Knowledge-Aware Cross-Semantic Alignment for Domain-Level Zero-Shot Recommendation

Junji Jiang
anjou_j@tju.edu.cn
College of Management and Economics, Tianjin University
Tianjin, China

Hongke Zhao*
hongke@tju.edu.cn
College of Management and Economics, Tianjin University
Tianjin, China

Ming He
heming01@foxmail.com
AI Lab at Lenovo Research
Beijing, China

Likang Wu
wulk@mail.ustc.edu.cn
School of Computer Science and Technology, University of Science and Technology of China & State Key Laboratory of Cognitive Intelligence
Hefei, China

Kai Zhang
sa517494@mail.ustc.edu.cn
Anhui Province Key Laboratory of Big Data Analysis and Application, University of Science and Technology of China & State Key Laboratory of Cognitive Intelligence
Hefei, China

Jianping Fan
jfan1@lenovo.com
AI Lab at Lenovo Research
Beijing, China

ABSTRACT

Recommendation systems have attracted attention from academia and industry due to their wide range of application scenarios. However, cold start remains a challenging problem limited by sparse user interactions. Some scholars propose to transfer the dense information from the source domain to the target domain through cross-domain recommendation, but most of the work assumes that there is a small amount of historical interaction in the target domain. However, this approach essentially presupposes the existence of at least some historical interaction within the target domain. In this paper, we focus on the domain-level zero-shot recommendation (DZSR) problem. To address the above challenges, we propose a knowledge-aware cross-semantic alignment (K-CSA) framework to learn transferable source domain semantic information. The motivation is to establish stable alignments of interests in different domains through class semantic descriptions (CSDs). Specifically, due to the lack of effective information in the target domain, we learn semantic representations of source and target domain items based on knowledge graphs. Moreover, we conduct multi-view K-means to extract item CSDs from the learned semantic representations. Further, K-CSA learns universal user CSDs through the designed multi-head self-attention. To facilitate the transference of user interest from the source domain to the target domain, we devise a cross-semantic contrastive learning strategy, grounded in the prototype distribution matrix. We conduct extensive experiments on several real-world cross-domain datasets, and the experimental results clearly demonstrate the superiority of our proposed K-CSA compared with other baselines.

*Corresponding Author

CIKM '23, October 21–25, 2023, Birmingham, United Kingdom

ACM ISBN 979-8-4007-0124-5/23/10…$15.00
https://doi.org/10.1145/3583780.3614945

CCS CONCEPTS

• **Information systems** → **Information retrieval**; • **Computing methodologies** → *Knowledge representation and reasoning*.

KEYWORDS

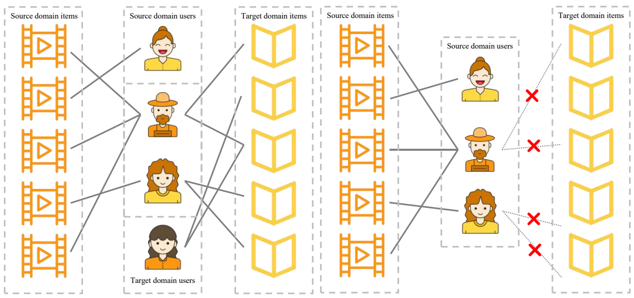
Zero-shot Recommendation, Knowledge Graph, Cross-domain Recommendation, Semantic Representation

ACM Reference Format:
Junji Jiang, Hongke Zhao, Ming He, Likang Wu, Kai Zhang, and Jianping Fan. 2023. Knowledge-Aware Cross-Semantic Alignment for Domain-Level Zero-Shot Recommendation. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (CIKM '23), October 21–25, 2023, Birmingham, United Kingdom.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3583780.3614945

1 INTRODUCTION

Recommendation systems aim to recommend items by learning users' personalized interests and providing basic network services to match user needs quickly. In the past few decades, many methods have been proposed on how to utilize users' historical behaviors to achieve better recommendations, such as collaborative filtering [14, 55], feature interaction [12, 40], sequential recommendation [16, 19] and cross-domain recommendation [3, 48]. However, improving the performance of cold-start recommendation systems remains a significant and challenging problem. Especially when a new domain appears, the interactions ae missing, and this cold start scenario is called domain-level zero-shot recommendation (DZSR) [54].

DZSR poses practical challenges in the realm of recommendation systems. When a new domain is introduced, the traditional single-domain recommendation methods prove ineffective due to the absence of historical user-item interactions specific to the target domain. Moreover, even cross-domain recommendation methods are constrained. As shown in Figure 1, general cross-recommendation assumes that there are overlapping users in the source domain and



(a) Cross-domain recommendation. (b) DZSR.

Figure 1: The difference between cross-domain recommendation and DZSR. Solid lines denote available interactions while dashed lines denote unavailable interactions.

the target domain and establishes relations based on this, while the target domain users and interactions in DSZR are completely unavailable. The main challenges can be summarised as 1) The first challenge originates from the discrepancies in the distribution of items across the source and target domains. The accumulated interaction records in the source domain cannot directly represent user interests in the target domain. This requires the development of techniques that can bridge the gap and proficiently capture user preferences; 2) The second challenge emerges from the reality that identical items may provoke diverse degrees of user interest across different domains. The shift in the contextual backdrop from the source to the target domains can induce considerable fluctuations in user preferences.

To solve the problem of sparse interaction in the target domain and provide high-quality recommendations, some scholars conduct knowledge transfer from the perspective of user behavior modeling in the source domain and try to establish connections between items in the source domain and the target domain. Previous research [7, 45] has proposed the usage of text as an intermediary. Nevertheless, textual information has been demonstrated to introduce bias, as the same text can hold different connotations depending on the context. For instance, the word "Python" in the title of a book might pertain to a tutorial on Python programming language in a programming context, whereas it could denote a book on animals in a natural science context. To address this issue, Tiger [54] introduced the knowledge graph to model richer knowledge information to eliminate bias. Therefore, it is intuitively plausible that user preferences can still propagate even if the original item sets from different domains do not directly overlap. KG contains a large number of structured and semantic triple-connected entities, which can serve as a bridge to project common knowledge. However, although Tiger jointly models the item semantics of source and target domains in the knowledge extraction stage, it fails to learn cross-domain representation information during the high-order feature interaction phase of user and item representations. This guides the model to learn only the semantic information of the source domain.

One effective way to address the aforementioned problems is to introduce the idea of class semantic descriptions (CSDs) from classic zero-shot learning problems [25, 36, 43, 47]. CSDs represent a kind of high-order semantic information, describing the core features that help to identify categories. For example, to distinguish animals, we can first define some CSDs like "swim", "wing" and "fur". Then, at the training stage, we can train classifiers to recognize these CSDs. For domain-level zero-shot recommendation, our goal is to obtain high-quality CSDs as auxiliary data for supervising knowledge from existing items in the source domain that are transferred to unseen classes in the target domain. In this paper, we generate semantic CSDs for item and user respectively, possessing these properties: 1) item-CSDs, we use the KG embedding method based on the item content, ensuring that the item's representation has the basic properties we proposed, that is, it has a good intra-domain and inter-domain generalization performance. 2) user-CSDs, we further integrate item CSDs into user CSDs learning, so that user information on a single domain can be effectively transferred.

In this paper, we propose a knowledge-aware cross-semantic alignment (K-CSA) framework to learn transferable source domain semantic information. Specifically, we first employ knowledge graphs to jointly learn semantic representations of items in source and target domains. To obtain generalized high-level information, prototypes are extracted from items through the multi-view K-means algorithm. The prototype extracts the mutual information of the source and target domain items, which avoids the tendency of the model to learn local information only by using source-domain item representations. Furthermore, we leverage the multi-head self-attention mechanism to learn high-order user CSDs with different view prototype information. To transfer user interest from the source domain to the target domain, we design cross-semantic contrastive learning based on the prototype distribution matrix. This effectively aligns item and user semantics across domains. In the model optimization stage, we introduce negative samples from the target domain for optimization. To verify the effectiveness of K-CSA, we conduct extensive experiments on four public datasets. The main contributions of this paper are summarized as follows:

- We are the first to introduce the class semantic descriptions (CSDs) modeling to domain-level zero-shot recommendation.
- To transfer user interest from the source domain to the target domain, we align the items through the knowledge graph. Then K-CSA constructs the item CSDs based on multi-view K-Means and extracts user CSDs based on the self-attention mechanism.
- We introduce cross-semantic contrastive learning through the prototype distribution matrix, to enhance the alignments of item and user CSDs across domains.
- Extensive experiments are conducted to verify the effectiveness of our proposed K-CSA and the applicability of produced representations for multiple recommendation scenarios.

The rest of this paper is organized as follows. Section 2 briefly introduces related work. Subsequently, Section 3 introduces the problem definition of DZSR. The details of our proposed K-CSA are presented in Section 4. Section 5 provides extensive experiment results and case studies to show the effectiveness of K-CSA. Finally, we summarize the paper and future plans in the final section.

2 RELATED WORK

Our paper is mainly related to three areas of research: 1) cold-start recommendation, 2) cross-domain recommendation, and 3) knowledge-aware recommendation.

2.1 Cold-start Recommendation

Many cold-start solutions have been developed for users or items with limited interactions, aiming to improve recommendation decisions. Various machine learning techniques have been employed such as meta-learning [8, 21, 27, 29] and leveraging additional data [4, 26, 30]. Meta-learning enables local updates of model parameters with a small number of samples, allowing for rapid adaptation to user preferences and estimation based on limited items. Recent studies have observed similarities between the intuitions of Zero-Shot Learning (ZSL) and Cold-Start Problem (CSP), as both involve two spaces: one for basic features and another for auxiliary descriptions. The goal is to predict unseen basic features for certain samples using auxiliary descriptions. Consequently, several zero-shot models for CSP have been proposed, such as LLAE [22] and MAIL [10]. LLAE employs a low-rank linear auto-encoder to map between user behavior and user attributes (i.e., user characteristics), while MAIL ensures the hidden features of user attributes and user behavior are highly aligned by performing cross-modal reconstruction between two autoencoders. However, existing cold-start work does not fully consider domain differences in the context of DSZR.

2.2 Cross-domain Recommendation

To address the widespread information sparsity in recommendation systems, transfer learning [1, 49] has been developed to fuse information for cross-domain recommendation (CDR). There have been various approaches to facilitate knowledge transfer for CDR. Some studies have explored cross-domain relevance at the attribute level, focusing on linking user and item features, such as user reviews[35], item tags [11], and knowledge graphs [15]. Others have delved into techniques like sharing embeddings [6, 53] or mapping embeddings [18, 28] of overlapping users/items. For instance, DTCDR [53] integrates multi-domain knowledge by sharing common user embeddings in the combination layer, while HeroGRAPH [6] enhances entity representations by merging in-domain and heterogeneous graph embeddings linked to multiple domains. Moreover, some methods, such as XPTRANS [17], have ventured into collaborative training by creating interconnections between models. However, traditional CDR approaches often still require sparse data within the target domain. Recently, an increasing number of attention has been paid to using large language models (LLMs) to solve the cold-start problem [44].

2.3 Knowledge-aware Recommendation

Knowledge graphs (KGs) are rich structured information repositories that can significantly enhance recommender systems. DKN [38] employs knowledge-aware convolutional neural networks (KCNN) for improved news representations using KGs, while KRED [24] refines article representations, aiding in news recommendations. Leveraging KGs' natural entity connections, RippleNet [37] propagates user preferences across KGs to address user-item interaction sparsity using a memory network. KGCN [39] and KGAT [41] harness knowledge graph convolution networks for item representations, and Studie [5] adopts a multi-task framework for both recommendation and KG embedding. In KG reasoning, KPRN [42] focuses on finding high-quality KG paths between nodes, employing a path encoder for selection. In contrast, PGPR [46] and ADAC [50] apply reinforcement learning for path-finding tasks in KGs. Lastly, in [23], a subgraph generator extracts significant KG subgraphs to infer relations between items.

3 PROBLEM FORMULATION

In this section, we first give the definition of the DZSR problem. There are source domain and target domain, denoted as $\mathcal{S}$ and $\mathcal{T}$ respectively. The objective is to transfer the user behavior information from the domain $\mathcal{S}$ to $\mathcal{T}$. $\mathcal{D}$ represents one of the domains in the paper. User set and item set can be denoted as $\mathcal{U}_{\mathcal{D}}$ and $\mathcal{V}_{\mathcal{D}}$. Different from traditional cross-domain recommendation, DZSR assumes that there are no feasible interactions in domain $\mathcal{T}$. So the relation between source domain users and target domain users is denoted as $\mathcal{U}_{\mathcal{T}} \in \mathcal{U}_{\mathcal{S}}$. For item, the item sets are totally different for the two domains, i.e., $\mathcal{V}_{\mathcal{T}} \cap \mathcal{V}_{\mathcal{S}} = \emptyset$. All possible interaction set between users and items is $\mathcal{I}_{\mathcal{D}} = \mathcal{U}_{\mathcal{D}} \times \mathcal{V}_{\mathcal{D}}$. For $\mathcal{I}_{\mathcal{T}}$, the interaction is unavailable for the training phase.

Task Description. With the above definitions, we define the task of DZSR as follows: **Input:** the source domain user-item interaction set $\mathcal{I}_{\mathcal{S}}$, the user and item set in the two domain $\{\mathcal{U}_{\mathcal{S}}, \mathcal{U}_{\mathcal{T}}, \mathcal{I}_{\mathcal{S}}, \mathcal{I}_{\mathcal{T}}\}$. **Output:** a trained model $\mathcal{F}(\cdot)$ that forecasts the interactions in target domain by $\hat{\mathcal{I}}_{\mathcal{T}} = \mathcal{F}(\mathcal{U}_{\mathcal{T}})$.

4 METHODOLOGY

In this section, we present the details of our K-CSA framework. Our method contains the following components: 1) Knowledge-Aware Representation Learning, 2) Cross-domain CSDs Construction, and 3) Cross-semantic Contrastive Learning. First, the Knowledge-Aware Representation Learning mechanism employs the item's title to align with the entities of the knowledge graph and subsequently extract the knowledge subgraph. Bert is introduced to obtain the semantic representation for each entity. After the information propagation of the subgraph, K-CSA reads out the representation of each item. Secondly, the Cross-domain CSDs Construction mechanism is developed to generate universal item and user CSDs. Specifically, we first obtain prototypes through multi-view K-means clustering and reconstruct item CSDs based on prototype representations. Additionally, user CSDs are constructed based on the designed multi-head self-attention mechanism. Finally, the Cross-semantic Contrastive Learning mechanism enhances the learned universal CSDs by comparing item CSDs through the prototype distribution matrix. The overall architecture of our method is shown in Figure 2.

4.1 Knowledge-Aware Representation Learning

Since items from $\mathcal{S}$ and $\mathcal{T}$ are separated due to the lack of interactions in $\mathcal{T}$. One of the main challenges for DZSR is to learn universal item representations. For instance, when watching the movie *La La Land*, the user notices that the main character IS reading the book *The Hero's Journey*. Then the user selects the *Hero's Journey* in the book domain. Consequently, the embedding for *La La Land* and *The Hero's Journey* should be similar in the cross-domain representation space. So we propose to use the knowledge graph to enhance the semantic representation. Specifically, we leverage ConceptNet [33], a semantic network that represents words and phrases as nodes connected by relationships, aiding natural language understanding. We utilize ConceptNet to extract universal

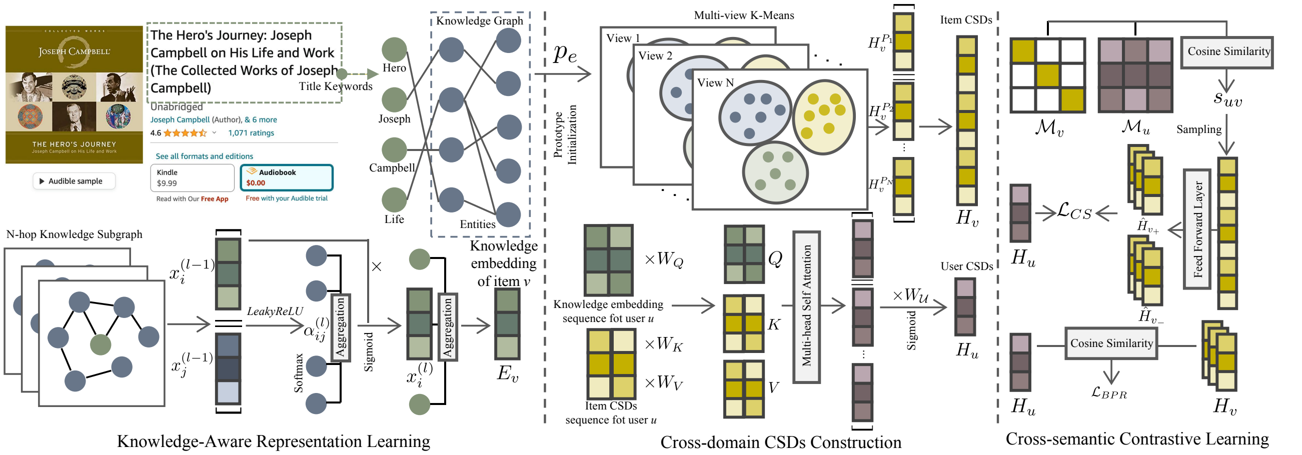


Figure 2: The overall framework of our proposed K-CSA model involves Knowledge-Aware Representation Learning through semantic learning based on knowledge graph and Cross-domain CSDs Construction.

semantic entities. The knowledge graph contains over 8 million entities and 21 million edges. The definition of knowledge graph is denoted as: $\mathcal{G} = \{(h, r, t)|h, t \in \mathcal{E}, r \in \mathcal{R}\}$, where (h, r, t) represents that a relation r from head h to tail t, and $\mathcal{E}$ and $\mathcal{R}$ denote the entity set and relation set respectively.

To associate items within the recommendation dataset with entities present in the knowledge graph, we adopt the methods proposed by KB4Rec [51] to obtain the relevant entities for each item. We extract the product title from the item's metadata and employ it as input for the Google Knowledge Graph Search API, retrieving the top-ranked entity. In this study, we utilize the initial 64 characters of the title to formulate the query keywords. Furthermore, through the related entity, the edges from items to entities are constructed, and we extract the n-hop subgraph from ConceptNet based on all the connected entities. A relation-based attention mechanism calculates the entity-level embeddings. And the initial entity embeddings are calculated by SBert [31]. The equations are as below:

$$\begin{aligned} x_i^{(l)} &= \sigma(\xi^{(l)} x_i^{(l-1)} + \textstyle\sum_{j \in N_i} \alpha_{ij}^{l} x_j^{(l-1)}), \\ \alpha_{ij}^{(l)} &= \frac{LeakyReLU(\mathrm{r}_{ij}^{\top} W_k^l [x_i^{(l-1)} || x_j^{(l-1)}])}{\sum_{j \in N_i} LeakyReLU(\mathrm{r}_{ij}^{\top} W_k^l [x_i^{(l-1)} || x_j^{(l-1)}])}, \end{aligned} \tag{1}$$

where x_i denotes the embedding for entity i, and $\mathcal{N}$ represents the neighbors for entity i. α_{ij} represents the attention score for the entity i and j. W_k is the learnable parameter. $LeakyReLU(\cdot)$ and $\sigma(\cdot)$ are the activation functions. ξ is an irrational number to enhance the expressiveness of the graph attention mechanism.

After the information propagation on knowledge CSDs, we aggregate the embedding to the item embedding.

$$E_v = \frac{1}{|\mathcal{N}_v|} \sum_{e \in \mathcal{N}_v} x_e, \tag{2}$$

where E_v is the aggregated item knowledge embedding and $\mathcal{N}_v$ represents the neighbor entities for item v.

4.2 Cross-domain CSDs Construction

The generation of item CSDs, predicated on the knowledge graph, enables the mapping of items from the source and target domains into a shared semantic space. The ensuing objective is the cross-domain learning of user CSDs. One idea is to aggregate the item CSDs of the source domain based on the item-user interest graph. However, due to disparities in the entity distribution of item titles between the source and target domains, directly learning graph representations on the source domain remains insufficient. Therefore, we propose to build prototypes based on the common entities of items on the source and target domains and build user CSDs based on the representation of the prototypes.

Specifically, we employ multi-view K-means to learn multiple prototypes. During the selection of the initial center point for the K-means algorithm, we choose to sample from the entities of the knowledge subgraph. To prevent local entities in the source domain from interfering with the generation of generic prototypes, we convert the proportion of entity-connected source and target domain items into sampling probabilities, and the formula is as follows,

$$p_e = \min((1 - |r_e^{\mathcal{S}} - r_e^{\mathcal{T}}|) \cdot p_{init}, p_\tau), \tag{3}$$

where p_e is the sampling probability for entity e. $r_e^{\mathcal{S}}$ and $r_e^{\mathcal{T}}$ represent the ratio of the source domain and target domains in the neighbors of entity e. p_{init} is a hyperparameter that controls the overall probability and $p_\tau > 1$ denotes the cut-off probability.

According to the sampling probability, we sample k entities for N views and use them as the embedding of the initial center point. After the K-means algorithm of each view converges, we connect the prototypes of N view groups as item CSDs.

$$H_v = W_{\mathcal{V}}[H_v^{P_1} | H_v^{P_2} | \cdots | H_v^{P_N}], \tag{4}$$

where $H_v^{P_i}$ is the corresponding prototype embedding for item v in the view i. $[\cdot|\cdot]$ means concat operation. This ensures that item CSDs can obtain common information across domains while avoiding being trapped in the representation of local information.

Given the reconstructed item CSDs, K-CSA further learns universal user CSDs. Based on the historical interactions, we design a self-attention structure that considers both fine-grained item representation and universal item CSDs. First, we use the item representation to compute the query matrix, and the item CSDs to compute the key and value matrices,

$$Q^u = W^Q \mathbf{E}^u,\ K^u = W^K \mathbf{H}^u,\ V^u = W^V \mathbf{H}^u, \tag{5}$$

where $\mathbf{E}^u$ and $\mathbf{H}^u$ represent the history knowledge embedding and item CSDs matrices for user u. Next, the result obtained by a single attention head can be obtained by the following formula,

$$head_i = \mathrm{softmax}(\frac{Q_i^u K_i^u}{\sqrt{d}})V_i^u. \tag{6}$$

Finally, we concat the multi-head attention vectors and calculate the user CSDs as follows,

$$H_u = LayerNorm(\sigma([head_1|head_2|\cdots|head_M]W_{\mathcal{U}})), \tag{7}$$

where H_u denotes the user CSDs and $W_{\mathcal{U}}$ is the hyperparameter matric. σ represents the activation function.

4.3 Cross-semantic Contrastive Learning

Our motivation is to learn universal item and user CSDs across source and target domains. However, directly adopting BPR loss for optimization still makes the model fall into the local optimum of the source domain. Therefore, we propose to address this issue through contrastive learning across semantics. The core problem is that only the item CSDs of the source domain are applied in the aggregation of user CSDs. Therefore, our goal is to make user CSDs similar to item CSDs of the same cluster category on the target domain according to the results of the multi-view prototype. For each user and item, we construct the prototype distribution matrix as follows,

$$\mathcal{M}_u = \begin{bmatrix} [m_u^{1,1}, \cdots, m_u^{1,K}] \\ \vdots \\ [m_u^{N,1}, \cdots, m_u^{N,K}] \end{bmatrix},\ \mathcal{M}_v = \begin{bmatrix} [m_v^{1,1}, \cdots, m_v^{1,K}] \\ \vdots \\ [m_v^{N,1}, \cdots, m_v^{N,K}] \end{bmatrix}, \tag{8}$$

where $\mathcal{M}_u$ and $\mathcal{M}_v$ represent the distribution matrix of user u and item v. Among them, $m_v^{i,j} \in \{0,1\}$ marks whether item v belongs to j-th category of i-th view, which is set to 1 if yes, 0 otherwise. $m_u^{i,j} \in [0,1]$ is the prototype distribution of normalized user history interaction items calculated by,

$$m_u^{i,j} = \frac{1}{|\mathcal{I}_u|}\sum_{v\in\mathcal{I}_u} m_v^{i,j}, \tag{9}$$

where $\mathcal{I}_u$ is the historical interaction of user u. Hence the similarity between user and item can be calculated by cosine similarity,

$$s_{uv} = \frac{M_u \cdot M_v}{||M_u||||M_v||}. \tag{10}$$

We set a threshold for cosine similarity to sample positive and negative examples from the target domain for contrastive learning. A 2-layer feed-forward neural network is utilized to map item CSDs to the size user CSDs. The loss function for cross-semantic contrastive learning is shown as follows,

$$\begin{gathered} \hat{H}_v = W_2^F(\sigma(W_1^F H_v + b_1^F)) + b_2^F, \\ \mathcal{L}_{CS} = \textstyle\sum_{u\in\mathcal{U}_\mathcal{S}}\sum_{v^+,v^-\in\mathcal{V}_\mathcal{T}} -\log\frac{\exp(H_u\cdot\hat{H}_{v^+}/\tau)}{\sum_{i^-\in\mathcal{V}_\mathcal{T}}\exp(H_u\cdot\hat{H}_{v^-}/\tau)}. \end{gathered} \tag{11}$$

4.4 Model Optimization

For the prediction results, we directly map the user CSDs on the source domain and perform inner product with the item CSDs on the target domain to predict the possibility of interaction between user u and item i.

$$\hat{y}_{uv} = \frac{H_u\hat{H}_v}{|H_u||\hat{H}_v|}. \tag{12}$$

In this paper, we utilize the Bayesian Personalized Ranking (BPR) loss [32] to directly capture information from interactions. The BPR loss is a ranking objective function that is commonly used in recommendation systems. Its design ensures that the predicted score of observed interactions is higher than that of sampled unobserved ones. The BPR loss is formulated as the following objective function:

$$\mathcal{L}_{BPR} = \sum_{u\in\mathcal{U}^\mathcal{S},\ i,j\in\mathcal{V}^\mathcal{S}} -log\sigma(\hat{y}_{ui} - \hat{y}_{uj}). \tag{13}$$

Optimization Objective. We define our optimization objective with the integration of the aforementioned losses and weight decay regularization term in the following equation,

$$\mathcal{L} = \mathcal{L}_{BPR} + \lambda_1\mathcal{L}_{CS} + \lambda_2||\Theta||^2, \tag{14}$$

where the parameters are denoted as Θ. λ_1 and λ_2 are the learning weights. The learning process of K-CSA is shown in Algorithm 1.

Algorithm 1 The Training Algorithm of K-CSA

Require: Source domain user-item interaction set $\mathcal{I}_\mathcal{S}$; Source domain user set $\mathcal{U}_\mathcal{S}$; Source domain item set $\mathcal{V}_\mathcal{S}$; Target domain item set $\mathcal{V}_\mathcal{T}$
Ensure: Parameters of K-CSA Θ

Random initialize model parameters Θ
while not converged **do**
  for $v \in \mathcal{V}_\mathcal{S}$ **do**
    Extract knowledge subgraph of v
    Calculate item knowledge embedding E_v based on Eq.1 and Eq.2
  end for
  for n **in** 1 **to** N **do**
    Sample initial prototypes based on Eq.3 and conduct K-Means Algorithm
    for $v \in \mathcal{V}_\mathcal{D}$ **do**
      Get the n-th view $H_v^{P_n}$
    end for
  end for
  Calculate the item CSDs based on Eq.4
  Construct the item prototype distribution matrix
  Calculate the user CSDs based on Eq.5, Eq.6 and Eq.7
  Construct the user prototype distribution matrix
  Calculate the $\mathcal{L}_{CS}$ based on Eq.11
  Calculate the $\mathcal{L}_{BPR}$ based on Eq.13
  Calculate the overall loss based on Eq.14
end while
return Parameters Θ

Table 1: Statistics of the experimental datasets.

Dataset	AB	AM	ML	LFM
#User	11,240	11,240	6,040	18,029
#Item	47,377	16,100	3,655	311,994
#Interaction	202,223	142,395	997,580	1,006,639
#Entity	3,599,000			
#Relation	2,089			
#Triple	32,372,637			

5 EXPERIMENTS

In this section, we perform extensive experiments on four public real-world datasets for model performance evaluation by answering the following research questions:

- **RQ1:** Compared with various state-of-the-art models, how does K-CSA perform for domain-level zero-shot recommendation?
- **RQ2:** Is K-CSA stable given different source domain data?
- **RQ3:** What is the impact of major components in K-CSA?
- **RQ4:** How do the key hyperparameters of K-CSA impact its performance with different settings?
- **RQ5:** Are the item and user CSDs interpretable?

5.1 Experimental Setting

1) Dataset. We follow the dataset used in Tiger [54]. We conduct experiments on the following four real-world datasets including different overlap levels of users and items:

- Amazon Movies TV (AM) and Amazon Books (AB) are two subsets of the Amazon datasets, which contain product reviews and metadata from Amazon and nowadays have become popular benchmark datasets for recommender systems. We use "reviewerID" to bridge users across the two datasets.
- Movielens(ML) dataset contains anonymous movie ratings to describe users' preferences on movies, which is widely used in the evaluation of recommender systems.
- LastFM(LFM) dataset contains music listening information from the world's largest online music service Last.fm. The ML and LFM datasets are used to extend the source domain and verify if Tiger can benefit from the public datasets without overlapping entities in the target domain

For the AB and AM datasets, we retain only users with at least five historical interactions in both datasets. User interactions on the source domain are preserved for training, while interactions on the target domain are assumed to be invisible to comply with the DZSR specification. Regarding the ML and LFM datasets, there is no overlap between users and Amazon data. They are used as supplementary datasets to verify the performance of K-CSA under different data richness. The data statistics are shown in Table 1.

2) Baselines. Regarding baselines, we benchmark K-CSA against a range of models, encompassing both existing methods and their variants. We meticulously reproduce these baselines in accordance with their original publications and open-source codes, striving to guarantee fair comparisons within our experiment. The baseline models incorporated in our study include:

Random: Random baseline is introduced as a common baseline in zero-shot learning problems. Any worthwhile model is expected to perform better than random outcomes.

Content-based methods: Textual content serves as an alternative way for universally representing items.

- **SBert [31]:** is a modification of the widely-used BERT architecture, specifically designed for sentence-level representation and similarity tasks.
- **DeBERTa [13]:** A model that introduces disentangled attention and a decoding-enhanced objective to improve the performance of the original BERT architecture.

Knowledge graph-based methods: Given that conventional knowledge graph embedding models acquire entity embeddings through a fully self-supervised approach, relying on the knowledge graph's structure, they can be considered as natural benchmarks for evaluating zero-shot recommendation performance.

- **TransE [2]:** is a prominent embedding model for knowledge graph representation learning, which encodes both entities and relations as continuous vectors in a shared latent space.
- **KGCN [39]:** is a powerful model for recommender systems, which leverage the rich information provided by knowledge graphs to enhance recommendation quality.
- **KGAT [41]:** is a model that integrates knowledge graphs by leveraging graph attention mechanisms.

Cold start-based methods: One way to solve the DZSR problem is to degenerate the problem into a cold-start recommendation on a single domain, that is, put all the items of the target domain into the source domain.

- **MvDGAE [52]:** extract multifaceted meaningful semantics on HINs as multi-views for both users and items, effectively enhancing user/item relationships on different aspects.
- **MetaKG [9]:** effectively captures the high-order collaborative relations and semantic representations, which could be easily adapted to cold-start scenarios.

DSZR methods: Most cross-domain recommendation methods are restricted due to the need for interaction with the target domain. The first work of DZSR is naturally added as a baseline.

- **Tiger [54]:** is a cutting-edge model designed for DSZR, which focuses on learning transferable interest graph embeddings.

3) Evaluation Metrics. We evaluate all models using two popular metrics: Hit Ratio (H@K), Normalized Discounted Cumulative Gain (N@K) and Mean Reciprocal Ranking (MRR@K), where K is obtained from the classical setting $\{10, 100\}$ taking into account both precision and recall properties. Higher values for all measures mean better performance. During the testing phase, all models are asked to rank all items that each user has not interacted with. To reduce the effect of random noise, each experiment was independently repeated 10 times under the same conditions, and the average performance is reported here.

4) Parameter Settings. All the methods are implemented by Pytorch, and we tune the hyperparameters of both our model and baselines by the Adam [20]. The embedding size produced by SBert is 1024 and further mapped into 512 by the feed-forward layer for efficiency. To ensure a fair comparison, all the other baselines are set the same and share the same batch size 1024. The size of H_u and

Table 2: Performance comparison for different methods. The results are obtained from 10 individual runs for every setting. The best is highlighted in bold.

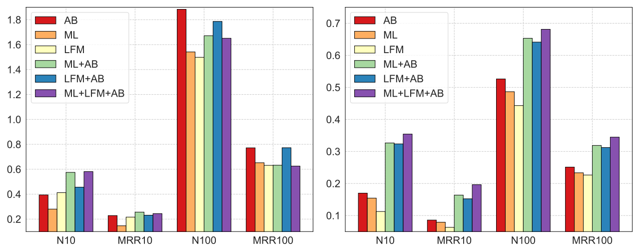
Model	Source	Target	H@10	N@10	MRR@10	H@100	N@100	MRR@100
Random	-	AM	0.061	0.0276	0.0148	0.6202	0.1294	0.0639
SBert	AB	AM	0.1299	0.0484	0.0272	1.2899	0.2583	0.1115
DeBERTa	AB	AM	0.2265	0.0866	0.0499	1.3154	0.2973	0.1236
TransE	AB	AM	0.3201	0.1571	0.0806	1.4852	0.3718	0.1653
KGCN	AB	AM	0.5366	0.2488	0.1278	3.6032	0.8149	0.3464
KGAT	AB	AM	0.5122	0.2235	0.1275	2.8814	0.8033	0.3495
MvDGA	AB	AM	0.7312	0.3211	0.1737	4.5213	1.0215	0.4225
MetaKG	AB	AM	0.7955	0.3446	0.1786	5.1387	1.1894	0.5171
Tiger	AB	AM	0.9311	0.3748	0.2194	7.3063	1.5401	0.7468
K-CSA	AB	AM	**0.9562**	**0.3941**	**0.2285**	**8.5633**	**1.8835**	**0.7718**
Random	-	AB	0.0208	0.0091	0.0054	0.2103	0.0436	0.0210
SBert	AM	AB	0.0503	0.0194	0.0100	0.4576	0.094	0.0408
DeBERTa	AM	AB	0.0545	0.0235	0.0123	0.4122	0.0933	0.0390
TransE	AM	AB	0.0619	0.0284	0.0166	0.3908	0.0909	0.0383
KGCN	AM	AB	0.0858	0.0483	0.0261	0.7111	0.165	0.0678
KGAT	AM	AB	0.0913	0.0512	0.0279	0.9934	0.1936	0.0905
MvDGA	AM	AB	0.1536	0.0814	0.0443	1.2352	0.2265	0.1031
MetaKG	AM	AB	0.2213	0.0922	0.0471	1.4993	0.3047	0.1447
Tiger	AM	AB	0.3052	0.1361	0.0700	1.9624	0.4511	0.2111
K-CSA	AM	AB	**0.3541**	**0.1699**	**0.0859**	**2.4714**	**0.5262**	**0.2512**

$\hat{H}_v$ are set to 512. The head of multi-head self-attention is 3. For the key hyperparameters, p_{init} and p_τ are set to 2 and 0.2 respectively. The layer L of the graph attention mechanism for semantic modeling is searched from $[1, 2, 3, 4, 5]$ and the number N of views for K-Means is searched from $[2, 4, 6, 8, 10]$. The hyperparameters λ_1 and λ_2 for loss function are searched from $[0.01, 0.1, 0.5, 1, 10]$.

5.2 Performance comparison (RQ1)

The performance of K-CSA compared with other baselines on two source and target domain settings is shown in Table 2. The results are obtained from 10 individual runs for every setting. From the evaluation results, we summarize the following observations:

- K-CSA constantly outperforms all baselines on the two different source-target domain settings. Through the cross-domain CSDs construction, K-CSA improves the generalization and robustness of DZSR by learning universal item and user CSDs. We attribute the significant performance gain of K-CSA to two key aspects: 1) K-CSA effectively alleviates the domain bias brought by direct learning representations on both source and target domains. 2) In addition to extracting the textual information of the item, the generalization performance of the model is enhanced to a certain extent by constructing the cross-domain item relationship through the graph.
- Knowledge graph-based methods perform better than content-based methods. This superiority could be attributed to the structured representation of concepts and their interrelations afforded by knowledge graphs, facilitating more intricate and comprehensive semantic modeling. In the context of DZSR, the source and target domains are intrinsically connected through the underlying semantics of items, rather than through explicit user-item interactions. While content-based approaches rely on item descriptions or user-generated content, which may vary drastically between domains, knowledge graphs establish robust semantic links that transcend these domain-specific characteristics. By capturing the overarching semantic structure, knowledge graph-based methods can create a more coherent and accurate mapping between source and target domain items.
- KGAT and KGCN demonstrate superior performance compared to TransE, illustrating the advantage of training entity embeddings using user behaviors from the source domain. KGAT and KGCN utilize user behaviors to inform the learning process, yielding embeddings that are much more relevant and tailored to specific tasks. Hence they successfully distill recommendation-oriented knowledge from the broader, structure-oriented knowledge graph, thereby enhancing the overall quality and relevancy of recommendations. Further, KGAT manifests additional advantages resulting from the graph attention mechanism.
- Compared with the best state-of-the-art method Tiger, the deployment of CSDs within K-CSA plays a crucial role in improving the method's performance, which fuses multi-view semantic information and alleviates the domain bias in DZSR. Meanwhile, cross-semantic contrastive learning further amplifies the efficacy of our proposed model. This learning paradigm seeks to understand and relate to different semantics across domains, facilitating a richer and more general representation of items. Furthermore, the fine-grained modeling of knowledge graphs employed by KCSA represents a significant advantage over the other baseline



(a) The target domain is AM. (b) The target domain is AB.

Figure 3: Performance comparison of K-CSA with different source domain datasets.

Table 3: Ablation studies for different mechanisms in K-CSA. The best is highlighted in bold.

Model	H10	N10	MRR10	H100	N100	MRR100
	AB (Source) - AM (Target)					
K-CSA-NK	0.8513	0.3655	0.2015	6.9832	1.3844	0.7369
K-CSA-NC	0.8212	0.3512	0.1903	6.1241	1.3062	0.7127
K-CSA-NS	0.9215	0.3784	0.2142	7.2447	1.4969	0.7413
K-CSA	**0.9562**	**0.3941**	**0.2285**	**8.5633**	**1.8835**	**0.7718**
Model	AM (Source) - AB (Target)					
K-CSA-NK	0.2818	0.1033	0.0693	1.9323	0.4142	0.1936
K-CSA-NC	0.2561	0.0982	0.0644	1.7182	0.3689	0.1774
K-CSA-NS	0.3156	0.1451	0.0712	2.2342	0.4692	0.2233
K-CSA	**0.3541**	**0.1699**	**0.0859**	**2.4714**	**0.5262**	**0.2512**

methods. By acknowledging and employing the vast and intricate connections within knowledge graphs, our method substantially improves the precision of semantic extraction.

5.3 Multi Source Domain Performance (RQ2)

We further test the performance of K-GCA on source domain data with different richness. We consider the cases of non-overlap users and partial-overlap users. This corresponds to the history available setting and history protected setting in Tiger [54], respectively. We show the results in Figure 3.

It is obvious that no matter whether the target domain is AB or AM, when the domain data of non-overlap users is fully used, the performance of the model is lower than that of partial-overlap. Intuitively, since the source domain does not overlap users with the target domain, there is a larger gap in the interest distribution of users. For example, in e-commerce platforms with different marketing strategies, users will prefer products with higher cost performance in promotional platforms. On the luxury e-commerce platform, users pay more attention to quality and brand. Although DZSR with non-overlap users is challenging, K-CSA still achieves promising results.

In the partial-overlap user scenario, when the target domain is AB, out-domain data can improve the performance of K-CSA on AB. This can be attributed that ML and LFM are video and audio respectively, which is more similar to partial-overlap AM. Therefore, consistent user CSDs can be obtained during training in the source domain, avoiding the noise introduced by different data distributions. Additionally, according to Figure 3(a), simply increasing the source domain data does not guarantee performance improvement.

5.4 Ablation study (RQ3)

In this section, we perform model ablation studies to evaluate the effects of different mechanisms of K-CSA in contributing to DZSR performance. In particular, 1) K-CSA-NK represents K-CSA without knowledge graph, we directly utilize the representations from SBert; 2) K-CSA-NC removes the item and user CSDs construction mechanism, and the user representation is learned through a common multi-head attention mechanism; 3) K-CSA-NS only utilizes the BPR loss for optimization. The performance for different variants is shown in Table 3.

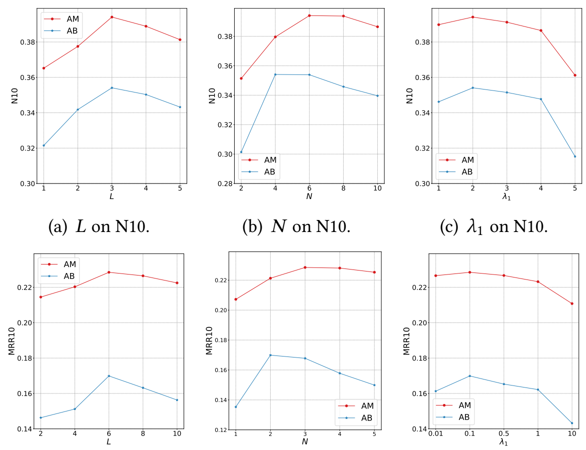
- For K-CSA-NK, performance drops in both settings. This is consistent with our conclusions in Section 5.2 that knowledge graph-based methods are better than content-based methods for K-CSA. The decline in performance stems from the lack of structural linkage of items between the two domains. This significantly hampers its ability to deliver accurate cross-domain recommendations, especially in a zero-shot setting. Therefore, these findings affirm the value of knowledge graphs K-CSA.
- K-CSA-NC performs the worst in the three variants. The reason is that directly applying and training semantic representations obtained from knowledge graphs leads to models falling into domain bias, especially in the absence of target domain interactions. This elucidates the pivotal role of CSDs in enhancing the robustness and accuracy of K-CSA.
- Based on the result, K-CSA-NS slightly performs slightly worse than K-CSA, which is mainly due to the lack of alignments for the CSDs in source and target domains. To elaborate, the alignments allow the model to relate and contrast the semantic features of items in different domains, thereby enhancing the comprehensiveness and generalization of item CSDs.

Our results show that K-CSA achieves the best performance compared to these variants, further emphasizing the benefits of learning universal CSDs and alignment of source and target domain items through cross-semantic contrastive learning.

5.5 Hyperparameter analysis (RQ4)

In this section, we examine the sensitivity of several important parameters of our K-CSA model as shown in Figure 4.

- **Effect of number of attention layer L for knowledge graph.** This hyperparameter controls the number of neighbor hops for extracting information from the knowledge graph. The larger L means that more knowledge graph information is extracted. Specifically, when the target domains are AB and AM, K-CSA achieves the optimal performance when L is set to 3. This shows that the model extracts favorable information when extracting 3-hop neighbors while preventing semantic noise introduced by too many neighbor hops.
- **Effect of the view N of prototypes.** We investigate the influence of the view count N on our model's performance. It can be instinctively understood that a larger N allows the model to encapsulate the representation of multiple item CSDs. Our findings suggest that the optimal value for N is dependent on the target domain; for the AM domain, the peak performance



(a) L on N10. (b) N on N10. (c) λ_1 on N10.

(d) L on MRR10. (e) N on MRR10. (f) λ_1 on MRR10.

Figure 4: Performance comparison of K-CSA with different source domain datasets.

was observed when $N = 6$, while for the FM domain, optimal results were obtained with $N = 4$. The rationale behind these observations is twofold. On one hand, when N falls short of the optimal threshold, augmenting the view count can enhance the depth and richness of information captured by the model. On the other hand, an excessively large N might lead to the inclusion of superfluous information, which, in turn, might compromise the quality of item CSDs.

- **Effect of the parameter λ_1 of contrastive learning loss.** Further, we validate the effect of hyperparameter λ_1, which coordinates the balance of $\mathcal{L}_{BPR}$ and $\mathcal{L}_{CS}$. According to the results, it is obvious that our method achieves the best performance when λ_1 is set to 0.1. When λ_1 is greater than 0.1, the loss of optimized contrastive learning affects the effect of the core BPR loss, thus causing the performance of the model to decline.

5.6 Case Study (RQ5)

In this section, we conduct case studies to investigate the effectiveness and interpretability of the CSDs.

- **Interpretability of item CSDs.** We perform a case study to show the quality of item CSDs. As shown in Figure 5, the result is derived from the setting that AM is the source domain and AB is the target domain. We observe that the item CSDs from the same user are similar, and the cosine similarity is 0.767. The similarity in CSDs across the same user affirms the strong alignment in item semantics within the same source domain. This inherent characteristic provides a dependable foundation for DZSR, promoting reliable semantic integration. On the contrary, the divergent CSDs among different users point out the individualized semantic interpretations, as denoted by the negative cosine similarity. This deviation presents a compelling narrative of individuality in item usage and perception.
- **Interpretability of user CSDs.** We perform the case study with sampled item and user CSDs as illustrated in Figure 6, to show the interpretability of the learned user CSDs.We show the items the user has interacted with in the source domain and their potential items in the target domain in different settings. To visualize the CSDs, we leverage the t-SNE algorithm [34] to map the mean representation into a two-dimensional vector. The scatterplot analysis in further corroborates that items genuinely engaged by users in the target domain exhibit semantic similarity with those interacted with in the source domain, which is further transferred to user CSDs. Simultaneously, K-CSA shows its adeptness in distinguishing items yet to be engaged within the target domain, thus reinforcing the potency of its discrimination capability.

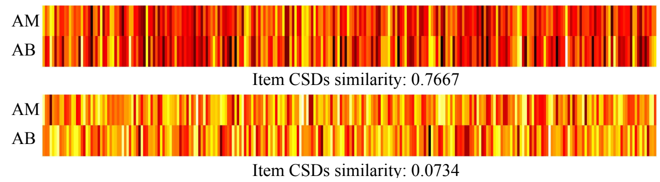
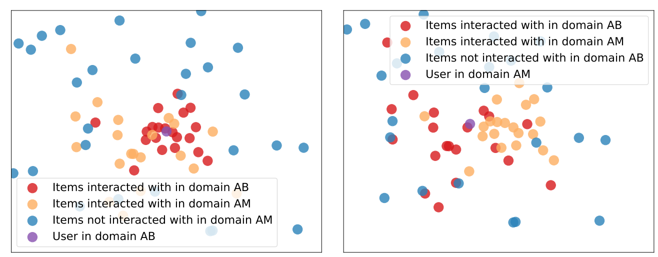


Figure 5: Case study of item CSDs. The upper case shows CSDs from different domains of the same user, and the lower case represents CSDs from different users.



(a) The target domain is AM. (b) The target domain is AB.

Figure 6: Case study of user CSDs. CSDs are mapped to a two-dimensional space through t-SNE.

6 CONCLUSION

In this paper, we demonstrate the main challenge of DZSR in modeling transferable representations in source and target domains. To tackle the problem, we propose a new framework called K-CSA to construct universal user and item CSDs. First, we leverage the knowledge graph to involve semantic information and aggregate the semantic representation through the graph attention layer. Secondly, we introduce the concept of CSDs and conduct multi-view K-means to learn multiple prototypes. Then the item CSDs are reconstructed based on the prototype representations. Furthermore, to alleviate the domain bias since lacking target domain interactions, we design the multi-head self-attention mechanism to integrate the fine-grained item representations and universal item CSDs. Thirdly, we propose the cross-semantic contrastive learning mechanism through prototype distribution. Our extensive experiments validate the effectiveness of our proposed model in different DZSR settings. In our future work, it is interesting to fuse multimodal data to enhance the performance and robustness in non-overlap DZSR. It will extend the application of K-CSA in industrial scenarios.

7 ACKNOWLEDGEMENT

This work was supported by the grant from the National Natural Science Foundation of China (Grant No. 72101176).

REFERENCES

[1] Shlomo Berkovsky, Tsvi Kuflik, and Francesco Ricci. 2007. Cross-domain mediation in collaborative filtering. In *User Modeling 2007: 11th International Conference, UM 2007, Corfu, Greece, July 25-29, 2007. Proceedings 11*. Springer, 355–359.

[2] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. *Advances in neural information processing systems* 26 (2013).

[3] Jiangxia Cao, Xixun Lin, Xin Cong, Jing Ya, Tingwen Liu, and Bin Wang. 2022. Disencdr: Learning disentangled representations for cross-domain recommendation. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 267–277.

[4] Yi Cao, Sihao Hu, Yu Gong, Zhao Li, Yazheng Yang, Qingwen Liu, and Shouling Ji. 2022. GIFT: Graph-guIded Feature Transfer for Cold-Start Video Click-Through Rate Prediction. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*. 2964–2973.

[5] Yixin Cao, Xiang Wang, Xiangnan He, Zikun Hu, and Tat-Seng Chua. 2019. Unifying knowledge graph learning and recommendation: Towards a better understanding of user preferences. In *The world wide web conference*. 151–161.

[6] Qiang Cui, Tao Wei, Yafeng Zhang, and Qing Zhang. 2020. HeroGRAPH: A Heterogeneous Graph Framework for Multi-Target Cross-Domain Recommendation.. In *ORSUM@ RecSys*.

[7] Hao Ding, Yifei Ma, Anoop Deoras, Yuyang Wang, and Hao Wang. 2021. Zero-shot recommender systems. *arXiv preprint arXiv:2105.08318* (2021).

[8] Manqing Dong, Feng Yuan, Lina Yao, Xiwei Xu, and Liming Zhu. 2020. Mamo: Memory-augmented meta-optimization for cold-start recommendation. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*. 688–697.

[9] Yuntao Du, Xinjun Zhu, Lu Chen, Ziquan Fang, and Yunjun Gao. 2022. Metakg: Meta-learning on knowledge graph for cold-start recommendation. *IEEE Transactions on Knowledge and Data Engineering* (2022).

[10] Philip J Feng, Pingjun Pan, Tingting Zhou, Hongxiang Chen, and Chuanjiang Luo. 2021. Zero shot on the cold-start problem: Model-agnostic interest learning for recommender systems. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. 474–483.

[11] Ignacio Fernández-Tobías and Iván Cantador. 2014. Exploiting Social Tags in Matrix Factorization Models for Cross-domain Collaborative Filtering.. In *CBRecSys@ RecSys*. 34–41.

[12] Huifeng Guo, Ruiming Tang, Yunming Ye, Zhenguo Li, and Xiuqiang He. 2017. DeepFM: a factorization-machine based neural network for CTR prediction. *arXiv preprint arXiv:1703.04247* (2017).

[13] Pengcheng He, Xiaodong Liu, Jianfeng Gao, and Weizhu Chen. 2020. Deberta: Decoding-enhanced bert with disentangled attention. *arXiv preprint arXiv:2006.03654* (2020).

[14] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural collaborative filtering. In *Proceedings of the 26th international conference on world wide web*. 173–182.

[15] Benjamin Heitmann and Conor Hayes. 2016. Semstim: Exploiting knowledge graphs for cross-domain recommendation. In *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*. IEEE, 999–1006.

[16] Balázs Hidasi and Alexandros Karatzoglou. 2018. Recurrent neural networks with top-k gains for session-based recommendations. In *Proceedings of the 27th ACM international conference on information and knowledge management*. 843–852.

[17] Meng Jiang, Peng Cui, Nicholas Jing Yuan, Xing Xie, and Shiqiang Yang. 2016. Little is much: Bridging cross-platform behaviors through overlapped crowds. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 30.

[18] SeongKu Kang, Junyoung Hwang, Dongha Lee, and Hwanjo Yu. 2019. Semi-supervised learning for cross-domain recommendation to cold-start users. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. 1563–1572.

[19] Wang-Cheng Kang and Julian McAuley. 2018. Self-attentive sequential recommendation. In *2018 IEEE international conference on data mining (ICDM)*. IEEE, 197–206.

[20] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[21] Hoyeop Lee, Jinbae Im, Seongwon Jang, Hyunsouk Cho, and Sehee Chung. 2019. Melu: Meta-learned user preference estimator for cold-start recommendation. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 1073–1082.

[22] Jingjing Li, Mengmeng Jing, Ke Lu, Lei Zhu, Yang Yang, and Zi Huang. 2019. From zero-shot learning to cold-start recommendation. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 33. 4189–4196.

[23] Danyang Liu, Jianxun Lian, Zheng Liu, Xiting Wang, Guangzhong Sun, and Xing Xie. 2021. Reinforced anchor knowledge graph generation for news recommendation reasoning. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. 1055–1065.

[24] Danyang Liu, Jianxun Lian, Shiyin Wang, Ying Qiao, Jiun-Hung Chen, Guangzhong Sun, and Xing Xie. 2020. KRED: Knowledge-aware document representation for news recommendations. In *Proceedings of the 14th ACM Conference on Recommender Systems*. 200–209.

[25] Huan Liu, Lina Yao, Qinghua Zheng, Minnan Luo, Hongke Zhao, and Yanzhang Lyu. 2020. Dual-stream generative adversarial networks for distributionally robust zero-shot learning. *Information Sciences* 519 (2020), 407–422.

[26] Siwei Liu, Iadh Ounis, Craig Macdonald, and Zaiqiao Meng. 2020. A heterogeneous graph neural model for cold-start recommendation. In *Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval*. 2029–2032.

[27] Yuanfu Lu, Yuan Fang, and Chuan Shi. 2020. Meta-learning on heterogeneous information networks for cold-start recommendation. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 1563–1573.

[28] Tong Man, Huawei Shen, Xiaolong Jin, and Xueqi Cheng. 2017. Cross-domain recommendation: An embedding and mapping approach.. In *IJCAI*, Vol. 17. 2464–2470.

[29] Feiyang Pan, Shuokai Li, Xiang Ao, Pingzhong Tang, and Qing He. 2019. Warm up cold-start advertisements: Improving ctr predictions via learning to learn id embeddings. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 695–704.

[30] Alan Vidotti Prando, Felipe G Contratres, Solange N Alves de Souza, and Luiz S de Souza. 2017. Content-based Recommender System using Social Networks for Cold-start Users.. In *KDIR*. 181–189.

[31] Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084* (2019).

[32] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2012. BPR: Bayesian personalized ranking from implicit feedback. *arXiv preprint arXiv:1205.2618* (2012).

[33] Robyn Speer, Joshua Chin, and Catherine Havasi. 2017. Conceptnet 5.5: An open multilingual graph of general knowledge. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 31.

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

[35] Maksims Volkovs, Guangwei Yu, and Tomi Poutanen. 2017. Dropoutnet: Addressing cold start in recommender systems. *Advances in neural information processing systems* 30 (2017).

[36] Ziyu Wan, Dongdong Chen, Yan Li, Xingguang Yan, Junge Zhang, Yizhou Yu, and Jing Liao. 2019. Transductive zero-shot learning with visual structure constraint. *Advances in neural information processing systems* 32 (2019).

[37] 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*. 417–426.

[38] Hongwei Wang, Fuzheng Zhang, Xing Xie, and Minyi Guo. 2018. DKN: Deep knowledge-aware network for news recommendation. In *Proceedings of the 2018 world wide web conference*. 1835–1844.

[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*. 3307–3313.

[40] Ruoxi Wang, Bin Fu, Gang Fu, and Mingliang Wang. 2017. Deep & cross network for ad click predictions. In *Proceedings of the ADKDD'17*. 1–7.

[41] 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*. 950–958.

[42] Xiang Wang, Dingxian Wang, Canran Xu, Xiangnan He, Yixin Cao, and Tat-Seng Chua. 2019. Explainable reasoning over knowledge graphs for recommendation. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 33. 5329–5336.

[43] Zheng Wang, Jialong Wang, Yuchen Guo, and Zhiguo Gong. 2021. Zero-shot node classification with decomposed graph prototype network. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. 1769–1779.

[44] Likang Wu, Zhaopeng Qiu, Zhi Zheng, Hengshu Zhu, and Enhong Chen. 2023. Exploring Large Language Model for Graph Data Understanding in Online Job Recommendations. *arXiv preprint arXiv:2307.05722* (2023).

[45] Tao Wu, Ellie Ka-In Chio, Heng-Tze Cheng, Yu Du, Steffen Rendle, Dima Kuzmin, Ritesh Agarwal, Li Zhang, John Anderson, Sarvjeet Singh, et al. 2020. Zero-shot heterogeneous transfer learning from recommender systems to cold-start search retrieval. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. 2821–2828.

[46] Yikun Xian, Zuohui Fu, Shan Muthukrishnan, Gerard De Melo, and Yongfeng Zhang. 2019. Reinforcement knowledge graph reasoning for explainable recommendation. In *Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval*. 285–294.

[47] Qin Yue, Jiye Liang, Junbiao Cui, and Liang Bai. 2022. Dual Bidirectional Graph Convolutional Networks for Zero-shot Node Classification. In *Proceedings of*

the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2408–2417.

[48] Chuang Zhao, Hongke Zhao, Ming He, Jian Zhang, and Jianping Fan. 2023. Cross-domain recommendation via user interest alignment. In *Proceedings of the ACM Web Conference 2023*. 887–896.

[49] Hongke Zhao, Yihang Cheng, Xi Zhang, Hengshu Zhu, Qi Liu, Hui Xiong, and Wei Zhang. 2022. What is Market Talking about Market-oriented Prospect Analysis for Entrepreneur Fundraising. *IEEE Transactions on Knowledge and Data Engineering* (2022).

[50] Kangzhi Zhao, Xiting Wang, Yuren Zhang, Li Zhao, Zheng Liu, Chunxiao Xing, and Xing Xie. 2020. Leveraging demonstrations for reinforcement recommendation reasoning over knowledge graphs. In *Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval*. 239–248.

[51] Wayne Xin Zhao, Gaole He, Kunlin Yang, Hongjian Dou, Jin Huang, Siqi Ouyang, and Ji-Rong Wen. 2019. Kb4rec: A data set for linking knowledge bases with recommender systems. *Data Intelligence* 1, 2 (2019), 121–136.

[52] Jiawei Zheng, Qianli Ma, Hao Gu, and Zhenjing Zheng. 2021. Multi-view denoising graph auto-encoders on heterogeneous information networks for cold-start recommendation. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. 2338–2348.

[53] 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*. 1533–1542.

[54] Jianhuan Zhuo, Jianxun Lian, Lanling Xu, Ming Gong, Linjun Shou, Daxin Jiang, Xing Xie, and Yinliang Yue. 2022. Tiger: Transferable Interest Graph Embedding for Domain-Level Zero-Shot Recommendation. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*. 2806–2816.

[55] Lixin Zou, Long Xia, Yulong Gu, Xiangyu Zhao, Weidong Liu, Jimmy Xiangji Huang, and Dawei Yin. 2020. Neural interactive collaborative filtering. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 749–758.