

MeKB-Rec: Personal Knowledge Graph Learning for Cross-Domain Recommendation

Xin Su^{1,†}, Yao Zhou^{1,†}, Zifei Shan^{1,*,†} and Qian Chen¹

¹WeiXin Group of Tencent, ShenZhen, China

Abstract

It is a long-standing challenge in modern recommender systems to make recommendations for new users, namely the cold-start problem. Cross-Domain Recommendation (CDR) has been proposed to address this challenge, but current ways to represent users' interests across systems are still severely limited. We introduce Personal Knowledge Graph (PKG) as a domain-invariant interest representation, and propose a novel CDR paradigm named MeKB-Rec. We first link users and entities in a knowledge base to construct a PKG of users' interests, named MeKB. Then we learn a semantic representation of MeKB for the cross-domain recommendation. Beyond most existing systems, our approach builds a semantic mapping across domains using Pretrained Language Models which breaks the requirement for in-domain user behaviors, enabling zero-shot recommendations for new users in a low-resource domain. We experiment MeKB-Rec on well-established public CDR datasets, and demonstrate that the new formulation achieves a new state-of-the-art that significantly improves HR@10 and NDCG@10 metrics over best previous approaches by 24%–91%, with a 105% improvement for HR@10 of zero-shot users with no behavior in the target domain. We deploy MeKB-Rec in WeiXin recommendation scenarios and achieve significant gains in core online metrics. MeKB-Rec is now serving hundreds of millions of users in real-world products.

Keywords

Cross-Domain Recommendation, Knowledge Graph, Entity Linking, Language Model

1. Introduction

The cold-start problem has been a core challenge for online Recommender Systems for decades. One promising line of methods to deal with the cold-start problem is Cross-Domain Recommendation (CDR) [1, 2], which transfers user and item information from other domains to assist recommendation for the target domain [3, 4, 5]. In this paper, we focus on a common scenario in CDR, where users partially overlap in two domains, and many users have rich behaviors in the source domain but no or few interactions in the target domain. How to effectively make recommendations for such users is a key challenge to improve modern Recommender Systems.

The central problem for CDR lies in learning transferable representations with the help of users that overlap between domains. However, most existing CDR approaches are constrained by a wide-spread hypothesis that embedding representations of users can be sufficiently learned from a few examples of overlapped users. We challenge this hypothesis by raising that, 1) the user representation with merely an embedding has limited

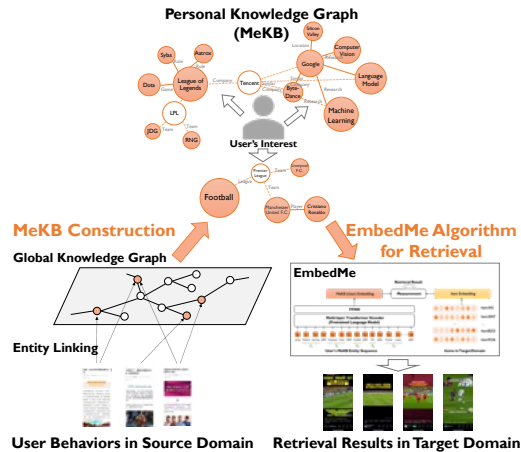


Figure 1: The MeKB-Rec system consists of two steps. The MeKB construction step builds a personal knowledge graph (MeKB) representing a user's interest as a sub-graph of entities. The EmbedMe algorithm step employs a pre-trained language model to map the entity sequence of MeKB for retrieval in the target domain.

transferability, as it tends to weaken the diverse semantics behind users' interactions in the source domain; 2) the data sparsity for overlapped users can severely hinder the learning effectiveness, without help of external knowledge.

To this end, we introduce the concept of Personal Knowledge Graphs (PKGs), originally proposed by Yang et al. [6] to model relations between users and real-world entities. We propose an interest-oriented PKG that mod-

Fifth Knowledge-aware and Conversational Recommender Systems (KaRS) Workshop @ RecSys 2023, September 18–22 2023, Singapore.

*Corresponding author.

[†]These authors contributed equally.

✉ levisu@tencent.com (X. Su); yaozhou@tencent.com (Y. Zhou); zifeishan@tencent.com (Z. Shan); qiancheng@tencent.com (Q. Chen)

0000-0002-5979-8702 (X. Su)

© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

els interest relations between users and entities, named **MeKB**. At the core of its advantages, (1) MeKB is semantically diverse, in that it leverages external information from large Knowledge Graphs (KGs) to connect and organize discrete user interests in the source domain; (2) MeKB is transferable, in that it forms a domain-agnostic graph of user interests that are semantically stable for cross-domain transfers.

To build MeKB, we first perform Entity Linking (EL) [7, 8] on items’ textual content to identify their entities of interest. Then we connect the entities with users through user-item interactions (e.g. likes or reviews) in the source domain. These steps establish MeKB, a sub-graph for the global KG centered around the user to express their interests. MeKB is essentially a general interest storage that represents user interests as world entities.

With MeKB constructed, it remains non-trivial to make cross-domain recommendations with this rich user representation. We notice that Pretrained Language Models (PLMs) are recognized as good few-shot learners [9, 10]. Here we propose **EmbedMe**, a representation learning mechanism to map semantic knowledge in MeKB to fit recommendation tasks in the target domain. By combining MeKB with PLMs, EmbedMe can elevate the semantic knowledge to significantly ease the learning process in CDR, a novel approach to bring world knowledge in PLMs into recommender systems. In particular, EmbedMe first flattens MeKB into a weighted sequence of entities, and then encodes the sequence with a PLM-powered transformer encoder to retrieve recommendation results for the target domain. The construction of MeKB and the EmbedMe mapping model together form the MeKB-Rec system, as illustrated in Figure 1.

We conduct experiments with MeKB-Rec on two real-world recommendation products. We first benchmark on three settings in the Amazon dataset which is commonly used for CDR evaluations [11, 12]. Compared with prior methods, MeKB-Rec achieves the state-of-the-art performance, outperforming other systems by a large margin on HR@10 and NDCG@10 metrics (varying from 24% to 91%). We will publish this processed dataset to facilitate future research in the camera-ready. Furthermore, we conduct offline and online experiments on WeiXin, a popular online platform with billions of users. Besides improvements in offline metrics, MeKB-Rec provides significant DAU and PCTR gains in online A/B testing of WeiXin Live recommendation. MeKB-Rec is now deployed into the product to serve hundreds of millions of users.

2. Related Work

Recommendation systems have widely become the key component for many online information systems in the

past decades. Two key challenges for the growth of recommendation systems are, (1) how to effectively recommend where user-item interactions are limited, namely the data sparsity problem; and (2) how to handle new users or items that enter the system, namely the cold-start problem. [13] To address these problems, Cross-Domain Recommendation (CDR) is an increasingly popular direction. The setting of CDR involves two recommendation domains, namely “source” and “target”, and the key idea is to transfer users’ representations from the source domain to assist recommendation in the target domain. [14]

According to different recommendation settings, CDR approaches are divided into sub-problems: user (or item) non-overlap, user (or item) partial overlap and user (or item) full overlap [2]. This paper studies the scenario of user partial overlap, which is an important case that occurs frequently in real-world recommender systems.

In the user partial-overlapped scenario, existing CDR approaches typically focus on building feature representations of users that can carry over from the source to the target domain. One common approach is Embedding & Mapping. Specifically, a user’s feature in the source domain is represented by an embedding U_i^s , and the user feature in the target domain is represented as U_i^t . During the transfer learning process, the model is trained by overlapped users to learn a mapping function f_U , which maps the feature U_i^s into a downstream-friendly representation in the target domain. [15]. The object is to establish the relationship between the latent space of domains:

$$\min_{\theta} \sum_{u_i \in U} L(f_U(U_i^s; \theta), U_i^t). \quad (1)$$

Embedding-and-mapping approaches aim at transferring representations across domains [16, 17, 18], but the single feature embedding become the bottleneck for expressing a user’s diverse interest. Besides, learning the embedding requires a large amount of examples for overlapped users. In real-world CDR problems, the variety of user features and the lack of overlapped users’ interactions can limit the application scope of the Embedding & Mapping method.

To utilize more user features across domains beyond the traditional feature embedding, another approach is building shared graphs to represent the relationships among users, items, attributes, and other factors. The constructed graph integrates the information over both source and target domains, and by using Graph Neural Networks, the users’ representations are embedded into a shared latent space [19, 20, 21, 22]. Although such methods successfully introduce more features in the source domain and improve CDR performance, they tend to ignore the rich semantics behind users’ discrete interest points, and the insufficient amount of overlapped users’ behavior leads to significant difficulty for optimization.

The goal of this paper is to break the limitation of traditional user representations in existing CDR approaches.

3. Approach

In this section, we demonstrate the proposed approach MeKB-Rec, a two-stage recommendation system, as illustrated in Figure 1. The first stage is building MeKB, a personal knowledge graph that connects users and entities via interest relations, derived from interactions in the source domain. The second stage is EmbedMe, an effective representation learning model that maps MeKB to a downstream embedding for retrieval in the target domain.

3.1. MeKB Construction

MeKB is an instance of Personal Knowledge Graphs (PKG) [23]. It distills user-item interactions into user-entity interest relations by mapping items into entities of interest. Different from prior works that learn latent interest representations, MeKB is an explicit interest store for a user’s interests in the form of entities. The overall process for MeKB construction is shown in Figure 2.

3.1.1. Knowledge Graph.

Knowledge Graphs (KGs) or Knowledge Bases (KBs) were proposed to represent structural world knowledge [24, 25, 26], and have been widely adopted for applications such as Search and Question Answering [27]. There have been efforts to use KGs for recommendation [28, 29], but few work used KGs in the CDR setting. Personal Knowledge Graphs (PKG) were proposed in 2019 [23] as a new concept to model user-entity interactions. Some work adopted PKGs for recommendations [6], though most work focused on news recommendations and few on the CDR problem.

In this paper, we use Wikidata [30] as our KG. The set of all entities in KG is denoted as E . Each $e \in E$ is an entity: persons, organizations, locations, works etc, usually with a Wikipedia page. The set of relations between entities is denoted as R . Each $r \in R$ is a triple between two entities: $(e1, rtype, e2)$ where $e1 \in E$ and $e2 \in E$. Each entity comes with a unique title $Title(e)$ for the given language. Note that in principle, one can replace Wikidata with another KG or augment it with domain-specific KGs to build MeKB.

3.1.2. Entity Linking on Items.

Built upon Wikidata as the global KG, the next step is to connect items with entities of interest in KG. Entity Linking (EL) is the task that links texts with entities in KG. In MeKB construction, it serves as the key step to connect

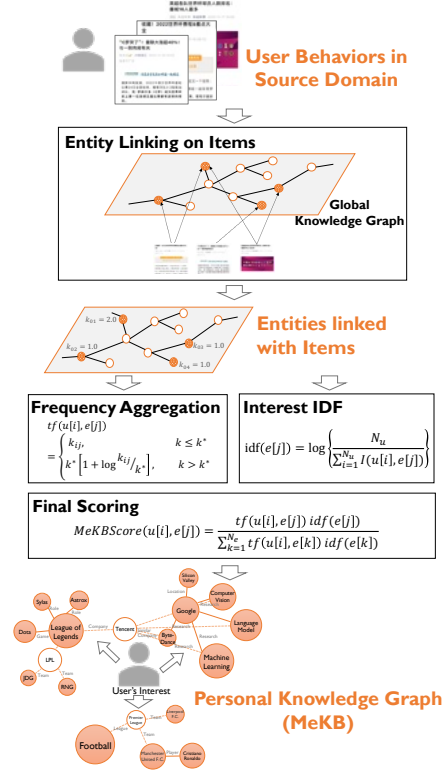


Figure 2: MeKB is constructed based on users’ behaviors in the source domain. We first extract a subgraph of the global KG by performing entity linking on items. Linked entities are weighted by Frequency Aggregation on user-item interactions, and an Interest IDF scoring function.

recommendation items with KG to build its semantic representation.

We formalize the definition of Entity Linking in our scenario. Given KG entities E and a recommendation item $i \in I$, denote the textual content of the item as $Text_i$. An Entity Linking system outputs as set of entities: $\{e_1, e_2, \dots, e_n\}$, where $\forall_{j \in [1, n]} e_j \in E$. Output entities are correlated with the item text, denoted $HasMention$:

$$\forall_{j \in [1, n]} HasMention(Text_i, e_j) \quad (2)$$

In this paper, we use off-the-shelf EL methods for MeKB construction. We follow previous work [31, 32, 33] to extract an Alias Table (AT) from Wikipedia, by parsing Wikipedia internal links, redirections, and page titles. An Alias Table defines the prior probability of a text mention m linking to an entity e , estimated with the formulation:

$$P(e|m) = \frac{freq(m, e)}{freq(m)}, \quad (3)$$

where $freq(m, e)$ denotes the number of anchor texts with the surface form m with reference to the entity page e

in Wikipedia, and $freq(m)$ denotes the total number of anchor texts with the surface form m .

To enable application on a large quantity of items, for the sake of simplicity and efficiency, we directly use the top-1 result from the Alias Table (AT@1) for disambiguation, which is considered a good EL benchmark across languages [34].

As our experiment scenarios span English and Chinese languages, we choose available EL methods in both languages: for English datasets, we first perform named entity recognition (NER) on the textual content of the item, powered by Spacy [35], then perform AT@1 matching [31, 33] with the constructed alias table to map the name to the entity. For Chinese datasets, we directly match against the item’s textual content against the alias table as released in Xu et al. [33].

3.1.3. User-Item Interaction Aggregation.

Naively, by aggregating all positive user-item interactions and collecting entities on items, we can already build the first version of MeKB. However, naive frequency-based aggregations assume that entities appearing more frequently better represent users’ interest, which is often inaccurate in actual recommender systems. To make MeKB better match the characteristics of real-world scenes, we devise an aggregation mechanism to adjust entity weights in the personal knowledge graph.

Concretely, the global KG reflects the global knowledge of the world. KG entities have equal rights, and the edges between entities only reflect the objective world relations. However, for the recommendation system, it remains challenging to model which entities play a better role in representing users’ interests. On one hand, counting the number of items that associate with each entity gives a first estimation of the interest’s strength. On the other hand, this frequency-based measurement may not be ideal as it can overestimate entities with higher global frequency. Therefore, for these two factors, we respectively integrate two interaction aggregation methods into the entity weight of MeKB: 1) Frequency Aggregation and 2) Interest IDF Scoring.

1) Frequency Aggregation. We first set a frequency score for each entity in the original personal knowledge graph, according to the number of user-item interactions that are associated with each entity. For example, a user clicks on k_1 videos associated with entity *Association_football* and k_2 videos with *Manchester_United_F.C.*. Then the initial frequency scores of the corresponding entities are set to k_1 and k_2 . Considering that some entities with extra-high frequency may dominate the head distribution, we perform a log smoothing for the frequency

greater than a certain threshold K , that is

$$s(k) = \begin{cases} k & k \leq k^* \\ k[1 + \log \frac{k}{k^*}] & k > k^* \end{cases} \quad (4)$$

2) Interest IDF Scoring. Considering the deviation of popularity distribution in actual recommendation scenarios, it is beneficial to reduce the weight of entities with higher exposure so that the scores in MeKB are more in line with the actual strength of users’ interests. We are inspired by the concept of IDF [36] from natural language processing, to judge the contribution of an interest entity in recommendation. Intuitively, IDF is correlated with information gain. An IDF-weighted interest score downweights popular interests and reduces exposure bias on the recommendation system. The IDF calculation formula for each entity is as follows.

$$idf(e[j]) = \log \left\{ N_u / \sum_{i=1}^{N_u} I(u[i], e[j]) \right\} \quad (5)$$

where $u[i]$ is the user i and the $e[j]$ is the entity j , and I is an indicator function for whether the interest relationship between $u[i]$ and $e[j]$ exists. We note that $tf(u[i], e[j])$ is the cooccurrence frequency of $u[i]$ and $e[j]$ as calculated by the Frequency Aggregation step, that is

$$tf(u[i], e[j]) = s(k_{ij}) \quad (6)$$

Then we perform IDF on the entity frequency score to obtain the final interest score of MeKB, denoted *MeKBScore*:

$$MeKBScore(u[i], e[j]) = \frac{tf(u[i], e[j]) \cdot idf(e[j])}{\sum_{k=1}^m tf(u[i], e[k]) \cdot idf(e[k])} \quad (7)$$

In short, the Interest IDF and Frequency Aggregation scores are combined with multiplication to compute the final interest weight for each entity in MeKB.

3.2. EmbedMe Algorithm

Although MeKB makes user interest semantically valid across domains, there is still an essential difficulty for learning to make accurate recommendations from MeKB to items in the target domain. The critical challenge is that, although MeKB adopts large knowledge graphs to bring in a large amount of auxiliary entity information and interest associations, the scarcity of CDR samples makes it difficult to implement end-to-end training with this amount of auxiliary features. On one hand, MeKB has a holistic view of a user’s interests as a set of entities in the global KG, each of which is semantically meaningful. On the other hand, the overlapped users’ behavior data is not enough to support end-to-end training of recommendation systems for all PKGs in the CDR setting.

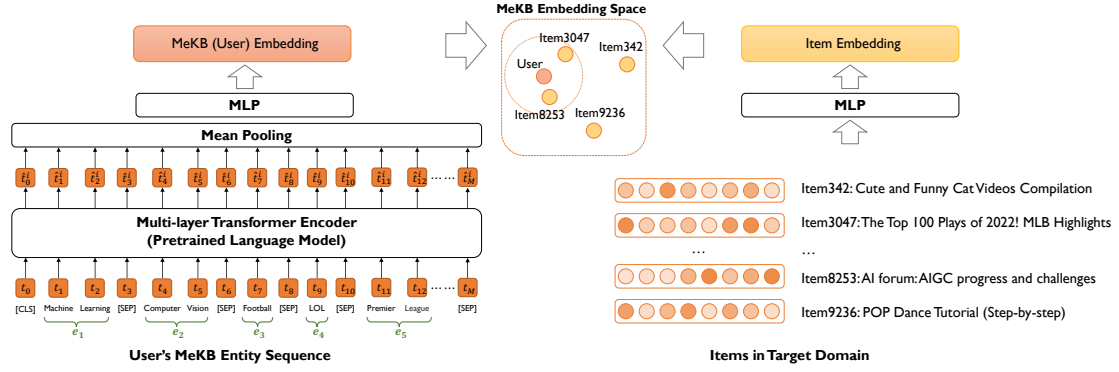


Figure 3: EmbedMe model structure. The user encoder is a PLM on the transformed MeKB sequence input, followed by an MLP projecting to a user embedding of dimension K . The item is simply represented by its ID embedding. Two towers are jointly optimized for retrieval over positive examples in the target domain.

To overcome this challenge, we present an algorithm to map a user’s personal knowledge graph, MeKB, into a dense embedding representation suitable for retrieval in the target domain. We name the algorithm EmbedMe.

To implement EmbedMe, we transform the personal knowledge graph into a semantically weighted sequence representation of entities, and then introduce a Pretrained Language Model (PLM) as the semantic encoder to map the entity sequence to an embedding. We fine-tune this semantic encoder together with the item representation in the target domain via a dual-encoder architecture.

3.2.1. Pretrained Language Model

Natural Language Processing has been enormously successful thanks to the help of Pretrained Language Models (PLMs) such as BERT and GPT [37, 38], leading to significant improvements in Search, Advertisement and Assistant [39]. PLMs are known for their few-shot learning ability [10], which makes them potentially suitable for CDR tasks with limited training examples.

Our core idea is applying PLMs on a textual representation of MeKB to model user interest. The advantage of the PLM is that it can absorb a wide range of corpus for robust semantic understanding, but there is still gap between the pretrained semantic representation and the specific recommendation system. Therefore, we devise an end-to-end fine-tuning method of the PLM for the CDR-specific scenario to extract task-oriented user embedding from MeKB.

3.2.2. MeKB Sequence Encoder.

We employ the PLM and build a user encoder that transforms MeKB to an output embedding of the user, named *MeKB Embedding*. Since MeKB is inherently a graph structure, it becomes challenging for PLMs to exploit the

structure. Therefore, we turn a user u ’s interest graph into a sequence of natural language tokens, ordered by their adjusted score $MeKBscore(u, e)$ in descending order. By encoding over the sequential input, the PLM can learn to encode cross-interest interactions to better fit recommendation tasks. Each entity is represented by its unambiguous title $Title(e)$ (Wikipedia title or name in domain-specific KGs), and tokenized via WordPiece following Devlin et al. [39]. Entities are separated by a special $[SEP]$ token. For efficiency, the entity sequence is truncated to a maximum of N tokens ($N = 256$). We add a mean-pooling layer over PLM’s sequence output on all tokens, followed by a Multi-layer Perceptron (MLP). The MLP transforms the output to an embedding dimension of K ($K = 256$), which is the user’s MeKB Embedding.

3.2.3. End-to-end Training.

For end-to-end training of MeKB Embedding with the specific CDR task, we employ a dual-encoder framework that aligns MeKB Embedding and the learned item’s embedding in the target domain, into the same representation space. The overall training architecture is illustrated in Figure 3.

Item representation. Since item representation is not the focus of the work, we choose to represent items simply by its ID embedding, which is modeled through a weight matrix with dimension $[I, K]$, where I is the number of items in the target domain, and K is the same embedding dimension.

Measurement. The user and item embedding are optimized with dot-product similarity. We define the score between a user and item s as their embeddings’ dot-product similarity:

$$s(u, i) = \phi(u)^T \psi(i), \quad (8)$$

where ϕ is the MeKB sequence encoder, and ψ is the item encoder, projecting user and item representations into

K -dimensional vectors respectively.

Optimization. The dual encoder is trained on positive interactions in the target domain, i.e. records in the target domain that have a non-empty MeKB input. In our experiments, MeKB can be constructed with user interactions in both source and target domains, which yields more available examples than traditional settings where only users with at least a number of actions in both domains are selected for training. [12]

For each positive example, we optimize for full softmax loss [40, 27] over all items with respect to the measurement:

$$l(u, i) = -\log \frac{\exp(s(u, i))}{\sum_{j=0}^I \exp(s(u, i_j))} \quad (9)$$

where I is the number of items in the target domain.

4. Experiments

In this section, we conduct various experiments to compare the recommendation performance of the proposed MeKB-Rec methodology and other CDR baselines. We present the detailed experiment settings, results, and analyses.

4.1. Dataset

We compare MeKB with other baseline methods on two cross-domain recommendation datasets, Amazon [11] and WeiXin.

Amazon Datasets. We selected three popular domains in Amazon: (1) Books, (2) Movies and TV, and (3) CDs and Vinyl, abbreviated as **Books, Movies and Music**. We build three CDR scenarios: *Amazon Books* \rightarrow *Movies*, *Amazon Books* \rightarrow *Music*, and *Amazon Movies* \rightarrow *Music*. We randomly select 20% of the overlapped users and place all their records into the test set, to verify our effectiveness on cold-start users.

WeiXin Datasets. We build large-scale CDR datasets extracted from WeiXin recommendation scenarios. The dataset comes from real user-item interaction on the WeiXin platform, involving three recommendation domains: Articles, Channels (short videos), and Live (live broadcasts). We set up two cross-domain recommendation scenarios that are *WeiXin Articles* \rightarrow *Videos* and *WeiXin Videos* \rightarrow *Live*, and construct CDR datasets with millions of records. The detailed parameters of the Amazon and WeiXin datasets used in the experiments are shown in Table 3. The Overlap Ratio is the fraction of overlapped users of two domains over the number of users in the current domain, i.e. $OR(X, Y) = |U_X \cap U_Y| / |U_X|$.

4.2. Experiment Settings

In the experiments, we randomly divide each dataset into training, validation and test sets according to the ratio of 8:1:1. We also put 20% overlapped users in test set of target domain for cold-start evaluation. In every experiment, we repeat the training and testing process five times and take the average as the final result. In terms of evaluation methods, we adopt **Hit Rate @K** and **NDCG @K** as the evaluation criteria for the recommendation systems. We choose $K=10$ and sample 1 positive item and 999 negative items to compute the metrics for all datasets.

For the implementation of MeKB-Rec, we distribute a total of 1024 examples into 8 NVIDIA A100 GPUs to accelerate training. We adopt Lamb optimizer [41] with a weight decay of $5e-4$, and set its initial learning rate to $5e-4$. We train MeKB-Rec for 15 epochs and choose cosine learning rate schedule with first epoch for warm-up. We set the length of sequence to 256 for all experiments.

4.3. Baseline

We compare our proposed MeKB-Rec method with the following state-of-the-art recommendation methods, among which the first two are general recommendation methods, while others are specifically for CDR:

(1) YoutubeDNN [42] is a popular and efficient DSSM model for recommendation retrieval tasks, not specifically designed for CDR.

(2) CMF: Collective Matrix Factorization [43] adopts a Matrix Factorization (MF) method to simultaneously factorize the rating matrices across the source and target domains, where user representations are shared between two domains.

(3) CoNet: Collaborative Cross Networks [5] replaces matrix factorization with neural networks in CDR problem, and builds a cross-mapping to connect hidden layers in two base networks.

(4) EMCDDR: Embedding-and-Mapping framework for CDR [15] is proposed for inter-domain recommendation. The method applies both MF and Bayesian Personalized Ranking (BPR) for generating latent factors for users and items. A mapping function with linear and non-linear MLP is used to project representations into the same space for overlapping users in both domains.

(5) DCDCSR: a Deep framework for both Cross-Domain and Cross-System Recommendations [44] is proposed to resolve the two challenges. The method is based on Matrix MF models and a fully connected Deep Neural Network (DNN).

(6) DDTCCR: Deep Dual Transfer Cross Domain Recommendation [3] uses a pretrained autoencoder for user and item representation. Users' within-domain preferences and cross-domain preferences are jointly modeled.

Table 1
Experiments on Amazon CDR Dataset

Methods	Books → Movies		Books → Music		Movies → Music	
	HR @ 10	NDCG @ 10	HR @ 10	NDCG @ 10	HR @ 10	NDCG @ 10
YouTubeDNN	0.1271	0.0762	0.1655	0.1134	0.1707	0.1185
CMF	0.1726	0.0995	0.2510	0.1659	0.2615	0.1529
CoNet	0.1364	0.0835	0.1948	0.1363	0.1758	0.1246
EMCDR	0.1755	<u>0.1103</u>	0.2815	<u>0.1819</u>	0.2683	0.1605
DCDCSR	0.1450	0.0826	0.1981	0.1563	0.1850	0.1479
DDTCDR	0.1717	0.0805	0.2839	0.1591	0.2956	0.1666
SSCDR	<u>0.1796</u>	0.0919	<u>0.2925</u>	0.1713	<u>0.3070</u>	<u>0.1807</u>
MeKB-Rec	0.2471	0.1370	0.4884	0.3298	0.5035	0.3444
Improved	+37.58%	+24.21%	+66.97%	+81.31%	+64.01%	+90.59%

Table 2
Experiments on WeiXin CDR Dataset

Methods	WeiXin Articles → Videos		WeiXin Videos → Live	
	HR@10	NDCG@10	HR@10	NDCG@10
YoutubeDNN	0.0733	0.0415	0.1375	0.0816
MeKB-Rec	0.1568	0.0842	0.2919	0.1855
Improved	+113.92%	+102.89%	+112.29%	+127.33%

Note that the auto-encoders are pre-trained on recommendation data without incorporating world semantic knowledge.

(7) SSCDR: Semi-Supervised framework for Cross-Domain Recommendation [45] is proposed to utilize both overlapping users and source-domain items to train the overlapping function.

4.4. Results

The evaluation results on Amazon datasets are shown in Table 1. Results on WeiXin datasets are in Table 2. On all datasets, MeKB-Rec significantly outperforms all other methods by a large margin.

4.4.1. Amazon Dataset

On Amazon Datasets, MeKB-Rec achieves uniform improvements on all settings, a 24%–91% gain over the second-best methods. The advantage in the *Books* → *Movies* scenario is smaller than that in the other two scenarios. We hypothesize that CDR settings with a larger domain gap or a lower user overlap ratio may observe more advantages by using our method.

We further conduct additional ablation studies to validate the effectiveness on PLMs, and show that the pre-trained parameters play a key role on the effectiveness of MeKB-Rec. We conduct a frequency-binned analysis on

user activity, and find that MeKB-Rec is particularly effective for zero-shot and few-shot users. These analyses are in Appendix B.

4.4.2. WeiXin Dataset

On the large-scale WeiXin dataset, we compare results between MeKB-Rec and the established YoutubeDNN baseline. [42] Other methods are not compared due to performance reasons at the scale of this dataset. As shown in Table 2, on both CDR settings tested, MeKB-Rec performs significantly better than YoutubeDNN, and the margin of advantage is larger than that of the Amazon dataset.

4.5. Analysis on MeKB-Rec’s Effectiveness

The proposed MeKB-Rec method uses the personal knowledge graph, MeKB, as a medium for expressing cross-domain interests. As the method strikes CDR performance to a new level, we analyze why it is effective from two standpoints:

1) Domain-agnostic interest representation of MeKB. MeKB uses global entities to represent user interests. Entities are domain-agnostic elements that are semantically stable, therefore can naturally serve as units of interests that are valid across domains, say from *Movies* to *Music*, or from *Articles* to *Videos*.

2) Few-shot knowledge transfer of PLMs. Traditional CDR methods suffer from limited training examples of

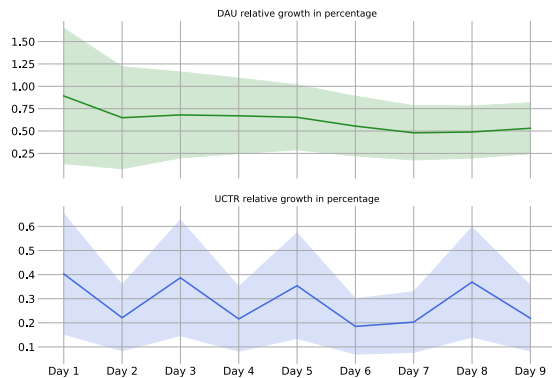


Figure 4: DAU and UCTR relative percentage change of an online A/B experiment using MeKB-Rec for WeiXin Live Recommendation.

overlapping users. As shown in subsection B.2, they fail to generalize to users with no or few activities in the target domain. MeKB-Rec on the other hand, can benefit world knowledge learned from PLMs to understand and transfer the semantics to better serve recommendations in the target domain. The EmbedMe training procedure is essentially a few-shot fine-tuning step of the original PLM to fit the target recommendation task. As PLMs are good few-shot learners, this approach is especially effective. In subsection B.1 we show that by removing the pretrained parameters, the performance drops by 25%–40%.

4.6. Online Experiments

We conduct A/B experiments in the WeiXin online recommendation system serving hundreds of millions of users. We report results in a major Live-stream recommendation scenario in the platform, where we deploy MeKB-Rec to leverage user behaviors in the short video domain to improve the recommendation for live streams, especially for cold-start users.

The control group is the current recommendation system in production. The experiment group is the same system with MeKB-Rec as an additional retrieval step for recommendation candidates. Experiments are run for 9 days, during which the item tower of MeKB-Rec is continuously trained on an hourly basis, to fit recently-popular live broadcasters. We focus on two metrics: (1) DAU of non-active users in Live, and (2) Overall UCTR for all users, which is the unique click-through rate at the user level.

The experimental result is shown in Figure 4. The X-axis is the day of the experiment, and Y-axis is the relative percentage improvement of the experiment group compared to the control. The metric shown in the DAU chart is accumulative while the UCTR chart shows a daily met-

ric. The error bar stands for the 95% confidence interval. Overall, the experiment group with MeKB-Rec achieves a significant gain of DAU +0.531% ($p < 0.01$) for non-active users, suggesting strong improvements over cold-start (zero-shot and few-shot) users. This pool of users accounts for a significant proportion and is critical for the growth of the recommendation scenario. The overall UCTR of all users is improved by +0.240% ($p < 0.01$), indicating that users are generally more in favor of the recommended content compared to the baseline.

As online experiments show significant gains, MeKB-Rec has been deployed to the production system for 100% of non-active users in this scenario.

5. Conclusion

In this paper, we propose a novel knowledge-graph-based method, MeKB-Rec, that significantly improves Cross-Domain Recommendation (CDR) performance and reduces the requirement for overlapped users’ information. MeKB-Rec is a two-phase method, firstly constructing MeKB, a user’s personal knowledge graph, and secondly mapping MeKB entities into a semantic embedding space for recommendations in the target domain, namely the EmbedMe algorithm. MeKB-Rec achieves state-of-the-art recommendation results on multiple CDR scenarios in Amazon and WeiXin. The method also achieved significant gains in online experiments in WeiXin Live recommendation, and is deployed in the product serving hundreds of millions of users.

Specifically, MeKB-Rec has achieved benefits in the following aspects: (1) obtained substantial advantages on large-scale CDR datasets: 38%–67% relative improvement of HR@10 on Amazon and 112%–114% gain of HR@10 on WeiXin datasets; (2) utilized the information contained in the global KG, to supplement the feature representation for user interests that carry across domains; (3) used PTMs to efficiently learn on insufficient training examples. We conclude that injecting global semantics via KG and PTM largely eases learning when the data is sparse. This assumption is validated by MeKB-Rec’s noticeable wins on zero-shot and few-shot user groups (respectively 105% and 66% HR@10 gains on Amazon) insufficient for traditional methods to learn a reliable user representation.

Regarding limitations, we note that the quality of the construction of MeKB is affected by the accuracy of the entity linking method. Specifically, the ambiguity of entity names or the insufficient coverage of entities will cause the representation deviation of MeKB. Future work may exploit more comprehensive entity linking models to improve MeKB’s expressiveness and MeKB-Rec’s overall results.

The work of this paper is an important attempt at

knowledge-graph-based methods in Cross-Domain Recommendations. Future work may focus on: (1) combining knowledge graphs with Large Language Models to further improve recommendations; (2) leveraging more sophisticated content understanding methods to improve the construction of MeKB; (3) better representation learning mechanisms to improve Embedme; and (4) exploring more usage for personal knowledge graphs such as MeKB.

References

- [1] A. I. Schein, A. Popescul, L. H. Ungar, D. M. Pennock, Methods and metrics for cold-start recommendations, in: Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval, 2002, pp. 253–260.
- [2] T. Zang, Y. Zhu, H. Liu, R. Zhang, J. Yu, A survey on cross-domain recommendation: taxonomies, methods, and future directions, arXiv preprint arXiv:2108.03357 (2021).
- [3] P. Li, A. Tuzhilin, Ddtcdr: Deep dual transfer cross domain recommendation, in: Proceedings of the 13th International Conference on Web Search and Data Mining, 2020, pp. 331–339.
- [4] M. Liu, J. Li, G. Li, P. Pan, Cross domain recommendation via bi-directional transfer graph collaborative filtering networks, in: Proceedings of the 29th ACM International Conference on Information & Knowledge Management, 2020, pp. 885–894.
- [5] G. Hu, Y. Zhang, Q. Yang, Conet: Collaborative cross networks for cross-domain recommendation, in: Proceedings of the 27th ACM international conference on information and knowledge management, 2018, pp. 667–676.
- [6] Y. Yang, J. Lin, X. Zhang, M. Wang, Pkg: A personal knowledge graph for recommendation, in: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2022, pp. 3334–3338.
- [7] X. Ling, S. Singh, D. S. Weld, Design challenges for entity linking, Transactions of the Association for Computational Linguistics 3 (2015) 315–328. URL: <https://www.aclweb.org/anthology/Q15-1023>. doi:10.1162/tacl_a_00141.
- [8] W. Zhang, W. Hua, K. Stratos, Entqa: Entity linking as question answering, arXiv preprint arXiv:2110.02369 (2021).
- [9] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, et al., Language models are few-shot learners, Advances in neural information processing systems 33 (2020) 1877–1901.
- [10] T. Gao, A. Fisch, D. Chen, Making pre-trained language models better few-shot learners, in: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Association for Computational Linguistics, Online, 2021, pp. 3816–3830. URL: <https://aclanthology.org/2021.acl-long.295>. doi:10.18653/v1/2021.acl-long.295.
- [11] J. Ni, J. Li, J. McAuley, Justifying recommendations using distantly-labeled reviews and fine-grained aspects, in: Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP), 2019, pp. 188–197.
- [12] J. Cao, X. Lin, X. Cong, J. Ya, T. Liu, B. Wang, Disencdr: Learning disentangled representations for cross-domain recommendation, in: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2022, pp. 267–277.
- [13] Y. Zhu, J. Lin, S. He, B. Wang, Z. Guan, H. Liu, D. Cai, Addressing the item cold-start problem by attribute-driven active learning, IEEE Transactions on Knowledge and Data Engineering 32 (2019) 631–644.
- [14] A. Da’u, N. Salim, Recommendation system based on deep learning methods: a systematic review and new directions, Artificial Intelligence Review 53 (2020) 2709–2748.
- [15] T. Man, H. Shen, X. Jin, X. Cheng, Cross-domain recommendation: An embedding and mapping approach., in: IJCAI, volume 17, 2017, pp. 2464–2470.
- [16] Y. Zhang, Y. Liu, P. Han, C. Miao, L. Cui, B. Li, H. Tang, Learning personalized itemset mapping for cross-domain recommendation., in: IJCAI, 2020, pp. 2561–2567.
- [17] P. Li, A. Tuzhilin, Dual metric learning for effective and efficient cross-domain recommendations, IEEE Transactions on Knowledge and Data Engineering (2021).
- [18] Y. Zhu, K. Ge, F. Zhuang, R. Xie, D. Xi, X. Zhang, L. Lin, Q. He, Transfer-meta framework for cross-domain recommendation to cold-start users, in: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2021, pp. 1813–1817.
- [19] X. Wang, X. He, L. Nie, T.-S. Chua, Item silk road: Recommending items from information domains to social users, in: Proceedings of the 40th International ACM SIGIR conference on Research and Development in Information Retrieval, 2017, pp. 185–194.

- [20] Q. Cui, T. Wei, Y. Zhang, Q. Zhang, Herograph: A heterogeneous graph framework for multi-target cross-domain recommendation., in: ORSUM@ Rec-Sys, 2020.
- [21] J. Li, Z. Peng, S. Wang, X. Xu, P. S. Yu, Z. Hao, Heterogeneous graph embedding for cross-domain recommendation through adversarial learning, in: International Conference on Database Systems for Advanced Applications, Springer, 2020, pp. 507–522.
- [22] K. Xu, Y. Xie, L. Chen, Z. Zheng, Expanding relationship for cross domain recommendation, in: Proceedings of the 30th ACM International Conference on Information & Knowledge Management, 2021, pp. 2251–2260.
- [23] K. Balog, T. Kenter, Personal knowledge graphs: A research agenda, in: Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval, 2019, pp. 217–220.
- [24] D. Vrandečić, M. Krötzsch, Wikidata: a free collaborative knowledgebase, *Communications of the ACM* 57 (2014) 78–85.
- [25] X. Chen, S. Jia, Y. Xiang, A review: Knowledge reasoning over knowledge graph, *Expert Systems with Applications* 141 (2020) 112948.
- [26] C. Zhang, C. Ré, A. Sadeghian, Z. Shan, J. Shin, F. Wang, S. Wu, Feature engineering for knowledge base construction, *IEEE Data Eng Bull* (2014). URL: <https://arxiv.org/abs/1407.6439>.
- [27] T. Févry, L. Baldini Soares, N. FitzGerald, E. Choi, T. Kwiatkowski, Entities as experts: Sparse memory access with entity supervision, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, Online, 2020, pp. 4937–4951. URL: <https://www.aclweb.org/anthology/2020.emnlp-main.400>. doi:10.18653/v1/2020.emnlp-main.400.
- [28] X. Wang, D. Wang, C. Xu, X. He, Y. Cao, T.-S. Chua, Explainable reasoning over knowledge graphs for recommendation, in: Proceedings of the AAAI conference on artificial intelligence, volume 33, 2019, pp. 5329–5336.
- [29] X. Wang, T. Huang, D. Wang, Y. Yuan, Z. Liu, X. He, T.-S. Chua, Learning intents behind interactions with knowledge graph for recommendation, in: Proceedings of the Web Conference 2021, 2021, pp. 878–887.
- [30] D. Vrandečić, Wikidata: A new platform for collaborative data collection, in: Proceedings of the 21st international conference on world wide web, 2012, pp. 1063–1064.
- [31] N. De Cao, G. Izacard, S. Riedel, F. Petroni, Autoregressive entity retrieval, in: International Conference on Learning Representations, 2021. URL: <https://openreview.net/forum?id=5k8F6UU39V>.
- [32] T. Févry, N. FitzGerald, L. B. Soares, T. Kwiatkowski, Empirical evaluation of pretraining strategies for supervised entity linking, in: Automated Knowledge Base Construction, 2020.
- [33] Z. Xu, Z. Shan, Y. Li, B. Hu, B. Qin, Hansel: A chinese few-shot and zero-shot entity linking benchmark, arXiv preprint arXiv:2207.13005 (2022).
- [34] J. A. Botha, Z. Shan, D. Gillick, Entity Linking in 100 Languages, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, Online, 2020, pp. 7833–7845. URL: <https://aclanthology.org/2020.emnlp-main.630>. doi:10.18653/v1/2020.emnlp-main.630.
- [35] M. Honnibal, I. Montani, S. Van Landeghem, A. Boyd, spaCy: Industrial-strength Natural Language Processing in Python (2020). doi:10.5281/zenodo.1212303.
- [36] S. Qaiser, R. Ali, Text mining: use of tf-idf to examine the relevance of words to documents, *International Journal of Computer Applications* 181 (2018) 25–29.
- [37] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805 (2018).
- [38] L. Floridi, M. Chiriatti, Gpt-3: Its nature, scope, limits, and consequences, *Minds and Machines* 30 (2020) 681–694.
- [39] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics, Minneapolis, Minnesota, 2019, pp. 4171–4186. URL: <https://www.aclweb.org/anthology/N19-1423>. doi:10.18653/v1/N19-1423.
- [40] J. Ling, N. FitzGerald, Z. Shan, L. B. Soares, T. Févry, D. Weiss, T. Kwiatkowski, Learning cross-context entity representations from text, arXiv preprint arXiv:2001.03765 (2020). URL: <https://arxiv.org/abs/2001.03765>.
- [41] Y. You, J. Li, S. Reddi, J. Hseu, S. Kumar, S. Bhojanapalli, X. Song, J. Demmel, K. Keutzer, C.-J. Hsieh, Large batch optimization for deep learning: Training bert in 76 minutes, arXiv preprint arXiv:1904.00962 (2019).
- [42] P. Covington, J. Adams, E. Sargin, Deep neural networks for youtube recommendations, in: Proceedings of the 10th ACM conference on recommender systems, 2016, pp. 191–198.
- [43] A. P. Singh, G. J. Gordon, Relational learning via collective matrix factorization, in: Proceedings of the 14th ACM SIGKDD international conference

Table 3
Statistics of the CDR datasets.

Dataset		Users	Overlap Ratio	Items	Records
Amazon	Books	163,189	4.77 %	34,820	2,431,128
	Movies	18,386	42.33 %	24,621	707,037
Amazon	Books	163,189	4.49 %	34,820	2,431,128
	Music	61,229	11.97 %	24,621	707,037
Amazon	Movies	18,386	42.26 %	24,621	707,037
	Music	61,229	12.69 %	55,219	826,490
WeiXin	Articles	2,270,800	52.19%	21,432	13,197,438
	Videos	1,260,491	94.02%	20,617	14,230,730
WeiXin	Videos	1,308,548	55.56%	20,617	10,275,617
	Live	818,773	88.79%	15,030	7,262,557

- on Knowledge discovery and data mining, 2008, pp. 650–658.
- [44] F. Zhu, Y. Wang, C. Chen, G. Liu, M. Orgun, J. Wu, A deep framework for cross-domain and cross-system recommendations, arXiv preprint arXiv:2009.06215 (2020).
- [45] S. Kang, J. Hwang, D. Lee, H. Yu, Semi-supervised learning for cross-domain recommendation to cold-start users, in: Proceedings of the 28th ACM International Conference on Information and Knowledge Management, 2019, pp. 1563–1572.

A. Dataset Details

In this section, we show more details about datasets in the experiments.

Amazon Datasets. The Amazon dataset is a widely used public recommendation dataset. We selected three popular domains in Amazon: (1) Books, (2) Movies and TV, and (3) CDs and Vinyl, abbreviated as **Books, Movies and Music**. In the CDR problem, we use these three domains of data to construct three cross-domain scenarios. Following Cao et al. [12], Liu et al. [4], two domains are combined into one CDR scenario. We combine the domains into three CDR scenarios: *Amazon Books* → *Movies*, *Amazon Books* → *Music*, and *Amazon Movies* → *Music*. The degree of the semantic gap between domains in these scenarios is different, which can verify the robustness of the methods tested. Positive interactions are defined as a review score of 4 or above. Following prior work [12], we filter out users and items based on the frequency of their interactions. Specifically, we filter out users with less than five positive interactions and filter out items with <50 positive interactions as the focus of this work is not on cold-start items. We randomly select 20% of the overlapped users and place all their records into the test set, to verify our effectiveness on cold-start users.

WeiXin Datasets. We build large-scale CDR datasets

extracted from WeiXin recommendation scenarios. The dataset comes from real user-item interaction on the WeiXin platform, involving three recommendation domains: Articles, Channels (short videos), and Live (live broadcasts). We set up two cross-domain recommendation scenarios that are *WeiXin Articles* → *Videos* and *WeiXin Videos* → *Live*, and construct CDR datasets with millions of records. To construct this dataset, we consider reading an article, fully watching a video, and watching a live stream for ≥ 30 seconds as positive behaviors. We randomly sample all positives within two months for training and the week after for testing. Similarly, we filter out items with <50 positives. When constructing the data set, we anonymized the user information, and only kept the user’s anonymous id and sampling behavior samples required for the experiment. Compared with the Amazon dataset, the WeiXin dataset covers a much larger scale of users and items. The detailed parameters of the Amazon and WeiXin datasets used in the experiments are shown in Table 3. The Overlap Ratio is the fraction of overlapped users of two domains over the number of users in the current domain, i.e. $OR(X, Y) = |U_X \cap U_Y| / |U_X|$.

Dataset-specific KG augmentation. Although Wikidata is a general KG with rich world information, complementing it with domain-specific knowledge can be helpful for the recommendation. We further incorporate dataset-specific entities into the KG to be considered in building MeKB, to increase item coverage and enrich user representations. For the Amazon dataset, we include three types of entities available in the dataset: brands, categories, and products, each represented by its textual name. They are considered as additional entities with a lower weight (set to 0.5, 0.3 and 0.1 respectively, where original Wikidata entities are assigned a weight of 1.0) Entity Linking on these entities is trivial, as each item in the dataset already have these entities denoted. For the WeiXin dataset, we include an additional keyword dictionary as domain-specific entities, such as hashtags in videos. Custom entities are associated with items via

Table 4

Impact of Pretrained Language Models in different CDR scenarios in the Amazon dataset.

Methods	Scenario 1		Scenario 2		Scenario 3	
	Books → Movies		Books → Music		Movies → Music	
	HR@10	NDCG@10	HR@10	NDCG@10	HR@10	NDCG@10
w/o PLM	0.1539	0.0818	0.3667	0.2283	0.3430	0.2118
MeKB-Rec	0.2471	0.1370	0.4884	0.3298	0.5035	0.3444
Improved	+60.56%	+67.48%	+33.18%	+44.45%	+46.79%	+62.61%

Table 5

Frequency-binned analysis on user activity. We split users in the Movies test set in the Amazon Books → Movies scenario, based on their frequency seen in training data, and report HR@10 numbers in each bucket. This analysis measures recommendation performance for users with limited or no behaviors for different approaches.

Action Frequency (# of Users)	Zero-shot Users: [0] (1,562)		Few-shot Users: [1, 10] (4,201)		Multi-shot Users: [10, +∞] (12,623)		All Users	
	HR@10	NDCG@10	HR@10	NDCG@10	HR@10	NDCG@10	HR@10	NDCG@10
YouTubeDNN	0.0041	0.0003	0.1225	0.0784	0.2512	0.1402	0.1271	0.0762
CMF	0.0642	0.0315	0.1433	0.0696	0.1912	0.0907	0.1726	0.0995
CoNet	0.0423	0.0196	0.1474	0.0610	0.1735	0.0861	0.1364	0.0835
EMCDR	0.0782	0.0589	0.1545	0.0815	0.2008	0.0945	0.1755	0.1103
DCDCSR	0.0526	0.0345	0.1360	0.0638	0.1807	0.0811	0.1450	0.0826
DDTCDR	0.0573	0.0323	0.1420	0.0704	0.1971	0.0911	0.1717	0.0805
SSCDR	0.0841	0.0521	0.1478	0.0775	0.2072	0.1053	0.1796	0.0919
MeKB-Rec	0.1724	0.0936	0.2563	0.1655	0.2554	0.1411	0.2471	0.1370
Improved	+104.99%	+58.91%	+65.89%	+103.07%	+1.67%	+0.64%	+37.58%	+24.21%

keyword matching.

Following prior work [12, 45], examples in both source and target domains can be used in CDR evaluation. Therefore we combine interactions in both domains with equal weights to construct a user’s MeKB, though in practice one can choose to only use the source domain to construct a pure out-of-domain MeKB for specific tasks. We only use examples in the training set to construct MeKB to make sure there is no data leakage on validation or test examples.

B. Additional Experiments

B.1. Impact of Pretrained Language Models

As mentioned above, MeKB has a strong representation capability of users’ interests, but this capability also brings the challenge of feature mapping for recommendation. Our proposed MeKB-Rec alleviates the problems of insufficient training records with the help of a pre-trained language model. In essence, MeKB embodies a large amount of world knowledge contained in the global KG to build a complete expression of users’ interests. The training records in recommendation scenarios can-

not cover the semantic information of all entities in the global KG. The introduction of the pre-trained language model enables the semantic understanding of entities in the global KG by pre-training on massive external corpora.

This hypothesis is confirmed in an ablation study shown in Table 4. We observe a significant decline in recommendation performance in MeKB-Rec, after dropping PLM parameters obtained during pretraining. The “w/o PLM” setting uses the same model architecture but replaces the PLM parameters with random initialization. This indicates that the world semantic knowledge obtained during pretraining is critical for modeling users’ interest in our model architecture. We suggest that future work may try to further pretrain the PLM on the corpus from the recommendation scenarios to improve the adaptation effect to the recommendation system.

B.2. Frequency-binned Analysis on User Activity

To understand how well our method performs for users with a limited amount of training data, we split users in the Movies test set based on their frequency seen in training data in the target domain, and report HR@10 numbers in each bucket. As shown in Table 5, MeKB-

Rec achieves uniform improvement in all user groups compared to all baselines. Particularly, MeKB-Rec shows a significant advancement for the zero-shot user group (appearing zero times in the training set) and few-shot users (seen less than ten times in training) compared to other methods, with **105%** and **66%** HR@10 gains respectively. MeKB-Rec performs less well on high-activity users with more than 10 training examples, as the baseline methods usually have a dedicated ID embedding that fits these users better after seeing many examples. As expected, MeKB-Rec is better for zero-shot and few-shot users which is a critical challenge in CDR that we aim to address.