Rethinking the Influence of Knowledge Graph Embedding Towards Graph Recommendation

Li Yuexin Li Yuwen Zheng Jinnan Wu Zijian

National University of Singapore

Abstract

To provide a more accurate, diverse, and explainable recommendation, it is compulsory to go beyond modeling user-item interactions and take side information into account. Our work mainly builds on the work of KGAT (Knowledge graph attention network), a paper investigating the utility of knowledge graph, which breaks down the independent interaction assumption by linking items with their attributes. It creates a hybrid structure of KG and user-item graphs to learn high-order relations –an essential factor for successful recommendation. KGAT explicitly models the high-order connectivities in KG (knowledge graph) in an end-to-end fashion, and recursively propagates the embeddings from a node's neighbors (which can be users, items, or attributes) to refine the node's embedding, and employs an attention mechanism to discriminate the importance of the neighbors. To better understand and intensify the utility of the knowledge graph on the KGAT model, we conduct studies on the learning of knowledge graph embedding in the pipeline. Besides TransR – a widely used parametrization method the paper employed, we try abundant methods including TransE, TransD, TransH, Complex, and DisMult on the original three public benchmarks with different information aggregation methods. Based on empirical results, we compare the performances among different methods and analyze the possible reason behind, focusing on the mechanism of these methods.

Background

Recommendation System plays a key role in modern information services, such as search engines, E-commerce platforms, and social medias. The successful application of recommendation systems hugely benefits from the tremendous number of user-item (UI) interactions that are easily obtained by recording the user behavior data. One of the commonly used models in recent years is collaborative filtering (CF), which evaluates similarities between users and items (user/item based CF) , or learns latent vector representations for users and items (model-based CF). However, CF-based methods fail to model the side information of items, such as the profile of a product, the director of a movie, or the genre of a book. In other words, this type of methods overlook higher-order relations of items, which makes them deficient in discovering intuitively relevant items for users. Hence, they are less likely to provide high-quality recommendations in the increasingly complex scenarios.

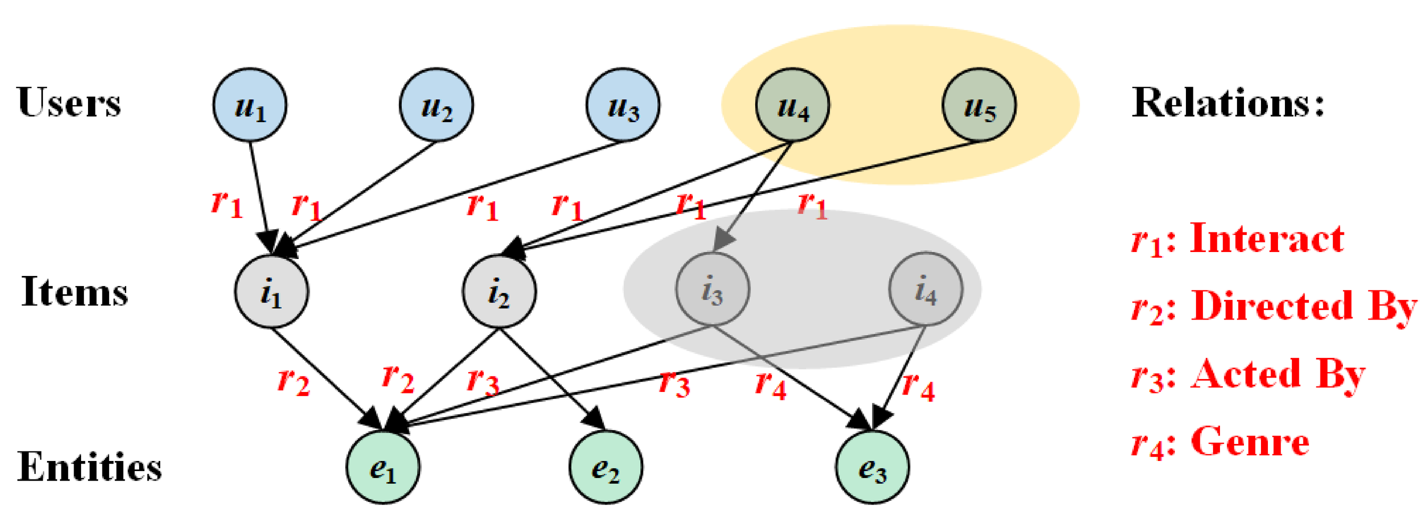

Figure 1. An illustration of collabrative knowledge graph (CKG).

A straightforward approach to utilize side information in recommendation task is introducing knowledge graphs (KGs) into the prediction model. KGs are graph-structured knowledge base that stores a number of real-world entities and their intra-relations. Each of the relational information is represented as a triplet (h, r, t) where $h, t \in \mathcal{E}$, $r \in \mathcal{R}$, $\mathcal{E}$ and $\mathcal{R}$ are the set of KG's entities and relations, respectively.

Essentially, the UI interactions $(u, Interact, i)$ can be abstracted as a bipartite graph, where $u \in \mathcal{U}$ (user set) and $i \in \mathcal{I}$ (item set, a subset of $\mathcal{E}$). It is thus reasonable to combine these two graphs together to better model the complex preference patterns of users. Here, we define collaborative knowledge graph (CKG) $\mathcal{G} = \{(h, r, t)|h, t \in \mathcal{E}', r \in \mathcal{R}'\}$, in which $\mathcal{E}' = \mathcal{U} \cup \mathcal{E}$ and $\mathcal{R}' = \mathcal{R} \cup \{Interact\}$, as it is shown in Figure 1.

In brief, the task can be described as: given a CKG G, how can we effectively generate potential item lists that user u would be most likely interested in?

Method

We now present the enhanced KGAT with various knowledge graph embedding components model, which exploits the influence of different knowledge graph embedding components towards high-order relations representation in an end-to-end fashion. Figure 2 illustrates the model framework with three main components:

1. **Embedding layer,** which exploits high-order connectivities and victories each node by preserving the structure of CKG.
2. **Attentive embedding propagation layer,** which recursively aggregates and propagates embeddings from a node's neighbors following message passing protocol and update its vectorized representation.
3. **Prediction layer,** which gathers user and item representation in a residual way from all propagation layers and outputs the final predicted score.

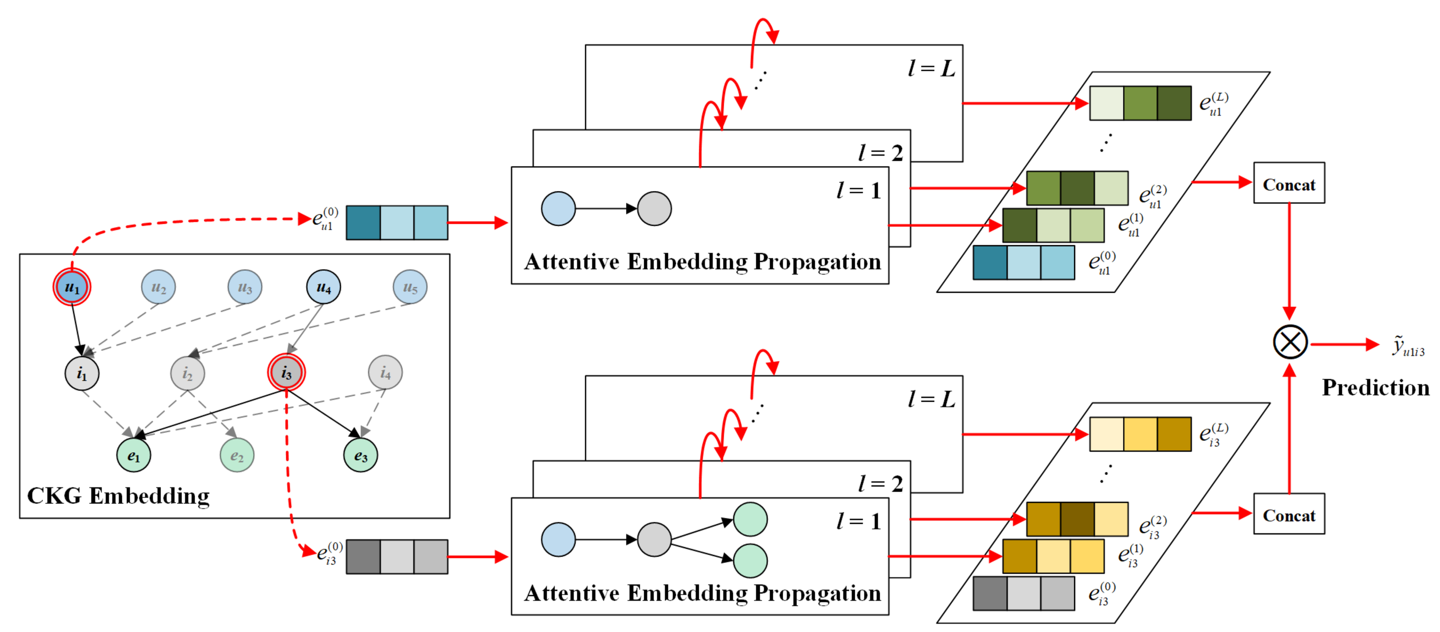

Figure 2. Illustration of knowledge-aware recommendation framework..

Currently, different knowledge graph embedding methods can only model a subset of all relation types efficiently. For instance, TransE has a relatively satisfying expression effect for combinational relations, while ComplEx is better at modeling asymmetric relations. Hence, Our proposed work explores the impact of different knowledge graph embedding methods on the expressiveness of the model.

Commonly used knowledge graph embedding methods are generally divided into translation-based and bilinear-based. To be more specific, translation-based methods learn entity and relation representation by optimizing the translation principle $\mathbf{e_h}^{\mathrm{v}} + \mathbf{e_r} \approx \mathbf{e_t}^{\mathrm{v}}$. The explored translation-based embedding methods' energy function can be formulated as follows:

TransE:

$$g(h, r, t) = \|\mathbf{e_h} + \mathbf{e_r} - \mathbf{e_t}\|_2^2 \tag{1}$$

TransR:

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

TransD:

$$g(h, r, t) = \left\|\mathbf{M_{r,h} e_h} + \mathbf{e_r} - \mathbf{M_{r,t} e_t}\right\|_2^2 \tag{3}$$

TransH:

$$g(h, r, t) = \left\|\mathbf{e_{h,r}} + \mathbf{e_r} - \mathbf{e_{t,r}}\right\|_2^2 \tag{4}$$

Bilinear-based methods directly multiply entity and relations presentations to get the match score:

ComplEx:

$$g(h, r, t) = \|\mathbf{Re}(\mathbf{e_h} \odot \mathbf{e_r} \odot \bar{\mathbf{e_t}})\|_1 \tag{5}$$

DistMult:

$$g(h, r, t) = \|\mathbf{e_h} \odot \mathbf{e_r} \odot \mathbf{e_t}\|_1 \tag{6}$$

Experiment

We evaluate six embedding methods on three datasets (amazon-book, last-fm and yelp2018) with three aggregation methods (bi-interaction, graphsage and GCN). Results of three metrics (precision, recall and NDCG) are reported. Here we only list the results on amazon-book with bi-interaction aggregation for space concern.

	Best Epoch	Precision@20	Recall@20	NDCG@20
TransE	120	0.0144	0.1376	0.0739
TransR	300	0.0150	0.1443	0.0758
TransD	210	0.0142	0.1377	0.0716
TransH	60	0.0147	0.1404	0.0753
Complex	280	0.0146	0.1404	0.0738
DisMult	310	0.0144	0.1397	0.0733

Table 1. Experiment Results of Different Embedding Methods.

Results show that TransR is the best in all metrics. But we think that the performance rank can be dataset dependent. Given the assumption that entity and relation are in the same space, TransE handles 1-to-1 relationship, while TransH handles n-to-1, 1-to-n and n-to-n relationships. Given the assumption that entity and relation are in different spaces, TransR holds that the mapping from entity to relation is only relation dependent, while TransD holds that the mapping depends on both entity and relation. For Complex and DisMult, they are bi-linear models which define scoring functions for triplets. Each method have their own emphasis, so they may be suitable for different datasets. Experiments are still ongoing on different datasets.

Conclusion

Our work builds on the work of KGAT (knowledge graph attention network for recommendation) to further explore the utility of knowledge graph embedding methods. KGAT explores high-order connectivity with semantic relations in CKG (collaborative knowledge graph) for knowledge-aware recommendation. It devised a new framework KGAT, which explicitly models the high order connectivities in CKG in an end-to-end fashion. At its core is the attentive embedding propagation layer, which adaptively propagates the embeddings from a node's neighbors to update the node's representation. We continue to study the effect of using different knowledge graph embedding methods on the performance and conduct analysis based on the mechanisms of these different methods. TransR is proved to be the best among all the KG embedding layers in our experiments.

References

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

[2] Guoliang Ji, Shizhu He, Liheng Xu, Kang Liu, and Jun Zhao. Knowledge graph embedding via dynamic mapping matrix. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 687–696, Beijing, China, July 2015. Association for Computational Linguistics.

[3] Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. Learning entity and relation embeddings for knowledge graph completion. In *Twenty-ninth AAAI conference on artificial intelligence*, 2015.

[4] Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. Complex embeddings for simple link prediction. In *International conference on machine learning*, pages 2071–2080. PMLR, 2016.

[5] Xiang Wang, Xiangnan He, Yixin Cao, Meng Liu, and Tat-Seng Chua. Kgat: Knowledge graph attention network for recommendation. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 950–958, 2019.

[6] Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. Knowledge graph embedding by translating on hyperplanes. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 28, 2014.

[7] Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. Embedding entities and relations for learning and inference in knowledge bases. *arXiv preprint arXiv:1412.6575*, 2014.