Knowledge-aware Neural Collective Matrix Factorization for Cross-domain Recommendation

Anonymous authors

Abstract

Cross-domain recommendation (CDR) can help customers find more satisfying items in different domains. Existing CDR models mainly use common users or mapping functions as bridges between domains but have very limited exploration in fully utilizing item-item relationships across domains. In this paper, we propose the incorporation of multi-domain knowledge graph (KG) to enable items in different domains sharing knowledge even without explicit common contents. To this end, we first construct a new dataset AmazonKG4CDR from the Freebase KG and a subset of Amazon Review Data in three domains: movies, books, and music. This new dataset facilitates linking knowledge to bridge within- and cross-domain items for CDR. Then we propose a new framework, KG-aware Neural Collective Matrix Factorization (KG-NeuCMF), leveraging KG to enrich item representations. It learns item embeddings by graph convolutional autoencoder to capture both domain-specific and domain-general knowledge from adjacent and higher-order neighbours in KG. To further improve KG-aware item embedding, we maximize the mutual information between representations learned from the KG and user-item matrix to establish cross-domain relationships for better CDR. We conduct extensive experiments on the newly constructed dataset and demonstrate that our model significantly outperforms the best-performing baselines.

1. Introduction

Cross-domain recommendation (CDR) [Fernández-Tobíás et al., 2012] is a promising solution to the data sparsity problem in recommender systems. Conventional single-target CDR models leverage information from a richer (source) domain to improve the recommendation performance in a sparser (target) domain [Hu et al., 2013, Yuan et al., 2019b, Berkovsky et al., 2007]. For performance improvement in both domains, recent dual-target CDR models [Man et al., 2017, Zhu et al., 2019, Li and Tuzhilin, 2020] enable bidirectional transfer across domains with dual-learning mechanism [Zhang et al., 2019, He et al., 2016].

Despite encouraging results from existing CDR models, several key issues remain [Zhu et al., 2021]. Firstly, current models, including the dual-target ones, can not simultaneously
Figure 1: Knowledge graph is a natural bridge that connects items from different domains. For example, “Lord of the Ring” in movies can get connected with “Harry Potter” in books via related genre *Fantasy*. Such inter-domain knowledge can reveal similar semantic relations among items from different domains to further improve cross-domain recommendation. This paper constructs a new dataset and proposes a new model to achieve this goal.

improve the performance in both source and target domains due to negative transfer [Pan and Yang, 2009]. In general, the knowledge learned from the sparser domain is less accurate than that learned from the richer domain. Thus, the recommendation performance in the richer domain tends to decline if the transfer direction is simply inverted. The limited information, especially in the sparse domain, is one of the bottlenecks. Secondly, current CDR models mainly use common users [Man et al., 2017, Zhu et al., 2019] or mapping functions [Li and Tuzhilin, 2020] to build connections between domains. In real-life scenarios, relationships between items within or across domains can characterize item-wise semantic relatedness to help understand user-item interaction patterns [Wang et al., 2020]. However, current CDR models are inadequate in capturing such useful item-item relationships.

In this paper, we aim to address this gap by leveraging knowledge graph (KG), a natural bridge for items from different domains [Wang et al., 2017]. KG can benefit the CDR task in multiple ways [Wang et al., 2018]. First, rich and explicit connections among items in a KG can help improve the recommendation performance in each domain, particularly the sparser domain. As shown in Fig 1, a user who has watched “Harry Potter and the Deathly Hallows” is very likely to have interest in the original novel, which can be recommended with the assistance of cross-domain knowledge. Second, domains often share some domain-general information. For example, genre can characterize both book and movie domains. “Lord of the Ring” (from movies), “Harry Potter” (from books), “J. K. Rowling” (from KG), and other entities from different hops of neighbors can be closely connected in KG via the related genre *Fantasy*. KG provides a natural way to leverage such inter-domain knowledge that can help models understand target items by associating rich semantic relatedness among items, explore their latent connections, and further improve recommendation performance.
To build KG-aware CDR, three unique technical challenges arise. (1) Though several
datasets exist for KG-aware single-domain recommendation, no publicly-available dataset
exists for KG-aware CDR. (2) To improve CDR, item (entity) embeddings (representations)
learned from KG should contain both domain-specific and domain-general information,
which typically comes from different hops of neighbors in KG. The second challenge is to
model both adjacent and higher-order relations in the item representation learning process.
(3) Item embeddings learned from KG and those from the user-item interaction matrix
should be closely related, e.g., highly correlated, so that cross-domain relationships can be
effectively established. This is the third challenge.

To address the challenges above, we construct a new dataset for KG-aware CDR and
propose a novel KG-aware Neural Collective Matrix Factorization (KG-aware NeuCMF)
model. More specifically, our contributions are:

- We construct a new dataset named Amazon product Knowledge Graph for CDR (AmazonKG4CDR)
  using a subset (movie, book, and music) of the Amazon Review Data
  (2018) [Ni et al., 2019] and the Freebase KG [Chah, 2017, Bollacker et al., 2008].
  To the best of our knowledge, this is the first dataset that links KG information for CDR.

- We propose a two-step KG-aware NeuCMF framework for KG-aware CDR.
  (1) We train a shared autoencoder using a relational graph convolutional network
      (RGCN) on KG following a contrastive learning-style [Kipf and Welling, 2016,
      Schlichtkrull et al., 2018]. GCN-based encoders learn a node’s embedding by
      aggregating information from its neighbors via non-linear transformation and
      aggregation [Kipf and Welling, 2017b]. Long-range node dependencies can be
      captured by stacking multiple GCN layers to propagate information for multiple
      hops [Xu et al., 2018]. This enables capturing both domain-specific and domain-
      general information from different hops of neighbors in KG.

  (2) To establish cross-domain relationships, the embeddings learned from KG should
      be highly coherent with those from the user-item interaction matrix. Therefore,
      we incorporate the mutual information (MI) estimation [Belghazi et al., 2018]
      into the neural collective matrix factorization (NeuCMF) framework to jointly
      learn the two cross-domain rating matrices by sharing the user embeddings. This
      mechanism allows our model to preserve both user-item interaction and KG
      information across items.

- Finally, we conduct extensive experiments on our newly constructed datasets and
demonstrate that our model significantly outperforms the best-performing baselines.

2. KG Construction for CDR

To develop a knowledge-aware CDR system, a key issue is how to obtain rich and struc-
tured knowledge information for items. Existing research works use side information from
the original recommender system, such as tags and reviews. We argue that the KG informa-
tion will provide additional useful information to the CDR task, since the intra-domain
relationship among items can be captu. In this paper, we present AmazonKG4CDR V1.0,
a new dataset linking KG information for CDR, which can be useful for researchers in the related areas to explore possible approaches with the rich KG information.

We use the widely used dataset, Amazon Review Data (2018) [Ni et al., 2019], covering various domains, from which we select a subset that includes two domain pairs: movie-music, movie-book, which are being linked together through a common user ID identifying the same user. On the KG side, we use the well-known KG: Freebase [Bollacker et al., 2008]. It stores facts by triples of the form \(<head><relation><tail>\). Since Freebase shut down its services, we use its latest public version. We map items into Freebase entities via title matching if there is a mapping available. Fig.2 shows the whole linkage process. Since we only have item Asins (IDs of Amazon products), we need to get items’ titles from the Amazon Review metadata first\(^1\). These titles are later used to get KG entity IDs from \textit{The Knowledge Graph Search API}, which are used to extract the graph information from Freebase. We take triplets that involve two-hop neighbor entities of items into consideration.

During the linkage process, we have dealt with several problems that will affect the quality of the extract knowledge graph. First, the correctness of the extracted KG entity IDs should be ensured. For example, a query is “Harry Potter” (a book name), returned results can be both movies and books. So, we filter returned results by their type and name to ensure extracted IDs are correct. To ensure the KG quality, we preprocess the extracted KG by filtering out infrequent entities (e.g., lower than 10 in both datasets) and retaining the relations appearing in at least 100 triplets.

3. KG-aware NeuCMF models

In this section, we present the technical details of our proposed CDR model, KG-aware Neural CMF (KG-NeuCMF) that aims to improve the performance of CDR by leveraging the KG. We first formulate the task, then present our proposed framework KG-NeuCMF.

3.1 Problem Statement

In this paper, we study the problem of KG-aware CDR. Formally, we are given two domains, a source domain \(S\) (e.g., movie recommendation) and a target domain \(T\) (e.g., book recommendation) that can be represented as two user-item interaction matrices \(R_S\) and \(R_T\), where \(r_{ui} = 1\) indicates that user \(u\) engages with item \(i\), otherwise \(r_{ui} = 0\). In real online shopping platforms (e.g., Amazon), users in domain \(S\) and domain \(T\) often overlap,

\(^1\) https://nijianmo.github.io/amazon/index.html
Figure 3: The framework of our model: KG-aware NeuCMF. It learns item representations from both KG (left) and user-item interaction matrices (right). Entity (item) representations learned from KG contain both domain-specific and domain-general information by utilizing graph autoencoding strategy, which can help assist the CDR task. Item embeddings are learned by a neural CMF model. To ensure the two types of embeddings are highly correlated, we maximize their MI by the neural mutual information estimator (middle).

meaning that they have purchased items in both domains. The set of users in both domains are shared, denoted by \( \mathcal{U} \) (of size \( m = |\mathcal{U}| \)). In our setting, there is no overlap of items between two domains and each item only belongs to one single domain. Denote the set of items in \( \mathcal{S} \) and \( \mathcal{T} \) by \( \mathcal{I}_S \) and \( \mathcal{I}_T \) with size \( n_S = |\mathcal{I}_S| \) and \( n_T = |\mathcal{I}_T| \) respectively. Additionally, we also have a knowledge graph \( \mathcal{G} \), a multi-relational graph, containing rich facts about items. Each fact in the KG is represented as a triple (head entity, relation, tail entity) \( ((h, r, t)) \), also called fact [Wang et al., 2017]. The KG can represent large-scale information from multiple domains [Ehrlinger and Wöß, 2016]. In recommendation scenarios, an item in the user-item interaction matrix corresponds to an entity in KG.

Given \( \mathbf{R}_S \) and \( \mathbf{R}_T \) as well as the knowledge graph \( \mathcal{G} \), we aim to predict whether user \( u \) will engage with item \( i \) with which the user has no interaction before. Our goal is to learn a prediction function \( y_{ui} = f(u, i | \Theta, \mathbf{R}_S, \mathbf{R}_T, \mathcal{G}) \), where \( y_{ui} \) denotes the probability (or the rating score) that user \( u \) will engage with item \( i \) and \( \Theta \) denotes the model parameters of function \( f \).

### 3.2 Methodology

In this subsection, we present the technical details of our proposed model, KG-aware Neural CMF (KG-NeuCMF) that aims to improve the performance of CDR by leveraging the KG. Fig.3 shows the overview of the proposed framework. In the first stage, we propose to learn KG-level representations by exploiting a multi-layer RGCN [Schlichtkrull et al., 2018] through the encode-decode paradigm by minimizing the reconstruction loss that follows a contrastive learning-style convention [Kipf and Welling, 2017a]. This step aims to learn item embeddings containing both domain-specific and domain-general information from different hops of neighbors in KG. In the second-stage, we learn item and user embeddings by borrowing ideas from the CMF framework [Singh and Gordon, 2008] and neural CF (NCF) [He et al., 2017]. Instead of jointly factorizing the two user-item interaction matrices directly as in CMF, we propose to utilize neural networks to jointly learn the two matrices by sharing user latent representations. Finally, item representations learned from KG and
user-item interaction matrix should be highly correlated. To quantify such correlation, we also exploit to maximize MI [Belghazi et al., 2018] between the two types of representations.

3.2.1 Entity embedding learning

To utilize KG in our task, we first need to learn entity representations. We do this by training a graph autoencoder model in the unsupervised fashion and learn representations in an encode-decode paradigm [Kipf and Welling, 2017a, Schlichtkrull et al., 2018]. We employ RGCN [Schlichtkrull et al., 2018] as our encoder that learns an entity embedding by aggregating information from its adjacent neighbors via non-linear transformation and aggregation dependent on the connecting relation, which can be denoted as

\[
f_{en}(e_i^{(l)}, e_j^{(l)}) = \sigma(W_0^{(l)} e_i^{(l)} + \sum_{r \in R} \sum_{j \in N_i^r} \frac{1}{c_{ij}} W_r^{(l)} e_j^{(l)}),
\]

where \( e_i^{(l)}, e_j^{(l)} \) are the hidden state of node \( i \) and node \( j \) in the \( l \)-th layer of the encoder, \( \sigma \) is an activation function such as ReLU, \( W_0^{(l)}, W_r^{(l)} \) are (learnable parameters) relation-specific transformation mapping matrix depending on the type of edge, \( c_{ij} \) is problem-specific normalization constant that can either be learned or chosen in advance, and \( N_i^r \) denotes the set of neighbors of node \( i \) under relation \( r \in R \). Through this operation, the local proximity structure and related semantic information can be successfully captured and stored in the new representation of each entity. Long-range node dependencies can be captured by stacking multiple graph encoder layers and this mechanism ensures that distinct domains can be connected via the information propagation.

The decoder can be any scoring function of KG embedding methods [Wang et al., 2017] that are used to measure the plausibility of each fact \((h, r, t)\). Following [Schlichtkrull et al., 2018], we use DisMult [Yang et al., 2015] factorization as the scoring function, which is well known for its simplicity and efficiency and a triple \((h, r, t)\) is scored as

\[
f_{de}(e_h, r, e_t) = e_h R_r e_t,
\]

where \( e_h, e_t \in \mathbb{R}^d \) are encoded features vector for entity \( h \) and \( t \), and each relation \( r \) is associated with a diagonal matrix \( R_r \in \mathbb{R}^{d \times d} \).

We train the encoder and decoder with negative sampling. We construct an equal number of negative samples by randomly replacing the head entity or tail entity of each positive sample and the overall set of samples are denoted by \( \mathcal{M} \). Then we minimize the cross-entropy loss of positive and negative node pairs

\[
\mathcal{L} = \sum_{(e_h, r, e_t, y) \in \mathcal{M}} (y \log f_{de}(e_h, r, e_t)) + (1 - y) \log (1 - f_{de}(e_h, r, e_t)).
\]
matrices by sharing user latent representations as shown in Fig. 3. The predicted scores in two domains are
\[ r_{si}^S = f_0(f_u(u), f_s(i_s^S)), \]
\[ r_{tj}^T = f_1(f_u(u), f_t(i_t^T)), \]
where \( u, i_s^S \) and \( i_t^T \) are represented one-hot vectors of users, items from domain \( S \) and domain \( T \) respectively, where only the element corresponding to that index is 1 and all others are 0. \( f_u, f_s \) and \( f_t \) can be multi-layer perceptron (MLP) that project the sparse representation to dense vectors. The obtained embeddings are then feed into two separate multi-layer neural architectures to map the latent vectors to predict scores \( r_{si}^S, r_{tj}^T \) for the two domains. Given \( R_S \) and \( R_T \), we minimize the two reconstruction losses \( L_S \) and \( L_T \) with the predicted scores.

The NeuCMF module connects two domains only by the common users, and fails to capture the relations among items. The item embedding learned from KG can capture both domain-specific and domain-general knowledge, thus will be effective for both single-domain and cross-domain recommendation. Intuitively, the learned item embedding from user-item interaction matrices should be highly correlated to the KG-level embeddings. Therefore, this motivates us to exploit to maximize MI [Belghazi et al., 2018] between the two types of representations to guarantee their highly correlated relationship. We design our neural mutual information estimator based on a discriminator \( D(x, y) \) for their pairwise relationships, to provide probability scores for sampled pairs. To be specific, we generate positive samples as \((e_i, i_i) \) (\( i \) can come from domain \( S \) and domain \( T \), half-half) and negative samples are generated by associating sampled items with fake embeddings based on shuffling strategy [Velickovic et al., 2019]. We define the loss function as:
\[ L_{mul} = -\frac{1}{N_{pos} + N_{neg}} \left( \sum_{i=1}^{N_{pos}} \mu(i_i, e_i) \log \sigma(i_i, e_i) + \sum_{i=1}^{N_{neg}} \mu(\tilde{i}_i, e_i) \log \sigma(\tilde{i}_i, e_i) \right), \]
where \( N_{pos}, N_{neg} \) denotes the number of positive and negative samples, \( \mu(\cdot) \) is an indicator function, e.g., \( \sum_{i=1}^{N_{pos}} \mu(i_i, e_i) = 1 \) and \( \sum_{i=1}^{N_{neg}} \mu(\tilde{i}_i, e_i) = 1 \) corresponds to positive and negative pair samples. We aim to minimize \( L_{mul} \), which is equivalent to maximize the mutual information, to jointly preserve the KG-level and user-item interaction information.

The final loss includes: the loss \( L_S \) of source and loss \( L_T \) of target recommendation with the mutual information maximization loss \( L_{mul} \). The objective is to minimize the overall loss \( L \) as follows:
\[ L = L_S(\Theta_S) + L_T(\Theta_T) + L_{mul}(\Theta_{mul}) + \lambda \|\Theta\|, \]
where \( \Theta = \Theta_S \cup \Theta_T \cup \Theta_{mul} \). Note that \( \Theta_S \) and \( \Theta_T \) share user embeddings. The objective function can be optimized by stochastic gradient descent (SGD) and its variants like adaptive moment method (Adam) [Kingma and Ba, 2015].

4. Experiment

**Baselines.** To validate the performance of the proposed model, we compare our model with five representative models, in which two single-domain RS models (MF, NCF) and three
Table 1: Statistics of the dataset.

<table>
<thead>
<tr>
<th>Domain: Music-Movie</th>
<th>Domain: Book-Movie</th>
</tr>
</thead>
<tbody>
<tr>
<td>Users</td>
<td>Books</td>
</tr>
<tr>
<td>4,196</td>
<td>3,977</td>
</tr>
<tr>
<td>Items</td>
<td>Movies</td>
</tr>
<tr>
<td>7,412</td>
<td>11,372</td>
</tr>
<tr>
<td>Interactions</td>
<td>Interactions</td>
</tr>
<tr>
<td>21,986</td>
<td>22,214</td>
</tr>
<tr>
<td>Entities</td>
<td>Entities</td>
</tr>
<tr>
<td>85,612</td>
<td>258,999</td>
</tr>
<tr>
<td>Relations</td>
<td>Relations</td>
</tr>
<tr>
<td>155</td>
<td>127</td>
</tr>
<tr>
<td>Triples</td>
<td>Triples</td>
</tr>
<tr>
<td>288,731</td>
<td>522,814</td>
</tr>
</tbody>
</table>

Table 2: Comparison of recommendation performance in Movie-Music (%).

The best results are in **bold** and the second best ones are underlined.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Movie</th>
<th>Music</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MAE</td>
<td>F1_Score</td>
</tr>
<tr>
<td>MF [Koren et al., 2009]</td>
<td>20.94</td>
<td>74.97±4.50</td>
</tr>
<tr>
<td>NCF [He et al., 2017]</td>
<td>19.01</td>
<td>88.93±0.05</td>
</tr>
<tr>
<td>CMF [Singh and Gordon, 2008]</td>
<td>20.23</td>
<td>89.09±0.36</td>
</tr>
<tr>
<td>CoNET [Hu et al., 2018]</td>
<td>18.22</td>
<td>88.68±0.70</td>
</tr>
<tr>
<td>DDTCDR [Zhu et al., 2019]</td>
<td>20.69</td>
<td>74.84±1.74</td>
</tr>
<tr>
<td><strong>Ours</strong></td>
<td><strong>14.23±0.97</strong></td>
<td><strong>90.69±0.22</strong></td>
</tr>
<tr>
<td><strong>Improvement (%)</strong></td>
<td><strong>21.28 %</strong></td>
<td><strong>1.80 %</strong></td>
</tr>
</tbody>
</table>

CDR models (CMF, CoNet, DDTCDR) using publicly released implementations. Please see appendix Section B for details of these methods and Section C for implementation details.

**Dataset.** We use the Amazon Review Data (2018) [Ni et al., 2019] that is widely used for product recommendation. It contains users’ rate (ranging from 1 to 5) for product from various domains. We select a subset that includes two domain pairs: movie-music (MM), movie-book (MB), which are being linked together through a common user ID identifying the same user. We construct the knowledge graph for each item by utilizing Freebase and the basic statistics details are presented in Table 1. We evaluate the recommendation performance based on MAE, F1_score (Threshold of positive rating is 4).

**Overall Performance of CDR.** We have conducted experiments on two cross domain tasks, movie-music (MM) and movie-book (MB), and the corresponding results of our model and baselines are shown in Table 2 and Table 3. We can see that our proposed model can consistently obtain the best performance across movie-music and movie-book recommendations in terms of MAE and F1_score. In particular, our model improves over the strongest baselines w.r.t. MAE by 21%, 15.18% in movie, music (Table 2) respectively, which justifies the effectiveness of our method in integrating items’ KG information. If we compare between these two tasks, MM and MB, the improvement on music in MM is more remarkable compared to the performance in MB. Possible reasons are 1) the data is more
Table 3: Comparison of recommendation performance in Movie-Book(\%).

The best results are in \textbf{bold} and the second best ones are \underline{underlined}.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Movie</th>
<th>Book</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MAE</td>
<td>F1_Score</td>
</tr>
<tr>
<td>MF [Koren et al., 2009]</td>
<td>24.17±1.32</td>
<td>73.64±0.74</td>
</tr>
<tr>
<td>NCF [He et al., 2017]</td>
<td>18.8±0.54</td>
<td>89.08±0.07</td>
</tr>
<tr>
<td>CMF [Singh and Gordon, 2008]</td>
<td>14.53±1.51</td>
<td>89.32±0.04</td>
</tr>
<tr>
<td>CoNET [Hu et al., 2018]</td>
<td>17.46±0.61</td>
<td>89.59±1.45</td>
</tr>
<tr>
<td>DDTCDR [Zhu et al., 2019]</td>
<td>20.17±0.56</td>
<td>82.60±2.37</td>
</tr>
<tr>
<td>Ours</td>
<td>\textbf{13.17±0.16}</td>
<td>\underline{90.60±0.37}</td>
</tr>
</tbody>
</table>

Improvement (%)

<table>
<thead>
<tr>
<th>Movie</th>
<th>Book</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.36 %</td>
<td>1.12 %</td>
</tr>
<tr>
<td>1.58 %</td>
<td>0.57%</td>
</tr>
</tbody>
</table>

Figure 4: Different ways to incorporate KG information for CDR.

sparse in the user-music interaction matrix, so leveraging KG information can greatly relieve the sparsity problem (we have verified this in the later experiments: Comparisions for cold-start item scenarios); 2) the extracted KG contains much useful information, especially for two closely related domains (movie and music both belong to multi-media datasets). Besides, CDR models (CMF, CoNet, DDTCDR) achieve better performance than SDR models (MF, NCF), indicating that utilizing extra information from other resources benefits the performance of recommendation.

**Different ways to incorporate KG information.** We explore different ways to combine item embeddings learned from KG and user-item interaction matrices. NMF$_{KG}$ takes KG-level embeddings as input, then incorporates them with item embeddings learned from user-item interaction matrices via an aggregation method, e.g., concatenation. NCMF$_{KG,T}$ tries to refine item embeddings learned from KG with a one-layer MLP and concatenates with embeddings learned from the user-item interaction matrix. NCMF$_{KG,mul}$ maximizes MI between the two types of representations to guarantee the highly correlated relationship. The results are shown in Fig. 4. Generally, refining the learned KG-level embeddings gets better performance than direct utilization. This is because in real-world KGs (e.g., Freebase) some noises are inevitably introduced in the process of automatically constructing large-scale KGs due to limited labour supervision [Xie et al., 2018, Jia et al., 2019].
NCMF\textsubscript{KG,mul} gets the best performance. The possible reason is that item embeddings jointly learn from the user-item rating matrix and entity embeddings from KG, which contain both domain-general and domain-specific knowledge and the neural mutual information estimator can ensure their correlation. Such design is more suitable for the cross-domain recommendation task.

**Comparisons for cold-start item scenarios.** KG is a natural bridge for items from different domains, which can further alleviate the item cold-start problem in RS. To validate this, we compare our methods with NCF, CMF under the code-start scenario. We set up the cold-start environment by sampling a subset of items for testing which are unseen in the training data. Results for cold-start items on movie-music datasets are shown in Fig. 5. NCF (the SDR model) is greatly influenced and gets the poorest performance, especially there are a large proportion new items. CMF (the CDR model) can leverage information from two domains, thus it can alleviate the cold-start problem in some extent. Our model goes further to learn representations for cold items from the KG, offering additional information beyond user-item interaction matrices.

5. Conclusion

In this paper, we constructed a new dataset AmazonKG\textsubscript{4CDR}, the first in the field linking KG information for cross-domain recommendation. Moreover, we proposed a KG-aware NeuCMF model to learn domain-specific and domain-general knowledge using graph autoencoding strategy to capture both adjacent and higher-order neighborhood information from KG. Our model unified item embeddings learned from user-item interaction matrices and KG with a neural collaborative filtering framework under a mutual information-based neural estimator. Through extensive experiments on real-world datasets, we demonstrated that KG-aware NeuCMF has achieved substantial gains over state-of-the-art baselines. For future work, we will build a larger dataset with more users, items, and domains, and explore the explainability of cross-domain recommendation.
References


Qingyu Guo, Fuzhen Zhuang, Chuan Qin, Hengshu Zhu, Xing Xie, Hui Xiong, and Qing He. A survey on knowledge graph-based recommender systems. IEEE Transactions on Knowledge and Data Engineering, 2020.


Xiang Wang, Xiangnan He, Yixin Cao, Meng Liu, and Tat-Seng Chua. KGAT: knowledge graph attention network for recommendation. In *SIGKDD*, 2019b.


A. Related Work

A.1 Cross-Domain Recommendation

Different from conventional single-domain recommendation, CDR can leverage information from source domain to improve the performance of target domain [Berkovsky et al., 2007, Fernández-Tobías et al., 2012], namely single-target CDR, which is a powerful tool to deal with the data sparsity problem. These approaches extend the single-domain recommendation models by utilizing same contents, such as tags, reviews [Fernández-Tobías and Cantador, 2014, Yuan et al., 2019a], common items or users [Singh and Gordon, 2008, Hu et al., 2018, Lian et al., 2017] as the bridge between and transfer information between domains [Hu et al., 2013, Loni et al., 2014, Sahebi and Brusilovsky, 2015, Sahebi and Walker, 2014].

The single-target CDR approaches only focus on how to leverage the source domain to help improve the recommendation accuracy on the target one, but not vice versa. Recently, dual-target CDR methods [Man et al., 2017, Zhu et al., 2019, Li and Tuzhilin, 2020] has been proposed to improve the performance on both source and target domains simultaneously by leveraging dual-transfer learning strategies [Zhang et al., 2019, He et al., 2016]. However, as referred to as Negative Transfer [Pan and Yang, 2009], this idea does not work, because, in principle, the knowledge learned from the sparser domain is less accurate than that learned from the richer domain, and thus the recommendation accuracy on the richer domain is more likely to decline by simply and directly changing the transfer direction. Therefore, dual target CDR demands novel and effective solutions. None of the current CDR models can indeed improve the performance on both domains simultaneously, and they are significantly hindered by limited information and connections between two domains.

A.2 Knowledge Graph for Recommendation

In recent years, introducing recommendations with the KG as side information has attracted considerable interest [Wang et al., 2018, 2017, 2019b]. A KG is a heterogeneous graph, where nodes represent as entities, edges represent relations between entities and a fact in KG is usually represented in the form of a triple (head entity, relation, tail entity) [Wang et al., 2017]. KGs contain rich semantic relatedness among items and incorporating KGs in RS can help explore the latent connections and provide explanations for recommended items [Guo et al., 2020]. Currently, KG-aware RS models are only for the single-domain RS [Catherine and Cohen, 2016, Wang et al., 2018, Tang et al., 2019, Zhao et al., 2019, Wang et al., 2017, 2019b]. While one bottleneck for CDR is lacking of connections between domains. since KGs can naturally connect different domains, it would be promising by incorporating KG in the user-item interaction matrix for better cross-domain recommendation performance.

B. Baselines

We compare the performance with five representative models, in which two single-domain RS models (MF, NCF) and three CDR models (CMF, CoNet, DDTCDR) using the publicly released implementations.
Knowledge-aware Cross-domain Recommendation

- **MF** [Koren et al., 2009]. Matrix Factorization (MF) is a classic latent factors CF approach which learns the user and item factors via matrix factorization in each domain separately.

- **NCF** [He et al., 2017]. Neural Collaborative Filtering (NCF) is a neural network architecture to model latent features of users and items using CF method. The NCF models are trained separately for each domain without transferring any information.

- **CMF** [Singh and Gordon, 2008]. Collective Matrix Factorization (CMF) jointly factorizes matrices of each domains. In our scenarios, the shared user factors enable knowledge transfer between cross domains.

- **CoNet** [Hu et al., 2018]. Collaborative Cross Networks (CoNet) enables dual knowledge transfer across domains by introducing cross connections from one base network to another and vice versa.

- **DDTCDR** [Li and Tuzhilin, 2020]. Deep Dual Transfer Cross Domain Recommendation (DDTCDR) learns latent orthogonal mappings across domains and provides cross domain recommendations by leveraging user preferences from all domains.

C. Implementation details

In the KG-pretrain step, we utilize a two-layer RGCN as the encoder to obtain entity embeddings with hidden dimension 16. In the NeuCMF module, we apply one-layer neural networks to project the one-hot vectors of users, and items to low-dimensional embedding vectors and $f_0$ and $f_1$ are two one-layer neural networks to map the latent vectors to predict scores. Throughout the experiments, the embedding size is tuned in the range of [8,16,32] and we use the Adam optimizer [Kingma and Ba, 2015] with learning rate 0.001, L2 regularization 0.0001. For each dataset, the ratio of training, evaluation, and test set is 6 : 2 : 2 [Wang et al., 2019a]. We employ the early stopping strategy based on the validation accuracy with a window size of 10 (we will stop training if the validation loss does not decrease for 10 consecutive epochs) and train 200 epochs at most. We report results over 20 runs with random weight matrix initialization. For a fair comparison, we set the same hyperparameters of the baselines as our model.