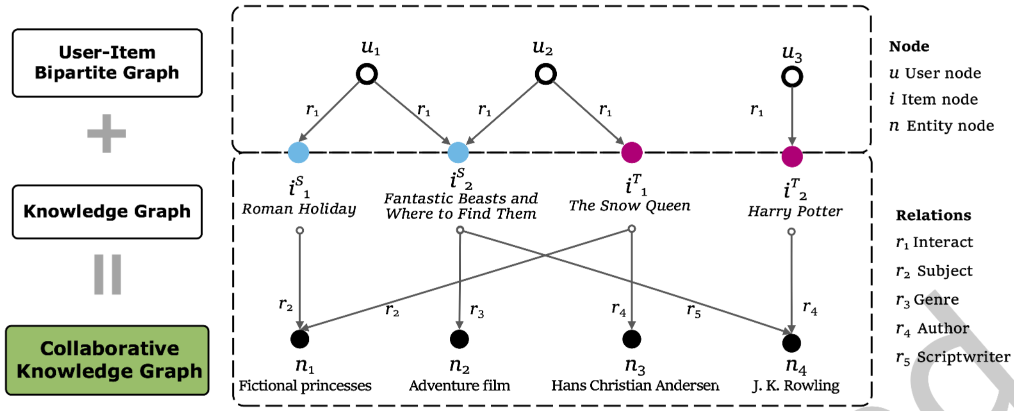



Fig. 5. An example of a cross-domain collaborative knowledge graph.

the structural information in CCKG. TransR is the embedding technique common in knowledge representation learning, and its principle is based on the structure of KG triples. Given a triple $(h, r, t) \in G$, TransR makes the learned embedding consistent with $e_h + e_r \approx e_t$. The scoring function measures the reality of $(h, r, t)$ as follows:

$$Score(h, r, t) = \|W_r e_h + e_r - W_r e_t\|_2^2 \tag{14}$$

where $W_r$ is the matrix for mapping entities to the relation space of $r$.

We take embedding learning as a ranking task that compares the relative positions of samples. The positive samples $(h, r, t) \in G$ represents the facts in KG; while the negative sample $(h, r, t') \notin G$ replaces the tail node with a randomly selected entity $t'$. The loss function $L_{KG}$:

$$L_{KG} = \sum_{(h,r,t,t')} -\ln \sigma \left(Score\left(h, r, t'\right) - Score(h, r, t)\right) \tag{15}$$

where $\sigma(\cdot)$ is the sigmoid activation function.

*3.3.2 Attentive neighbor aggregation layer.* As shown in Fig. 4, we design the Attentive Neighbor Aggregation Layer to capture the high-order connectivity across domains, which is motivated by the work[42]. The main idea is to propagate information following the path in the graph and recursively aggregate the information from multi-hop neighbors. However, for a central node, the number of its multi-hop neighbors increases sharply as the order increases, and the information contributions of these neighbors vary with the correlation between nodes. So we adopt attention mechanism to distinguish the importance of neighbor information and thus explore more relevant information about the central node.

The Attentive Neighbor Aggregation strategy consists of three steps: information propagation, attention calculation, and information aggregation. For a central node $h$, its first-order connected structure is defined as $N_h = \{(h, r, t) \mid (h, r, t) \in G\}$, and its embedding $e^{(0)}_{N_h}$ is calculated as follows:

$$e^{(0)}_{N_h} = \sum_{(h,r,t) \in N_h} \alpha(h, r, t) e_t \tag{16}$$

where $\alpha(h, r, t)$ is the weight to control the information propagated from entity $t$ to entity $h$. We adopt the attention mechanism to learn the weight $\alpha(h, r, t)$ as in Eq. (17). Note that the normalized $\alpha(h, r, t)$ is beneficial

Just Accepted

to find useful propagation information for each specific node.

$$\alpha'(h,r,t) = (W_r e_t)^T \tanh(W_r e_h + e_r)$$
$$\alpha(h,r,t) = \frac{\exp(\alpha'(h,r,t))}{\sum_{(h,r',t')\in N_h}\exp(\alpha'(h,r',t'))} \tag{17}$$

where $\tanh(\cdot)$ is the activation function.

Next, neighbor information aggregation updates the embedding of the central node. Node $h$ obtains the embedding $e_h^{(1)}$ through first-order information:

$$e_h^{(1)} = f\left(e_h^{(0)}, e_{N_h}^{(0)}\right) \tag{18}$$

where $f(\cdot)$ is the aggregation function. We adopt the Bi-Interaction aggregation function[42] that considers various interaction information between features to spread more information between similar entities, as follows:

$$\begin{aligned} f\left(e_n, e_{N_n}\right) = & \text{ LeakyReLU }\left(W_1\left(e_n + e_{N_n}\right)\right) \\ & + \text{LeakyReLU }\left(W_2\left(e_n \odot e_{N_n}\right)\right) \end{aligned} \tag{19}$$

where LeakyReLU $(\cdot)$ is a nonlinear activation function. $W_1$ and $W_2$ are weight matrices, and $\odot$ is the element-wise product between matrices.

Repeating the above three steps several times, when the execution reaches the $l$-th time, $e_h^{(l)}$ denotes the representation of the $l$-th-order information, as follows:

$$e_h^{(l)} = f\left(e_h^{(l-1)}, e_{N_h}^{(l-1)}\right) \tag{20}$$

Finally, all the obtained vectors $\left\{e_h^{(0)}, \ldots, e_h^{(l)}\right\}$ are concatenated to obtain the embedding $e_h$, which contains multi-order connectivity information as in Eq. (21).

$$e_h = e_h^{(0)} \| \cdots \| e_h^{(l)} \tag{21}$$

We denote the embedding of the user node as $e_u$, and the item embedding containing structure-aware domain knowledge as $e_i^{graphic}$.

*3.3.3 Feature fusion strategy.* The previous two modules model cross-domain knowledge from the perspective of structural connectivity. Further, we design the feature fusion layer to combine the semantic and structural information, aiming to fully explore the cross-domain correlation and enhance the expressiveness of learned representations. However, we acknowledge that the proposed method has a non-end-to-end limitation, as we can only perform the fusion operation after the training of DAFSE is completed. As shown in Fig. 4, $e_i^{graphic}$ denotes an item feature generated by the attentive neighbor aggregation layer, which contains high-order connectivity information; $e_i^{semantic}$ denotes an item feature generated by DASFE, which contains semantic information. We design three feature fusion strategies with different ideas, denoted as $FFL_{init}$, $FFL_{concat}$, and $FFL_{add}$.

- $FFL_{init}$: this strategy inputs the semantic feature $e_i^{semantic}$ into CD-KGAT for item vector initialization (i.e., $e_h^{(0)}$), which will participate in the subsequent information propagation and aggregation.
- $FFL_{concat}$: this strategy concatenates $e_i^{graphic}$ and $e_i^{semantic}$ into a long vector, and then employs a fully connected layer to transform the long vector to the specified dimension. It is worth noting that we use L2 regularization after concatenation, which aims to unify the magnitude of the combined vector for efficient model learning.



Just Accepted

- $FFL_{add}$: this strategy uses two fully connected layers to linearly transform $e_i^{graphic}$ and $e_i^{semantic}$ into the same dimension. The two transformed features are then added after L2 normalization to obtain the final item embedding.

We experimentally compare the three strategies, and the results are presented in Section 4.5.2.

*3.3.4 Optimization.* We denote the complete item features generated by the feature fusion layer as $e_i$. Next, CD-KGAT predicts $u$'s preference for item $i$ by measuring the match between $e_u$ and $e_i$:

$$\hat{y}(u,i) = {e_u}^\top * e_i \tag{22}$$

In the sample dataset $H$, if the user $u$'s interactions contain item $i$, we denote $(u,i)$ as a positive sample; by randomly selecting $j$ that is not in the interactions, we denote $(u,j)$ as a negative sample. We define the loss function $L_{CF}$ based on Bayesian Personalization Ranking loss[30]:

$$L_{CF} = \sum_{(u,i,j)\in H} -\ln \sigma(\hat{y}(u,i) - \hat{y}(u,j)) \tag{23}$$

where $\sigma(\cdot)$ is the sigmoid activation function. $L_{CF}$ enables positive samples representing user preferences to receive higher prediction scores than negative samples.

CD-KGAT performs overall optimization by jointly learning in correlating cross-domain information and capturing collaborative information:

$$\mathbb{E}(\Theta) = L_{KG} + L_{CF} + \lambda\|\Theta\|_2^2 \tag{24}$$

where $L_{KG}$ and $L_{CF}$ are determined by Eq. (15) and Eq. (23), respectively; $\Theta$ is all parameters involved in CD-KGAT. $\lambda$ is the regularization weight to avoid overfitting.

In summary, we combine the encoded high-order connectivity with the extracted semantic information to realize our knowledge-correlated cross-domain recommendation method.

## 4 EXPERIMENTS

In this section, we conduct experiments to answer the following research questions:

- **RQ1** From the perspective of baseline comparison and internal mechanism, how does our method perform in the cold start scenario?
- **RQ2** How effectual is domain adaptation on knowledge transfer?
- **RQ3** Does the consideration of loss functions address the class imbalance issue?
- **RQ4** How do different designed modules contribute to CD-KGAT performance?
- **RQ5** How does CD-KGAT perform with different feature fusion strategies?

### 4.1 Datasets

We conduct the following experiments on two datasets from Facebook and Amazon, respectively. The Facebook dataset contains user interactions and item-related information on the platform and is provided by The Information Retrieval Group[5] in the work[7]. The Amazon dataset[6] contains user ratings and product metadata (description, category information, etc.), and we take ratings greater than 3 as positive interactions. DBpedia is a well-known knowledge graph constructed based on Wikipedia, as an external knowledge base to provide auxiliary information. For constructing the knowledge graph, we use the item metadata from DBpedia in our dataset, which is obtained through syntactic matching and SPARQL query techniques. The title in the item description is used to retrieve the KG entity ID from the DBpedia search API. To ensure the correctness of the retrieved information, we specify

[5] http://ir.ii.uam.es/
[6] https://nijianmo.github.io/amazon/index.html

Just Accepted

Table 1. Statistics of the recommendation with KG datasets.

| *Domain* | *Items* | *Users* | *Interactions* | *Categories* | *Avg. items/cates* | *Sparsity* | *Entities* | *Triples* | *Overlap users* | *CD-connectivity* |
|---|---|---|---|---|---|---|---|---|---|---|
| Facebook-Movie | 5,383 | 57,008 | 1,495,145 | 2,203 | 6.81 | 99.51% | 26,924 | 96,122 | 5187 | 19.21% |
| Facebook-Book | 4,411 | 7,084 | 108,554 | 1,034 | 3.19 | 99.65% | 16,130 | 38,376 | | |
| Amazon-Movie | 4,664 | 5,684 | 68,313 | 6,270 | 12.11 | 99.63% | 56,862 | 176,281 | 3257 | 15.61% |
| Amazon-Book | 5,886 | 3,257 | 684,878 | 8,921 | 3.45 | 96.43% | 29,780 | 77,608 | | |

the item type (e.g., dbo:Book and dbo:Film [7]) to avoid the problem of similarly named items from different domains, and then based on the item title to syntactic match with the label of its DBpedia entity. Since the data of the knowledge graph is organized in triples $(h, r, t)$ as shown in Definition 3.1, we use each linked entity as a head entity $h$ to get item metadata. It is worth noting that since our goal is item recommendation, we follow the previous work [7] by specifying relation $r$ as those relevant to relate the common preferences of different users, such as genres and directors of movies, and more detailed information can be found in [7]. In addition, we take a two-step filtering operation to ensure the quality of the datasets: 1). $N$-core filtering: filter out users and items with less than $N$ interactions to alleviate data sparsity[34]. We set $N$=10 for the Facebook dataset and $N$=5 for the Amazon dataset; 2). Items-aligning: exclude items with no matched entities in DBpedia for simplicity[17, 38, 39]. Table 1 shows the statistics of datasets, where "*Avg. items/cates*" denotes the average number of categories contained in each item. We also make a quantitative analysis of the cross-domain connectivity in the datasets, where "*CD-connectivity*" calculates the proportion of triples of common entities in the total triples. The Amazon-KG dataset we have processed is available at GitHub[8] for further comparisons.

## 4.2 Experimental setup

*4.2.1 Evaluation methodology.* In the offline experiments, we simulate a user cold-start scenario in the target domain based on the work[7], using a modified user-based five-fold cross-validation strategy with the following main steps:

Step 1. The users in the target domain are divided into five equal-sized user groups. In each cross-validation phase, four groups are training users, while the fifth group is the test users.

Step 2. For each test user $u$ in the fifth group, we randomly select $profileSize$ interactions (we set $profileSize$= 2) and add them to the training set, with the remaining interactions as test data. Thus, user $u$ becomes a new user with very few interactions.

Step 3. Extend the training set with all data from the remaining four groups of users and all source domain data.

Finally, the training set contains three types of data: all data in the source domain, all data of the training users, and $profileSize$ data of the test users in the target domain. The test set contains only the remaining interaction data of the test users.

Considering that a satisfactory recommendation can discover users' potential preferences beyond their expressed interests, we evaluate the method's performance in terms of both accuracy and diversity. For accuracy, we adopt three widely-used evaluation metrics: Recall@$K$, Precision@$K$, and MRR@$K$. By default, we set $K = 20$. Specifically, Precision is the mean probability that an item retrieved among *top-K* recommendations is relevant to the user, and Recall is the mean probability that relevant items are successfully retrieved among *top-K* recommendations. MRR indicates the order in which the first item relevant to the user appears on average, and is

[7]Namespace for dbo, http://dbpedia.org/ontology.
[8]https://github.com/WangYuhan-0520/Amazon-KG-dataset

Just Accepted

calculated as follows:

$$MRR = \frac{1}{|U|} \sum_{i=1}^{|U|} \frac{1}{rank_i} \tag{25}$$

where $rank_i$ is the highest rank of relevant results of $i$-th user in the *top-K* recommendation list, and if the list does not contain the relevant items then $rank_i \to \infty$. For diversity, we use the metric *BinomDiv*[37], which measures the diversity of items in the recommendation list based on the category information. *BinomDiv* is defined as the product of two components:

$$BinomDiv(R) = Coverage(R) \cdot NonRed(R) \tag{26}$$

where $R = \{i_1, i_2, ..., i_K\}$ denotes an output recommendation list, $Coverage(R)$ measures the comprehensiveness (coverage of the item categories) of $R$ during predicting the user's interests, and $NonRed(R)$ denotes the non-redundancy, which penalizes the score of $R$ that over-represent a certain category. We average the scores of the recommendation list for all test users as the final *BinomDiv* result. For the above four metrics (Recall@$K$, Precision@$K$, MRR@$K$, and *BinomDiv*), higher values mean better recommendation performance. We report the average results for the test set samples obtained through the five-fold cross-validation.

*4.2.2 Methods in comparison and parameter setting.* We compare the proposed method with the following representative baselines in single-domain and cross-domain recommendation methods:

- **Pop**: Item popularity-based recommendation method, where popularity represents the number of interactions. We randomly select N recommendations for a user among the top 100 popular target items.
- **UserCF**: User-based Collaborative Filtering method for single-domain recommendation. The user similarity is calculated by *Jaccard* similarity.
- **ItemCF**: Item-based Collaborative Filtering method for single-domain recommendation. The item similarity is calculated by *Jaccard* similarity.
- **BPR**[30]: Bayesian Personalized Ranking is a pairwise ranking method based on matrix factorization, which optimizes based on the relative preferences of users.
- **IMF**[16]: Matrix factorization recommendation method for single-domain recommendation with implicit (positive-only) feedback.
- **DTCDR**[49]: This work is the first to propose the dual-target CDR (methods that enable bidirectional transfer across domains with a dual-learning mechanism), which designs an embedding-sharing strategy integrating multiple sources of content information (e.g., reviews and tags).
- **CoNet**[13]: Collaborative Cross Networks is a cross-domain method that enables dual knowledge transfer based on the designed cross-stitch networks.
- **EMCDR**[27]: EMCDR learns latent factors for each domain separately based on matrix factorization and then uses a multilayer perceptron to bridge the cross-domain latent factors.
- **CMF**[33]: CMF jointly factorizes the rating matrices of two domains with the same users, assuming that all domains have a shared global user embedding matrix.
- **PTUPCDR**[51]: A meta-network is learned by fed with shared user preferences in the source and target domains to construct personalized bridge functions, aiming at personalized transfer of user preferences.
- **CD-MFs**[7]: Including three cross-domain recommendation methods based on matrix factorization model with different regularization designs, namely **SimMF**, **NieghborMF**, and **CentroidMF**. They regularize the model by computing the semantic similarity of cross-domain item metadata.

In addition, we also conduct experiments on UserCF, ItermCF, and IMF in the cross-domain scenario, and use CD- as the prefix to denote the cross-domain versions of these methods, namely **CD-UserCF**, **CD-ItermCF** and **CD-IMF**.

Just Accepted

Table 2. Overall Performance Comparison.

| **Method** | **Facebook Movie→Book** | | | | **Amazon Book→Movie** | | | |
|---|---|---|---|---|---|---|---|---|
| | *Accuracy* | | | *Diversity* | *Accuracy* | | | *Diversity* |
| | MRR | Recall | Precision | BinomDiv | MRR | Recall | Precision | BinomDiv |
| Pop | 10.36% | 5.00% | 3.53% | 4.98% | 3.42% | 1.91% | 1.10% | 3.77% |
| BPR | 8.45% | 2.93% | 2.06% | 4.36% | 3.56% | 2.13% | 1.00% | 3.63% |
| UserCF | 13.82% | 4.05% | 2.87% | 4.99% | 5.97% | 2.98% | 1.54% | 3.97% |
| ItemCF | 12.93% | 4.07% | 2.77% | 4.73% | 5.26% | 3.50% | 1.46% | 3.93% |
| IMF | 14.36% | 5.67% | 3.86% | 4.95% | 5.40% | 3.26% | 1.61% | 3.55% |
| CD-UserCF | 17.65% | 7.03% | 4.97% | 5.37% | 6.61% | 3.34% | 1.72% | 3.97% |
| CD-ItemCF | 10.63% | 6.23% | 4.43% | 6.43% | 4.47% | 3.25% | 1.33% | 3.84% |
| CD-IMF | 17.80% | **9.19%** | 6.40% | 5.74% | 6.43% | **3.84%** | **1.92%** | 3.24% |
| DTCDR | 8.77% | 4.30% | 3.10% | 5.04% | 4.88% | 2.90% | 1.54% | 4.19% |
| CoNet | 9.31% | 3.42% | 2.42% | 4.70% | 3.35% | 1.95% | 0.97% | 3.72% |
| EMCDR | 12.43% | 4.02% | 2.86% | 5.05% | 5.05% | 2.50% | 1.23% | 3.79% |
| CMF | 14.40% | 4.78% | 3.36% | 5.15% | 6.59% | 3.56% | 1.62% | 3.88% |
| PTUPCDR | 1.69% | 0.62% | 0.44% | 3.72% | 0.68% | 0.41% | 0.19% | 3.33% |
| SimMF | 16.52% | 9.13% | 6.38% | 5.85% | 6.70% | 3.78% | 1.83% | 3.53% |
| NeighborMF | 17.74% | 9.17% | **6.41%** | 5.84% | 6.63% | 3.76% | 1.82% | 3.54% |
| CentroidMF | 16.26% | 9.01% | 6.31% | 5.74% | 6.70% | 3.77% | 1.82% | 3.52% |
| **CD-KGAT** | **29.21%** | 8.83% | 6.34% | **6.84%** | **6.89%** | 3.28% | 1.46% | **4.20%** |

Just Accepted

The parameter settings of our method are as follows. For the parameters of DASFE, the dimension of Word2Vec embedding is 100, and the three convolution kernels of Text-CNN are 256 in number and 2, 3, and 4 in size. We set $\gamma = 2$, $\alpha = 0.25$ in the loss function (Eq. (6)), due to the best performance in the experiments of the original paper[24]. We use stochastic gradient descent optimization to train DASFE. For the parameters of CD-KGAT, the information propagation layer depth is searched in $\{1, 2, 3, 4\}$ (see Section 4.6 for detailed analysis), and the dimensions of each layer are 64, 32 and 16, respectively. We use the best-performing $FFL_{add}$ as the feature fusion strategy (see Section 4.5.2 for experimental comparison) and adopt the Adam optimization to train CD-KGAT. For the comparison fairness of all methods above, we restrict the maximum training epoch to 300, the batch size to 256, the embedding size to 64, the dropout ratio to 0.1, the regularization weight s searched in $\left\{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}\right\}$, and the learning rate is tuned in $\{0.01, 0.001, 0.0005, 0.0001\}$. The remaining parameters of baselines are fine-tuned based on the optimal values in the original paper.

## 4.3 Overall Performance and Case study (RQ1)

We evaluate the performance of our proposed model in terms of accuracy and diversity, and then intuitively analyze the mechanism of the method through a case study.

*4.3.1 Overall comparison.* Table 2 shows the comparison results in terms of accuracy and diversity. We have the key observations as follows.

*For our proposed CD-KGAT.* It shows promising performance in both accuracy and diversity metrics. In the MRR results. CD-KGAT has the best performance, outperforming the single-domain methods by 14.85~20.76% and the cross-domain methods by 11.41~27.52% on the Facebook dataset, and outperforming the single-domain methods by 0.92~3.47% and the cross-domain methods by 0.19~6.21% on the Amazon dataset. In the results of Recall and Precision, CD-KGAT outperforms most recommendation methods and can achieve comparable results to the best-performing ones. For the diversity metric *BinomDiv*, CD-KGAT improves by 0.41~3.12% compared with other recommendation methods on the Facebook dataset and is 0.01~0.96% on the Amazon dataset. Overall, these results show that, even in cold-start scenarios, our proposed method is capable of exploring not only the expressed but also the diverse potential ones for user interest. This confirms the effectiveness of our method in correlating cross-domain knowledge: it utilizes the cross-domain transferable semantic features extracted by DASFE, which ensures the recommendation accuracy; at the same time, it uses the graph information propagation in CCKG to obtain high-order connectivity across domains to encode users' multifaceted preferences, thus showing the advantages in diversity.

*Cross-domain methods vs. single-domain methods.* For the single-domain methods, Pop and BPR both perform poorly due to the fact that the former only provides non-personalized recommendations, while the latter struggles in coping with the cold-start. It is worth noting that the cross-domain version of UserCF shows improvements in both accuracy and diversity. However, ItemCF and IMF do not always improve after incorporating cross-domain information, and the reason may be that CD-ItemCF and CD-IMF fail to balance different domain information, making the auxiliary data become noise that weakens performance. It suggests that CDRS should focus on the efficient transfer and utilization of multi-source data. For methods that are inherently cross-domain recommendations, we can see that most have significantly better diversity than the single-domain methods. Among them, the accuracy performance of the dual-target methods DTCDR and CoNet is unsatisfactory, which may be attributed to their requirement for common users to have an equal amount of behaviors across domains, thus hardly handling the gap between sparse and dense data. The matrix factorization-based methods EMCDR and CMF focus more on the overlapping information between domains and achieve relatively better results. PTUPCDR achieves the worst performance, although the method expects to learn personalized preference transfer for users, it requires a training set of shared users with enough interactions in both domains to train the parameters of the meta-network, leading to the method's difficulty in cold-start scenarios. In contrast, the three methods of CD-MFs have competitive performance, which is attributed to the successful utilization of semantic information. However, their diversity is lower than our CD-KGAT due to the lack of consideration of the structural information.

Together these results provide important insights into the capabilities of CDRS. With the prerequisite of effective knowledge transfer, cross-domain methods increase the available data in the target domain, thereby alleviating the problem of cold-start and data sparsity. Moreover, since the single-domain methods may be vulnerable to similar and redundant item types, introducing varied data from other domains enhances the diversity of recommendation results. In summary, our experiments verify the superiority of cross-domain recommendation methods.

*4.3.2 Case study.* We present a case study to discuss our method in detail. We first randomly sample a user $u_{100030}$ from the Facebook dataset and scrutinize his historical interactions (i.e., the data in the training set) as shown in Table 3, and the Top-10 recommendations provided by CD-KGAT as shown in Table 4.

Table 3 provides that the user is interested in genres such as "Fantasy", "Adventure", and "Children", and also occasionally watch "Science" movies. Table 4 shows that CD-KGAT captures the user's primary interests, providing him with *Twilight*, *City of Ashes*, and *Chosen* in the "Fantasy" category, and *The Hunger Games* and *The Alchemist* in the "Adventure" category. A notable observation is that one of the user's interests, the genre keyword "Children", CD-KGAT recommends the target domain item *Speak*, whose genre keyword is "Young adult". This shows that CD-KGAT is able to align cross-domain features by analyzing the content information of items from a semantic perspective. Meanwhile, CD-KGAT also provides "Science" items *Pretties* and *Uglies* for users' secondary interests. Therefore, it verifies the accuracy of the method in learning user's domain-specific preferences.



Just Accepted

Table 3. $u_{100030}$'s history interactions.

| | Item Name | Genre |
|---|---|---|
| Movie | *Percy Jackson and The Lightning Thief* | Fantasy \| Adventure |
| Book | *Inkheart* | Fantasy \|Speculative \| Children |
| Movie | *Shrek* | Children \| Adventure \|Action |
| Movie | *The Final Destination* | Horror\|Paranormal |
| Movie | *Iron Man* | Science \|Action\|War |
| Movie | *Bride Wars* | Romance\| Comedy |
| Movie | *The Karate Kid* | Education\|Drama \|Romance |
| Book | *A Series of Unfortunate Events* | Children \| Comedy |
| Movie | *Law Abiding Citizen* | Crime\|Thriller |

Table 4. Recommendations for $u_{100030}$.

| | Item Name | Genre |
|---|---|---|
| Book | *Twilight* | Fantasy \|Speculative \|Horror\|Romance |
| Book | *The Hunger Games* | Adventure \| Science fiction \| Dystopian\|Post-apocalyptic |
| Book | *City of Ashes* | Fantasy \| Young adult |
| Book | *To Kill a Mockingbird* | Racism\|Ethnicity \|Discrimination |
| Book | *Warriors* | Science fiction \| Fantasy |
| Book | *Chosen (A House of Night novel)* | Fantasy \|Horror \|Speculative |
| Book | *Pretties* | Science fiction \|Dystopian \| Young adult |
| Book | *Speak* | Young adult \|Crime |
| Book | *Uglies* | Science fiction \|Dystopian \|Speculative |
| Book | *The Alchemist* | Adventure \| Fantasy |
| ⋮ | ⋮ | ⋮ |

Another notable observation is that the recommendation list also includes items that differ from the genres of the user's historical preference, such as the book *To Kill a Mockingbird.* This result may be explained by the fact that this book contains content plots about "Law" and "Judge", while *Law Abiding Citizen* in the historical interactions starts the storyline based on "Judicial procedure" related content. So there is a potential correlation between the two items. Since CD-KGAT captures high-order relations across domains from the connectivity perspective, it makes multi-category items of users' interests more reachable, thus enhancing diversity performance.

Without loss of generality, we randomly select another user $u_{100032}$. Same as above, we check his history and recommendations provided by CD-KGAT as shown in Table 5 and 6, respectively. We can see that CD-KGAT also captures the user's primary interests, such as "Comedy", "Science" and "Fantasy". However, there are also some items in the recommendation list that are untraceable as they contain genres that are beyond intuition. We analyze this probably because the movie domain has some domain-specific information, as the red boxed ones in the table: "Computer-animated", "Film soundtracks" and "Musicals". Similarly, the book domain has specific information like the description of writing techniques "Narrative". We recognize that CD-KGAT focuses on capturing cross-domain consistency without considering single-domain peculiarities, resulting in domain-specific features also involved in the learning process of content semanticity and structural connectivity, which becomes noise for training the model to a certain extent. In future work, we will discriminatively model the domain-shared and domain-specific information (using techniques such as disentanglement) to improve transfer learning across domains.

## 4.4 Model Analysis of DASFE

Next, we evaluate the ability of DASFE in knowledge transfer and the effectiveness of model design through qualitative and quantitative experiments.

*4.4.1 Effect on knowledge transfer (RQ2).* A good representation for cross-domain transfer is **domain-invariant**, which means that the algorithm cannot learn to identify its domain of origin[2, 3]. We adopt t-SNE[36] to visualize the distribution of the extracted semantic features from different domains on the Facebook dataset, aiming to

Just Accepted

Table 5. $u_{100048}$'s history interactions.

| | Item Name | Genre |
|---|---|---|
| Movie | *Finding Nemo* | Computer-animated \|Children \|Film soundtracks \|Adventure |
| Movie | *The Stepford Wives* | Comedy \| Science \|Technology |
| Movie | *Office Space* | Comedy \| Film soundtracks |
| Movie | *Miss Congeniality* | Comedy \| Film soundtracks |
| Movie | *The Prestige* | Science \|Thriller\|Revenge |
| Movie | *What Women Want* | Fantasy \|Romance \|Advertising |
| Movie | *The Phantom of the Opera* | Romance\| Musicals \|Drama |
| Movie | *Happy Feet* | Children\| Computer-animated \|Environmental |
| Movie | *Avatar* | Science \|Technology |
| Book | *I Capture the Castle* | Comedy \|Romance |
| Book | *The Historian* | Historical\|Horror \|Speculative\|Gothic |
| Movie | *Sleepy Hollow* | Horror\|Crime\|Ghost |
| Movie | *The Princess Bride* | Fantasy \|Adventure\| Comedy |

Table 6. Recommendations for $u_{100048}$.

| | Item Name | Genre |
|---|---|---|
| Book | *The Hitchhiker's Guide to the Galaxy* | Comedy \|Speculative \| Science fiction |
| Book | *A Series of Unfortunate Events* | Comedy \|Children |
| Book | *Brave New World* | Science fiction \|Dystopian \|Speculative |
| Book | *Wuthering Heights* | Fantasy \|Victorian\|Speculative |
| Book | *Watership Down* | Fantasy \|Children\|Speculative |
| Book | *Memoirs of a Geisha* | Historical\|Sexuality |
| Book | *Jane Eyre* | Victorian\|Gothic\|Orphans |
| Book | *The Count of Monte Cristo* | Maritime\|Adventure\|Revenge |
| Book | *The Shack* | Christian |
| Book | *Paradise Lost* | Christian\| Narrative |
| Book | *Twilight* | Fantasy \|Speculative \|Horror\|Romance |
| Book | *Walk Two Moons* | Children |
| ⋮ | ⋮ | ⋮ |

evaluate the effect of our method on semantic knowledge transfer. The proposed DASFE uses domain adaptation to align item features in a unified feature space. For comparison, we set another model named Source-Only, which uses only source domain data for training. In addition, *perplexity* is a hyperparameter of t-SNE and is considered as a smooth measure of the number of effective neighbors[36]. We investigate the visualization of *perplexity* at different values[9]. As shown in Fig. 6, we color-code samples by domains, where the blue circles indicate the source domain and the red triangles correspond to the target domain. It can be seen that the features learned by Source-Only are scattered and disordered, with no obvious manifold distribution, and the features of the two domains hardly overlap. The result can be explained by the heterogeneity in the content information of different domains, leading to differences in extracted features. In contrast, the features of the two domains extracted by our DASFE have a high degree of overlap and exhibit similar manifold distributions even under different perplexities. Since the domain adaptation effect largely represents the classification accuracy of the model for the target domain[8], this effect is strongly related to the feature overlap in visualization. Therefore, the experiments demonstrate that Source-Only is not suitable for the knowledge transfer task, while DASFE can utilize domain adaptation to obtain **domain-invariant** semantic features, which is beneficial for cross-domain knowledge correlation.

In addition, the extracted semantic features should be **discriminative**, which implies the ability to perform the classification task at the semantic level. We select several representative items in the target and source domains and visualize their extracted features using t-SNE, as shown in Fig. 7. It can be seen that the book *Harry Potter* is the closest to the movie *Harry Potter*, which is consistent with their content facts. Secondly, the distance between the movie *Toy Story* and the book *The Little Prince* is closer than other items, which is consistent with the fact that they both belong to the genres of "Children" and "Fantasy". While the movie *The Avengers* is far from the other items, probably because it belongs to the distinctive genres of "Science fiction" and "Adventure".

Further, we perform quantitative analysis by calculating the cosine similarity between the extracted item features, as shown in Table 7, where (M) in the name suffix means movie and (B) means book. As can be seen in Table 7, whether in the same domain or different domains, DASFE can learn closer feature representations for

[9]The learning rate of t-SNE is fixed to 10, and we use the default settings provided by Scikit-learn for the other parameters.

Just Accepted

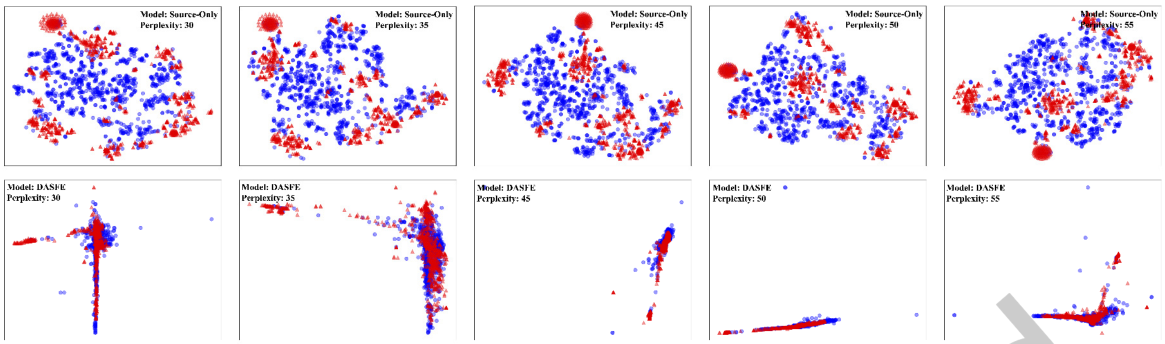



Fig. 6. The effect of adaptation shown by t-SNE visualizations of source and target domains with different *perplexities*. The images in the first row are the results of Source-Only model and the second row is the results of DASFE model. The blue circles indicate the source domain samples and the red triangles correspond to the target domain samples.

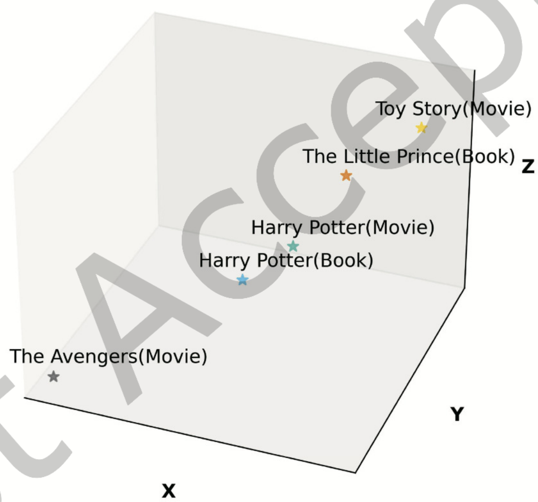



Fig. 7. A case study on t-SNE visualization of item semantic features.

Just Accepted

items with semantically similar content information, which is also consistent with the results of the previous visual analysis. Therefore, it verifies that our model can effectively extract the content semanticity of items, and the learned features are **discriminative**, which helps fully mine the relationships between cross-domain items.

*4.4.2 Improved design for class imbalance (RQ3).* Since our method utilizes the category attributes of items as training labels in semantic feature extraction, we conduct statistics on the distribution of the labels. Table 8 shows the distribution of some category labels on the Facebook dataset. We can see that the occurrence frequency of labels varies greatly, for example, the ratio of "Comedy" to "Toys" in the movie domain even reaches 108.5:1. So it is necessary to handle the class imbalance to ensure effective learning.

Table 7. Cosine similarity matrix of the item features.

| | *HarryPotter*(B) | *HarryPotter*(M) | *TheLittlePrince*(B) | *ToyStory*(M) | *TheAvengers*(M) |
|---|---|---|---|---|---|
| *HarryPotter*(B) | 1.0000 | - | - | - | - |
| *HarryPotter*(M) | 0.8508 | 1.0000 | - | - | - |
| *TheLittlePrince*(B) | 0.5864 | 0.5699 | 1.0000 | - | - |
| *ToyStory*(M) | 0.3614 | 0.4037 | 0.5553 | 1.0000 | - |
| *TheAvengers*(M) | 0.3611 | 0.3929 | 0.1485 | 0.1722 | 1.0000 |

Table 8. Distribution of partial categories within the Facebook-movie and Facebook-book domains.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Movie** | Category | *Comedy* | *Romance* | *Crime* | *Children* | $\cdots$ | *War* | *Toys* |
| | Samples Num. | 1302 | 556 | 457 | 269 | $\cdots$ | 96 | 12 |
| **Book** | Category | *Speculative* | *Science* | *Political* | *Business* | $\cdots$ | *Psychology* | *Animals* |
| | Samples Num. | 991 | 213 | 129 | 66 | $\cdots$ | 34 | 7 |

Table 9. Comparison of DASFE using different loss functions.

| | Recall | Precision | F2 |
|---|---|---|---|
| DASFE-FL | **76.64%** | **98.97%** | **80.26%** |
| DASFE-BCE | 17.68% | 74.86% | 20.87% |

We deal with the class imbalance issue from the point of loss function optimization: for the label predictor of DASFE, we employ Focal Loss (FL)[24] as the loss function of the output layer (Eq. (6)), denoted as DASFE-FL. We replace the loss function with the common Binary Cross Entropy loss (BCE) in the comparison model and denote it as DASFE-BCE. We use three metrics: Precision, Recall, and F-Score, to evaluate the classification performance of the label predictor. F-Score is a tradeoff of Recall and Precision as follows:

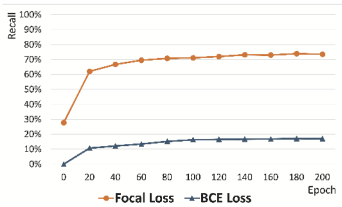
$$F_\beta = \frac{\left(\beta^2 + 1.0\right) \cdot \text{ Precision } \cdot \text{ Recall}}{\beta^2 \cdot (\text{ Precision } + \text{ Recall })} (0 < \beta < +\infty) \tag{27}$$

where $\beta = Recall/Precision$ is used to adjust the relative weights of the two metrics. Since the goal of DASFE is to predict all category labels contained in an item, Recall plays a more critical role in the evaluation. Therefore, we use $\beta = 2$ and denote this metric as F2.

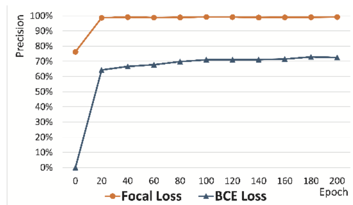
Table 9 shows the evaluation results of the two models on the test set after training reaches convergence. We can see that DASFE-BCE achieves extremely unsatisfactory results, while DASFE-FL significantly outperforms it in all three metrics. We investigate the model training process in detail to further analyze the impact of class imbalance. Fig. 8 shows the training results obtained after 200 epochs. From Fig. 8(a) and 8(b), we can see that although DASFE-BCE can achieve a relatively high Precision, its Recall is consistently at a low value. The reason for this is that in the scenario of class imbalance, BCE biases the label predictor towards popular labels while ignoring other less frequent labels. However, this bias is not suitable for the optimization goal of DASFE, which leads to the failure of inductive learning on samples. In contrast, DASFE-FL improves performance with increasing training times. It shows that under the guidance of FL, DASFE is able to learn the semantic features by continuously fitting the samples, thus realizing stable progress in the classification task. Fig. 8(c) also shows that

Just Accepted

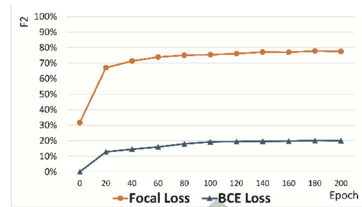
DASFE-FL is superior to DASFE-BCE significantly in the comprehensive metric F2. Therefore, this comparative experiment verifies the effectiveness of our improved design for class imbalance.




(a) Recall




(b) Precision




(c) F2

Fig. 8. Analysis of the training process of DASFE using different loss functions.

## 4.5 Model Analysis of CD-KGAT

We perform ablation and contrast experiments on the model to gain deep insight into the design of CD-KGAT.

*4.5.1 Ablation study (RQ4).* We conduct an ablation study to evaluate the impact of knowledge graph embedding, attention mechanism, and semantic feature fusion. For CD-KGAT, we disable the fusion of semantic features in the representation learning of CD-KGAT (see Section 3.3.3), denoted as "w/o Semantic"; we disable the adaptation of cross-domain item features in semantic knowledge extraction (see Section 4.4.1), denoted as "w/o DA"; we disable the knowledge graph embedding module (Eq. (15)), denoted as "w/o KGE"; we disable the attention mechanism (Eq. (17)), denoted as "w/o Attentive". The experimental results are shown in Fig. 9 and 10.

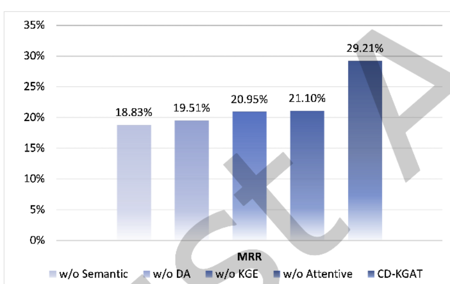



Fig. 9. Ablation results on Facebook dataset.

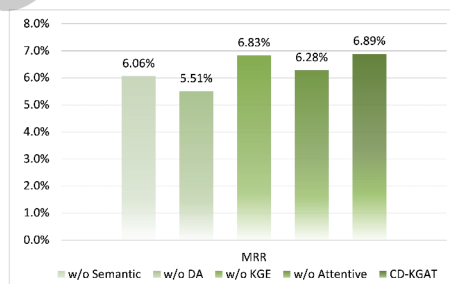



Fig. 10. Ablation results on Amazon dataset.

From the results, we find that the removal of four modules degrades the performance. On the Facebook dataset, removing the attention mechanism degrades the performance by 8.11%, which indicates that not differentiating the contribution of neighbor information may create noise to affect feature learning. Removing the knowledge embedding module leads to an 8.26% reduction in MRR, which demonstrates its importance for predicting user preferences in digging for cross-domain connectivity. Compared with the previous two modules, it can be seen that semantic feature fusion has a more significant impact on accuracy with 10.38%. The performance of "w/o DA" is also unsatisfactory due to the lack of cross-domain semantic alignment, which confirms that transferable semantic information across domains helps to build complete item profiles and enhances representation learning. On the Amazon dataset, the ablation of the four modules shows a similar trend. Therefore, this experiment verifies the effectiveness of our model design.

Just Accepted

*4.5.2 Comparison of feature fusion strategies (RQ5).* In order to make full use of the KG information, we design three feature fusion strategies, namely $FFL_{init}$, $FFL_{concat}$, and $FFL_{add}$. We compare the performance of CD-KGAT with different strategies as shown in Fig. 11 and 12. What stands out is that $FFL_{add}$ is significantly improved compared to $FFL_{init}$ by 10.02% and $FFL_{concat}$ by 10.47% on the Facebook dataset, and 2.96% and 0.60% on the Amazon dataset. It shows that $FFL_{add}$ can combine the advantages of semantic and structural information more appropriately. Then we experiment on $FFL_{add}$ without L2 normalization, denoting it as $FFL_{add}$ $w/o$ $reg$. We can see that not using normalization weakens the performance. The reason may be that the magnitudes differ between different types of features, directly adding them may lead to distortion of useful features. This experiment verifies that our method fits user-item interactions well, thus accurately discovering user preferences.

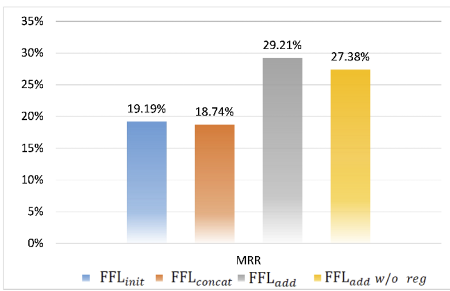



Fig. 11. Comparison of fusion strategies on Facebook dataset.

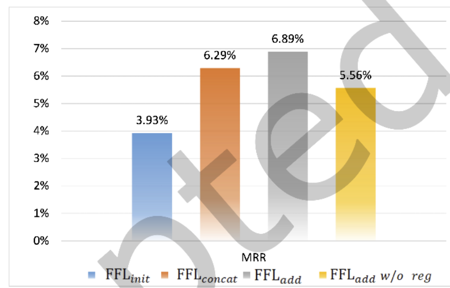



Fig. 12. Comparison of fusion strategies on Amazon dataset.

*4.5.3 Comparison on KG properties.* Since our proposed CD-KGAT is KG-dependent, we reprocess the KG data of the Facebook dataset from the properties of sparsity and reliability to gain insight into the usefulness of KG for the proposed method. First, we randomly remove triples with a percentage of 40% for the original KG $G_{kg}$, denoted as "sparse-KG". Second, we generate the fake KG by reconstructing the triples in $G_{kg}$. Specifically, we generate a fake triple by randomly replacing the tail entity with the original triples: $(h, r, t') = fake((h, r, t))$, where $t' \neq t$. We perform a one-to-one $fake$ operation on all triples in $G_{kg}$ to generate "fakeKG" with the same density; then we randomly remove triples with a proportion of 40% for "fakeKG", denoted as "sparse-fakeKG"; further, we perform random connection between nodes to enrich the triples with a proportion of 40% for "fakeKG", denoted as "dense-fakeKG". The information statistics of the above KG variants are shown in Table 10[10], where we utilize two density metrics for KG information proposed by Jay Pujara et al.[29], defined as the average triples per relation (relational density, *RD*) and the average triples per entity (entity density, *ED*):

$$RD = \frac{\|T\|}{\|R\|}, \quad ED = \frac{2\|T\|}{\|E\|} \tag{28}$$

where $\|T\|$ is the number of all triples in KG.

We conduct experiments on the proposed CD-KGAT utilizing variants of KG as shown in Fig. 13. In terms of KG sparsity, we can see that the effect of CD-KGAT on "sparse-KG" decreases compared to the original "KG", which is attributed to the fact that the removal of information makes "sparse-KG" less helpful for structural connectivity of the method across domains. In terms of KG reliability, "sparse-KG" still outperforms all three fakeKGs. It is interesting to note that MRR decreases as the density of fakeKG increases, implying that the reliability of KG is proportional to the model performance. The reason may be that the false connections in the graph increase the noise data, which makes CD-KGAT merge useless information during neighbor aggregation. More dense fakeKG

[10]Here we merge the original KG triples from the two domains for density calculations, noting that the "Triples" statistic is not exactly equal to the summation of the value in the "Triples" column of Table 1 due to the overlap of the triples from the two domains.



Just Accepted

results in more noise, thus "dense-fakeKG" generates the weakest effect. In addition, since CD-KGAT has certain guarantees from the perspective of content semanticity[11], it still outperforms other baselines even on fakeKGs, which confirms the advantage of learning from our dual perspectives in robustness.

Table 10. Statistics of KG variants.

| KG variants | Triples | *RD* | *ED* |
|---|---|---|---|
| dense-fakeKG | 176,899 | 5,706 | 9 |
| fakeKG | 126,357 | 4,076 | 6 |
| sparse-fakeKG | 75,815 | 2,446 | 4 |
| sparse-KG | 75,815 | 2,446 | 4 |
| KG | 126,357 | 4,076 | 6 |

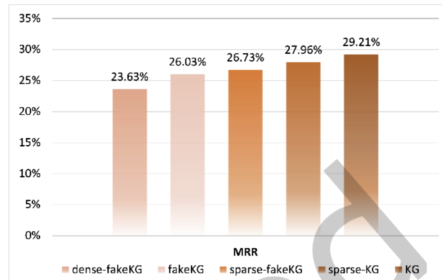



Fig. 13. Comparison on KG properties.

### 4.6 Parameter sensitivity

CD-KGAT uses the information propagation mechanism to capture high-order neighbor information. We vary $l$ in Eq. (20) - (21) while keeping other parameters fixed, to further analyze the effect of propagation depth on the cross-domain recommendation. The experimental results with different propagation layers are shown in Fig. 14. The left side of the figure shows the convolution kernel size settings for different layers. We set the reduced size to avoid possible noise impacts of high-order distant nodes[48]. It is apparent from this figure that the performance improves with the propagation depth and reaches an optimum at $l = 3$, while the performance degrades when $l = 4$. We have the following understanding: CD-KGAT can accurately model high-order relationships across domains, and the information propagation mechanism can improve the model performance. However, possible noise in over-high-order information may affect the representation learning.

## 5 CONCLUSION AND FUTURE WORK

In this paper, we innovatively propose a knowledge-correlated cross-domain recommendation method, which uses KG as common knowledge to link different domains, and explores the contribution of KG from dual perspectives of content semanticity and structural connectivity. We first mine cross-domain transferable semantic information based on domain adaptation, then capture cross-domain high-order structural information based on the graph neural network. In particular, we design the feature fusion strategy to combine the two types of information, thereby enhancing the representation learning ability of the model. We conduct experiments under the cold-start scenario and use a series of qualitative and quantitative analyses to verify the rationality and effectiveness of the method. The results demonstrate that our method has the advantages of accuracy and diversity in preference prediction.

This research focuses on the analysis of the relationship between cross-domain knowledge and user preferences, thus enriching the theory of cross-domain recommendation research and providing implications for coping with the cold-start problem in reality. Furthermore, our method can extend in future contributions by exploring more application scenarios (e.g., the dual-target CDR). And we will consider more available information such as user reviews, social relations, and user-related information.

Just Accepted

[11]We have not made changes to the semantic module DASFE, even though the semantic information of KG is used in this paper. The reason is that common recommendation datasets contain item-related attributes that can also provide semanticity.

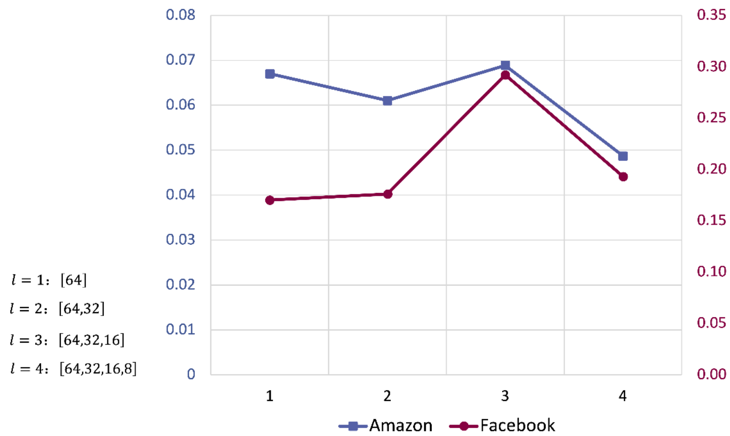



Fig. 14. Effect of information propagation depth.

Just Accepted

## ACKNOWLEDGMENTS

This research is partially supported by National Natural Science Foundation of China (Grant No.62276196), Natural Science Foundation of Chongqing (Grant No. cstc2021jcyj-msxmX1013), Key Research Program of Hubei (Grant No. 2021BAA030).

## REFERENCES

[1] Mustafa Azak and Aysenur Birturk. 2010. Crossing Framework - A Dynamic Infrastructure to Develop Knowledge-based Recommenders in Cross Domains. In *Proceedings of the 6th International Conference on Web Information Systems and Technologies*, Vol. 2. INSTICC Press, 125–130.

[2] Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Wortman Vaughan. 2010. A theory of learning from different domains. *Mach. Learn.* 79, 1-2 (2010), 151–175.

[3] Shai Ben-David, John Blitzer, Koby Crammer, and Fernando Pereira. 2006. Analysis of Representations for Domain Adaptation. In *NIPS*. MIT Press, 137–144.

[4] Shlomo Berkovsky, Tsvi Kuflik, and Francesco Ricci. 2008. Mediation of user models for enhanced personalization in recommender systems. *User Model. User Adapt. Interact.* 18, 3 (2008), 245–286.

[5] Iván Cantador, Ignacio Fernández-Tobías, Shlomo Berkovsky, and Paolo Cremonesi. 2015. Cross-Domain Recommender Systems. In *Recommender Systems Handbook*. Springer, 919–959.

[6] Ignacio Fernández-Tobías, Iván Cantador, Marius Kaminskas, and Francesco Ricci. 2011. A generic semantic-based framework for cross-domain recommendation. In *Proceedings of the 2nd International Workshop on Information Heterogeneity and Fusion in Recommender Systems*. ACM, 25–32.

[7] Ignacio Fernández-Tobías, Iván Cantador, Paolo Tomeo, Vito Walter Anelli, and Tommaso Di Noia. 2019. Addressing the user cold start with cross-domain collaborative filtering: exploiting item metadata in matrix factorization. *User Model. User Adapt. Interact.* 29, 2 (2019), 443–486.

[8] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor S. Lempitsky. 2016. Domain-Adversarial Training of Neural Networks. *J. Mach. Learn. Res.* 17 (2016), 59:1–59:35.

[9] Chen Gao, Xiangning Chen, Fuli Feng, Kai Zhao, Xiangnan He, Yong Li, and Depeng Jin. 2019. Cross-domain Recommendation Without Sharing User-relevant Data. In *The World Wide Web Conference*. ACM, 491–502.

[10] Qingyu Guo, Fuzhen Zhuang, Chuan Qin, Hengshu Zhu, Xing Xie, Hui Xiong, and Qing He. 2022. A Survey on Knowledge Graph-Based Recommender Systems. *IEEE Transactions on Knowledge and Data Engineering* 34, 8 (2022), 3549–3568.

[11] Jia He, Rui Liu, Fuzhen Zhuang, Fen Lin, Cheng Niu, and Qing He. 2018. A General Cross-Domain Recommendation Framework via Bayesian Neural Network. In *2018 IEEE International Conference on Data Mining*. IEEE, 1001–1006.

[12] Geoffrey E. Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2012. Improving neural networks by preventing co-adaptation of feature detectors. arXiv:1207.0580

[13] Guangneng Hu, Yu Zhang, and Qiang Yang. 2018. CoNet: Collaborative Cross Networks for Cross-Domain Recommendation. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management, CIKM 2018, Torino, Italy, October 22-26, 2018.* ACM, 667–676.

[14] Guangneng Hu, Yu Zhang, and Qiang Yang. 2019. Transfer Meets Hybrid: A Synthetic Approach for Cross-Domain Collaborative Filtering with Text. In *The World Wide Web Conference*. ACM, 2822–2829.

[15] Liang Hu, Jian Cao, Guandong Xu, Longbing Cao, Zhiping Gu, and Can Zhu. 2013. Personalized recommendation via cross-domain triadic factorization. In *22nd International World Wide Web Conference*. ACM, 595–606.

[16] Yifan Hu, Yehuda Koren, and Chris Volinsky. 2008. Collaborative Filtering for Implicit Feedback Datasets. In *Proceedings of the 8th IEEE International Conference on Data Mining*. IEEE, 263–272.

[17] Jin Huang, Wayne Xin Zhao, Hongjian Dou, Ji-Rong Wen, and Edward Y. Chang. 2018. Improving Sequential Recommendation with Knowledge-Enhanced Memory Networks. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. ACM, 505–514.

[18] Sang-Young Jo, Sun-Hye Jang, Hee-Eun Cho, and Jin-Woo Jeong. 2019. Scenery-Based Fashion Recommendation with Cross-Domain Geneartive Adverserial Networks. In *International Conference on Big Data and Smart Computing*. IEEE, 1–4.

[19] Marius Kaminskas, Ignacio Fernández-Tobías, Iván Cantador, and Francesco Ricci. 2013. Ontology-Based Identification of Music for Places. In *Information and Communication Technologies in Tourism 2013*. Springer, 436–447.

[20] Yoon Kim. 2014. Convolutional Neural Networks for Sentence Classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*. ACL, 1746–1751.

[21] Yehuda Koren. 2008. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 426–434.

[22] Anil Kumar, Nitesh Kumar, Muzammil Hussain, Santanu Chaudhury, and Sumeet Agarwal. 2014. Semantic clustering-based cross-domain recommendation. In *2014 IEEE Symposium on Computational Intelligence and Data Mining*. IEEE, 137–141.

[23] Zhetao Li, Pengpeng Qiao, Yuanxing Zhang, and Kaigui Bian. 2019. Adversarial Learning of Transitive Semantic Features for Cross-Domain Recommendation. In *2019 IEEE Global Communications Conference*. IEEE, 1–6.

[24] Tsung-Yi Lin, Priya Goyal, Ross B. Girshick, Kaiming He, and Piotr Dollár. 2017. Focal Loss for Dense Object Detection. In *IEEE International Conference on Computer Vision*. IEEE, 2999–3007.

[25] Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015. Learning Entity and Relation Embeddings for Knowledge Graph Completion. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence*. AAAI Press, 2181–2187.

[26] Antonis Loizou. 2009. *How to recommend music to film buffs: enabling the provision of recommendations from multiple domains.* Ph. D. Dissertation. University of Southampton.

[27] Tong Man, Huawei Shen, Xiaolong Jin, and Xueqi Cheng. 2017. Cross-Domain Recommendation: An Embedding and Mapping Approach. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*. ijcai.org, 2464–2470.

[28] Tomás Mikolov, Ilya Sutskever, Kai Chen, Gregory S. Corrado, and Jeffrey Dean. 2013. Distributed Representations of Words and Phrases and their Compositionality. In *Advances in Neural Information Processing Systems*. 3111–3119.

[29] Jay Pujara, Eriq Augustine, and Lise Getoor. 2017. Sparsity and Noise: Where Knowledge Graph Embeddings Fall Short. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. ACL, 1751–1756.

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

[31] Giuseppe Sansonetti, Fabio Gasparetti, and Alessandro Micarelli. 2019. Cross-Domain Recommendation for Enhancing Cultural Heritage Experience. In *Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization*. ACM, 413–415.

[32] Yue Shi, Martha A. Larson, and Alan Hanjalic. 2011. Tags as Bridges between Domains: Improving Recommendation with Tag-Induced Cross-Domain Collaborative Filtering. In *User Modeling, Adaption and Personalization*, Vol. 6787. Springer, 305–316.

[33] Ajit Paul Singh and Geoffrey J. Gordon. 2008. Relational learning via collective matrix factorization. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 650–658.

[34] Zhu Sun, Di Yu, Hui Fang, Jie Yang, Xinghua Qu, Jie Zhang, and Cong Geng. 2020. Are We Evaluating Rigorously? Benchmarking Recommendation for Reproducible Evaluation and Fair Comparison. In *Proceedings of the 14th ACM Conference on Recommender Systems*. ACM, 23–32.

[35] Grigorios Tsoumakas, Ioannis Katakis, and Ioannis P. Vlahavas. 2010. Mining Multi-label Data. In *Data Mining and Knowledge Discovery Handbook*. Springer, 667–685.

[36] Laurens Van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of machine learning research* 9, 11 (2008), 2579–2605.

[37] Saúl Vargas, Linas Baltrunas, Alexandros Karatzoglou, and Pablo Castells. 2014. Coverage, redundancy and size-awareness in genre diversity for recommender systems. In *Proceedings of the 8th ACM Conference on Recommender Systems*. ACM, 209–216.


Just Accepted

[38] Hongwei Wang, Fuzheng Zhang, Jialin Wang, Miao Zhao, Wenjie Li, Xing Xie, and Minyi Guo. 2018. RippleNet: Propagating User Preferences on the Knowledge Graph for Recommender Systems. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. ACM, 417–426.
[39] Hongwei Wang, Miao Zhao, Xing Xie, Wenjie Li, and Minyi Guo. 2019. Knowledge Graph Convolutional Networks for Recommender Systems. In *The World Wide Web Conference*. ACM, 3307–3313.
[40] Jun Wang, Shijun Li, Sha Yang, Hong Jin, and Wei Yu. 2017. A New Transfer Learning Model for Cross-Domain Recommendation. *Chinese Journal of Computers* 40, 10 (2017), 2367–2380.
[41] Quan Wang, Zhendong Mao, Bin Wang, and Li Guo. 2017. Knowledge Graph Embedding: A Survey of Approaches and Applications. *IEEE Trans. Knowl. Data Eng.* 29, 12 (2017), 2724–2743.
[42] Xiang Wang, Xiangnan He, Yixin Cao, Meng Liu, and Tat-Seng Chua. 2019. KGAT: Knowledge Graph Attention Network for Recommendation. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 950–958.
[43] Yuhan Wang, Qing Xie, Lin Li, and Yongjian Liu. 2021. An Empirical Study on Effect of Semantic Measures in Cross-Domain Recommender System in User Cold-Start Scenario. In *Knowledge Science, Engineering and Management*, Vol. 12816. Springer, 264–278.
[44] Yang Xu, Zhaohui Peng, Yupeng Hu, and Xiaoguang Hong. 2018. SARFM: A Sentiment-Aware Review Feature Mapping Approach for Cross-Domain Recommendation. In *Web Information Systems Engineering*, Vol. 11234. Springer, 3–18.
[45] Deqing Yang, Yanghua Xiao, Yangqiu Song, Junjun Zhang, Kezun Zhang, and Wei Wang. 2014. Tag propagation based recommendation across diverse social media. In *23rd International World Wide Web Conference*. ACM, 407–408.
[46] Feng Yuan, Lina Yao, and Boualem Benatallah. 2019. DARec: Deep Domain Adaptation for Cross-Domain Recommendation via Transferring Rating Patterns. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*. ijcai.org, 4227–4233.
[47] Qian Zhang, Jie Lu, Dianshuang Wu, and Guangquan Zhang. 2019. A Cross-Domain Recommender System With Kernel-Induced Knowledge Transfer for Overlapping Entities. *IEEE Trans. Neural Networks Learn. Syst.* 30, 7 (2019), 1998–2012.
[48] Cheng Zhao, Chenliang Li, and Cong Fu. 2019. Cross-Domain Recommendation via Preference Propagation GraphNet. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. ACM, 2165–2168.
[49] Feng Zhu, Chaochao Chen, Yan Wang, Guanfeng Liu, and Xiaolin Zheng. 2019. DTCDR: A Framework for Dual-Target Cross-Domain Recommendation. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management (CIKM '19)*. ACM, 1533–1542.
[50] Feng Zhu, Yan Wang, Chaochao Chen, Guanfeng Liu, Mehmet A. Orgun, and Jia Wu. 2018. A Deep Framework for Cross-Domain and Cross-System Recommendations. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*. ijcai.org, 3711–3717.
[51] Yongchun Zhu, Zhenwei Tang, Yudan Liu, Fuzhen Zhuang, Ruobing Xie, Xu Zhang, Leyu Lin, and Qing He. 2022. Personalized Transfer of User Preferences for Cross-domain Recommendation. In *The Fifteenth ACM International Conference on Web Search and Data Mining*. ACM, 1507–1515.

Just Accepted