Title
CogTrans: A Cognitive Transfer Learning-based Self-Attention Mechanism Architecture for Knowledge Graph Reasoning

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 45(45)

Authors
Wu, Yuejia
Zhou, Jiantao

Publication Date
2023

Peer reviewed

CogTrans: A Cognitive Transfer Learning-based Self-Attention Mechanism Architecture for Knowledge Graph Reasoning

Yuejia Wu and Jian-tao Zhou✉
College of Computer Science, Inner Mongolia University
National & Local Joint Engineering Research Center of Intelligent Information Processing Technology for Mongolian
Engineering Research Center of Ecological Big Data, Ministry of Education
Inner Mongolia Engineering Laboratory for Cloud Computing and Service Software
Inner Mongolia Key Laboratory of Social Computing and Data Processing
Inner Mongolia Key Laboratory of Discipline Inspection and Supervision Big Data
Inner Mongolia Engineering Laboratory for Big Data Analysis Technology
Hohhot, China
wuyuejia@imudges.com and cszjtao@imu.edu.cn

Abstract

Knowledge Graph Reasoning (KGR) is an effective method to solve the incompleteness and sparsity problems of Knowledge Graph (KG), which infers new knowledge based on existing knowledge. Especially, the Graph Convolution Network (GCN)-based approaches can obtain state-of-the-art effectiveness, but there are still some problems such as weak reasoning ability, incomplete local information acquisition, insufficient attention score, and high learning cost, which lead to limited prediction accuracy. This paper proposes a multi-head self-attention mechanism architecture based on cognitive transfer learning, named CogTrans, to make effective improvements in the above problems. Shaped like a cross, CogTrans horizontally includes intuition and reasoning stages, which can achieve a faster convergence rate and obtain prediction results that are more in line with human intuition. Furthermore, CogTrans longitudinally includes source and target domains, and benefit from transfer learning, it can not only obtain the advantages of the horizontal architecture but also can "draw inferences from one instance", which is more conducive to realizing the human brain-like reasoning effect of the architecture. Extensive experimental results show that our CogTrans architecture can obtain the most advanced accuracy of current GCN-based methods.

Keywords: Knowledge Graph; Knowledge Graph Reasoning; Cognitive Science Theory; Attention Mechanism; Transfer Learning

Introduction

A Knowledge Graph (KG) is essentially a semantic network that reveals the relationships between entities. Existing KGs such as Freebase (Bollacker, Evans, Paritosh, Sturge, & Taylor, 2008), DBpedia (Auer et al.,2007), and WordNet (Miller,1992) have been widely used for Knowledge-Based Question Answering (Ye et al., 2021), Semantic Analysis (Al-Obeidat et al., 2020), Personalized Recommendation Systems (Ye et al., 2021), etc. However, the KG is usually incomplete and sparse, which leads to it having some limitations in the above applications.

Knowledge Graph Reasoning (KGR) is one of the main methods to solve these problems. It inferences new knowledge based on the existing knowledge in the KG (Lin, Socher & Xiong, 2018), and its main objectives include link prediction (Zhang & Chen, 2018), fact prediction (Wang et al., 2022), entity prediction (Wickramarachchi, Henson & Sheth, 2022) and relation prediction (Zhao, Li, Hou, & Bai, 2022), etc. In this paper, we will focus on the link prediction approaches. Recent studies on KGR have shown that GCN-based methods can achieve state-of-the-art results, which: (i) modeling relations (Schlichtkrull et al., 2017; Tian et al., 2020; Wang, Zhong & Wang, 2021); (ii) modeling entities and relations jointly (Zhang et al., 2020; Vashishth, Sanyal, Nitin & Talukdar, 2019; Yao, Mao & Luo, 2019). However, there are still exist four kinds of problems. First, a more intellective ability for reasoning at the level of knowledge is lacking. Second, the acquisition of domain information is incomplete. Third, the calculation of the attention score is single. Finally, the cost of learning is significant.

In this paper, we present a cross-shaped multi-head self-attention mechanism architecture based on cognitive transfer learning, referred to as CogTrans. From a horizontal perspective, CogTrans consists of two stages: intuition and reasoning, which can make CogTrans' training more in line with human intuition and carry out knowledge sharing. From a longitudinal perspective, CogTrans has two domains: source and target, which can make CogTrans' training more quickly and effectively learn new knowledge. The main contributions of this paper are as follows:

-We use the dual channel theory of cognitive science to achieve the prediction effect closer to the human brain. In the intuition stage, the graph after hierarchical knowledge expansion is pre-trained by random walk; The reasoning stage integrates and deduces by using a fine-grained multi-head self-attention mechanism to obtain scores for each layer of entities and relations.

-We further establish new associations for the entities and relations of the same level in the hierarchical process, so as to expand the scale of the initial KG, which is more conducive to the acquisition of local information.

-We propose a hierarchical fine-grained multi-head self-attention mechanism encoding pattern, which obtains scores for each layer of entities and relations to fully learn and train our model.

-We apply transfer learning to the GCN-based KGR by treating the KG as a hierarchical structure and taking

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society*.

advantage of this structural similarity to faster and better learn new knowledge, which is conducive to reasoning.

-We conduct plentiful comparison experiments on four benchmark datasets and the results show that our proposed CogTrans architecture can obtain optimal experimental results of existing GCN-based models.

Related Work

In this section, we mainly stated KG's incompleteness and sparsity (Pujara, Augustine & Getoor, 2017) and briefly introduced the solution to these problems based on KGR methods. As shown in Figure 1, we give the general manifestations of incompleteness and sparsity in KG: (i) incompleteness of entities and relations: entity incompleteness is represented by missing knowledge about people related to "David Beckham" and others; relation incompleteness is instantiated in the absence of the "Father & Son" relation of "David Beckham" and "Romeo James Beckham". (ii) Sparsity of entities and relations: it can be obviously found from Figure.1 that the ratio of entities' kind (6) and relations' kind (3) to the number of facts (10) is relatively small.

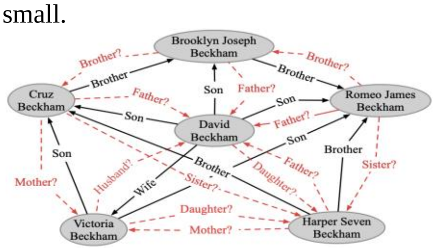

Figure 1: An example of KG.

KGR is one of the main methods to improve the above problems and recent studies have shown that GCN-based methods can obtain the most advanced reasoning results. Next, we give a brief categorical overview of the solutions in these methods: (i) modeling relations, which introduced parameter sharing techniques, strengthened sparsity constraints, and applied them to multigraphs with a large number of relations to better extract topological relation features (Schlichtkrull et al., 2017; Tian et al., 2020; Wang, Zhong & Wang, 2021). (ii) Modeling entities and relations jointly, which jointly embedded both entities and relations in a KG and shared relation embeddings across layers that can effectively aggregate the local neighborhood information of each entity to alleviate the problem of over-parameterization (Zhang et al., 2020; Vashishth, Sanyal, Nitin & Talukdar, 2019; Yao, Mao & Luo, 2019).

Although the existing GCN-based KGR methods can reason knowledge effectively, there are still some problems. Firstly, reasoning ability at the knowledge level is excessively lacking, resulting in insufficient prediction results. Secondly, the acquisition of local information is deficient so that loss much effective reasoning information. Thirdly, the single-layer calculation of the attention score may affect the encoding effect. Finally, it is too expensive to learn directly from scratch in the target domain information. These problems will be described in detail in Problems 1, 2, 3, and 4 in Section.3.2, respectively. This paper focuses on these problems and proposes a cognitive transfer learning-based architecture CogTrans that first expands and pre-trains hierarchical KG, next presents a fine-grained encoding pattern to obtain local domain information effectively, then uses a decoder to accomplish the KGR task, and last use transfer learning achieves quick and efficient learning.

Methodology

Preliminaries

Definition 1 (Knowledge Graph). A Knowledge Graph $\mathcal{G} = (\mathcal{E}, \mathcal{R}, \mathcal{F})$, where $\mathcal{E}$ is the entity set, $\mathcal{R}$ is the relation set, and $\mathcal{F}$ is the fact set. If $\exists h, t \in \mathcal{E}, h \xrightarrow{r} t \in (h, r, t)$, the fact set can be expressed as $\mathcal{F} = \{(h, r, t) | h, t \in \mathcal{E}, r \in \mathcal{R}\}$.

Definition 2 (Link Prediction). Given a partial subset $p \in \mathcal{R}$, a scoring function $f(h, r, t)$ is designed to judge the possibility that any given edge $(h, r, t) \notin p$ belongs to $\mathcal{R}$. In this paper, we use the same $f(h, r, t)$ score as R-GCN, which perform well on standard link prediction bench:

$$f(h, r, t) = p_h^T \mathcal{R}_r p_t \tag{1}$$

And we use the loss function as follows (Schlichtkrull et al., 2017):

$$\mathcal{L} = -\frac{1}{(1+\omega)|\hat{\mathcal{R}}|} \sum_{(h,r,t,y)\in\mathcal{T}} \log l\big(f(h, r, t)\big) y$$
$$+ (1 - y) \log\left(1 - l(f(h, r, t))\right) \tag{2}$$

where ω is a negative set, $\hat{\mathcal{R}}$ is incomplete subset of relations, $\mathcal{T}$ is the whole set of real and corrupted triples, l is a logistic sigmoid function, and the y is an indicator, which set to 0 for negative triples and set to 1 for positive triples.

Problem Presentation and Solution

Based on the problems of GCN-based methods mentioned in the above sections, we provide their detailed description and specific solutions in this section.

Problem 1. When reasoning with the existing GCN-based KGR algorithms, the computer may only find local fragments. So, its ability to reason at the knowledge level and take a panoramic view of the KG is somewhat deficient. Recently, the dual process theory (Ding et al., 2019) is used to improve this problem. As shown in Figure 2, there are two systems in the cognitive system of the human brain: System 1 is an intuitive system, which can find answers quickly and simply through an intuitive match of relevant information; System 2 is an analytic system, which finds answers through certain reasoning and logic.

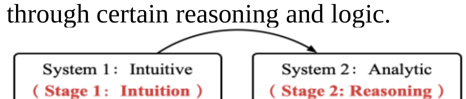

Figure 2: Dual process theory in the cognitive science.

Solution 1. To ameliorate **Problem 1**, we retain the causal structure of the two systems but reorganize their internal components and divide them into two new stages.

As shown in Figure 2, in this paper, the intuitive system corresponds to the intuition stage, which can gain the reasoning ability closer to the human brain by using random walking on hierarchical KG. The analytic system is represented as the reasoning stage, which can achieve more fine-grained knowledge-sharing information and more accurate reasoning results through encoding and decoding.

Problem 2. Most of the existing GCN-based KGR models trained triples independently and simply, as shown in "Pattern 1" in Figure 3, so that the local information cannot be captured adequately, which furthermore limited the reasoning accuracy. Although Zhang et al. (Zhang et al., 2020) collectively trained entities under the same relation to improve the interpretability of the model, as shown in "Pattern 2" in Figure 3, there is still the problem of incomplete local information acquisition.

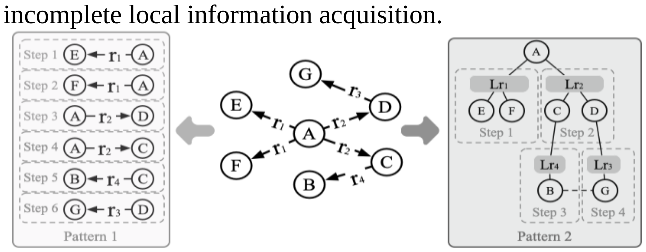

Figure 3: Different training patterns in GCN-based methods.

Solution 2. To improve **Problem 2**, we present to expand the knowledge between entities and relations under the same layer in the pre-training phase to enhance the knowledge-sharing ability and obtain more local information for reasoning.

Problem 3. In the encoding phase, the coarse-grained calculation of attention score only for entities and relationships independently, which may lose some knowledge-sharing information, thus affecting the encoding effect.

Solution 3. To refine **Problem 3**, we propose an encoding structure based on the fine-grained multi-head self-attention mechanism to achieve more accurate score aid reasoning by calculating attention scores for each layer of entities and relations respectively after pre-training.

Problem 4. In a large-scale KG, because of the expensive cost to learn the target domain directly from scratch, it is expected to use the existing relevant knowledge to assist in learning new knowledge as soon as possible, which is the lack of the current GCN-based KGR method.

Solution 4. To solve **Problem 4**, we propose to use the similarity of the hierarchical structure of KG to construct a bridge between old and new knowledge by transferring learning to learn better and faster adaptively.

The Proposed CogTrans Architecture

Overview Based on the solutions mentioned above, we propose the CogTrans architecture to approach KGR. As shown in Figure 4 and Figure 5, the proposed CogTrans architecture can be viewed from two perspectives: horizontal

and longitudinal. Horizontally, as shown in Figure 4, the CogTrans is equipped with a three-phase structure: (i) the first phase is pre-training, which will be elaborated on in the next subsection. It regards KG as a relational hierarchical

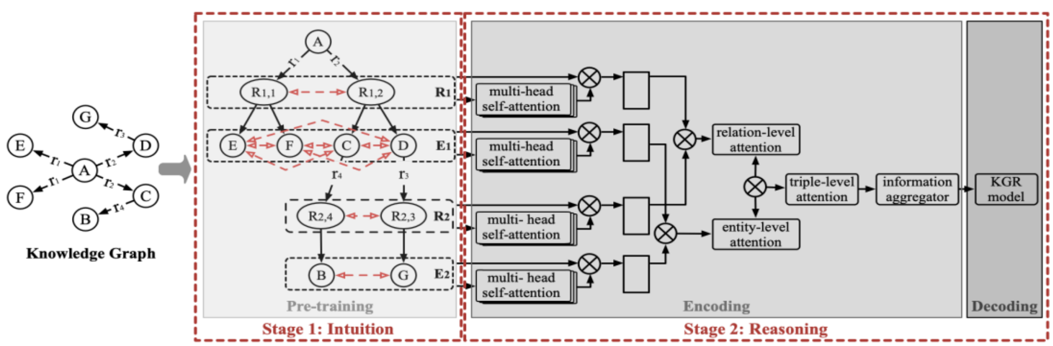

Figure 4: Overview of the CogTrans Architecture with horizontal perspective.

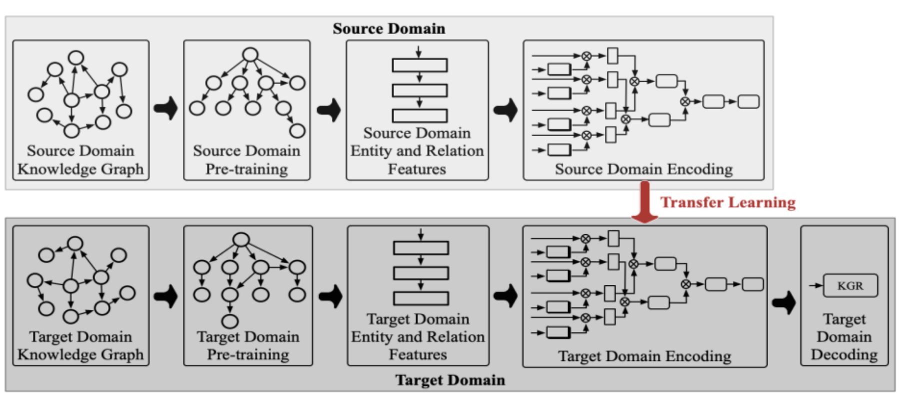

Figure 5: Overview of the CogTrans Architecture with longitudinal perspective.

structure and achieves more abundant local information by extending knowledge under the same hierarchical level, which is more in line with human intuition; (ii) the second phase is encoding, which will be presented in the next subsection. It constructs fine-grained multi-head self-attention for each layer of entities and relations to obtain fine-grained knowledge-sharing information and feature scores. (iii) the third phase is decoding, which will be described in the next subsection. It can be replaced by various existing KGR models that can guarantee our architecture's flexibility and extendibility, i.e., can adapt to the changes in parameters or other variables. Longitudinally, as shown in Figure 5, the CogTrans contains two domains: source and target. By using the similarity of hierarchical KG, it transfers the source domain learning to the target domain and adaptively performs target reasoning.

Pre-training Phase Pre-training is the phase of the intuition stage in CogTrans. As shown in Figure 4, we first regard the KG as a hierarchical structure under the same relation, different from the Zhang et al. (Zhang et al., 2020), then we extend KG's knowledge scale by adding new knowledge between each entity and relation in the same layer, which is marked as red dashed line, such as ("R1,1", "same_layer_rel", "R1,2"), ("E", "same_layer_ent", "F"), ("F", "same_layer_ent", "C") and so on. Finally, we sample the KG by using the random walk strategy, that is, in each step of sampling, an entity $\boldsymbol{e_j}$ is randomly selected from the neighbor entities of the current entity $\boldsymbol{e_i}$ as the next entity to be sampled, with a probability of $\boldsymbol{p_{i,j}}$ as follows:

$$p_{i,j} = \begin{cases} \frac{1}{\mathcal{N}_{e_i}}, & if\ e_j\ is\ a\ neighbor\ of\ e_i \\ 0, & otherwise \end{cases} \tag{3}$$

where $\mathcal{N}_{e_i}$ denotes the neighbor set of e_i.

Encoding Phase Encoding is the first phase of reasoning stage in CogTrans. As shown in Figure 4, based on the previous work of Zhang et al. (Zhang et al., 2020), we propose the hierarchical fine-grained self-attention model to gain more granular scores to help reason, which embeds multi-head self-attention mechanisms for each entity layer and relation layer respectively to obtain composite scores to further facilitate knowledge sharing among triples.

To be specific, the input embedding H, which is created by the pre-training phase, is projected onto corresponding representations Q,K, and V by three matrices $W_Q \in \mathbb{R}^{d\times d_Q}$, $W_K \in \mathbb{R}^{d\times d_K}$ and $W_V \in \mathbb{R}^{d\times d_V}$, where d is hidden dimension. Then, the self-attention is calculated as follows:

$$Q = HW_Q,\ K = HW_K,\ V = HW_V \tag{4}$$

$$Attn(H) = softmax(\frac{QK^T}{\sqrt{d_K}})V \tag{5}$$

where $\frac{QK^T}{\sqrt{d_K}}$ is used to prevent the result from being too large.

Each set of self-attention is used to map the input to a different sub-representation space, which allows the architecture to focus on different locations in diverse sub-representation spaces. The process of multi-head self-attention calculation is shown in Figure 6, which can be expressed as follows:

$$M(Q,K,V) = Concat(head_1, head_2, \ldots, head_n)W^O \tag{6}$$

$$head_i = Attn(QW_i^Q, KW_i^K, VW_i^V) \tag{7}$$

where n is the number of scaled dot-product attention.

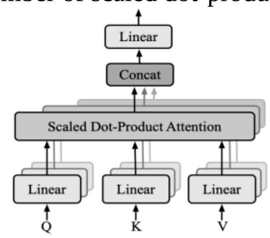

Figure 6: Overview of multi-head self-attention mechanism.

After calculating the hierarchical fine-grained attention scores of different layers, for entity h, the relation-level attention is defined as follows:

$$a_{h,r} = W_1[h \parallel v_r] \tag{8}$$

$$\alpha_{h,r} = softmax_r(a_{h,r}) = \frac{\exp(\sigma(p\cdot a_{h,r}))}{\sum_{r'\in\mathcal{N}_h} exp(\sigma(p\cdot a_{h,r'}))} \tag{9}$$

where W_1 , v_r and p are training parameters, $\parallel$ is a concatenation operation, $\mathcal{N}_h$ is the neighboring relations of h, σ is the LeakyReLU function.

Next, the entity-level attention is computed as follows:

$$b_{h,r,t} = W_2[a_{h,r} \parallel t] \tag{10}$$

$$\beta_{r,t} = softmax_t(b_{h,r,t}) = \frac{\exp(\sigma(q\cdot b_{h,r,t}))}{\sum_{t'\in\mathcal{N}_{h,r}} exp(\sigma(q\cdot b_{h,r,t'}))} \tag{11}$$

where W_2 and q are training parameters, $\mathcal{N}_{h,r}$ is the tail entities of h under relation r.

Then, the triple-level attention is obtained as follows:

$$\mu_{h,r,t} = \alpha_{h,r} \cdot \beta_{r,t} \tag{12}$$

Finally, the information aggregator aggregates the local information to the central entity, and combines it with h, which can be expressed as follows:

$$\hat{h} = \sum_{r\in\mathcal{N}_h} \sum_{t\in\mathcal{N}_{h,r}} \mu_{h,r,t}\, b_{h,r,t} \tag{13}$$

$$h' = \frac{1}{2}\left(\sigma\left(W_3(h+\hat{h})\right) + \sigma\left(W_4(h\odot\hat{h})\right)\right) \tag{14}$$

where W_3 and W_4 are the training parameter, and $\odot$ is the Hadamard multiplication.

Decoding Phase Decoding is the second phase of the reasoning stage in CogTrans. The decoding method can be replaced by most of the existing KGR models to ensure the flexibility and extendibility of our proposed CogTrans architecture. In this paper, we use the RotatE (Sun, Deng, Nie & Tang, 2018) method to approach the link prediction task after encoding to evaluate the performance of our proposed models.

Experiments and Results

Datasets and Experimental Settings

To sufficiently verify the effectiveness of the CogTrans architecture proposed in this paper, we use four different KGR standard datasets: FB15K, WN18, FB15K-237, and

Table 1: Overview of the experimental datasets

Datasets	#Entity	#Relation	#Train	#Test	#Valid
FB15K	14,951	1,345	483,142	59,071	50,000
WN18	40,943	18	141,442	5,000	5,000
FB15K-237	14,541	237	272,115	20,466	17,535
WN18RR	40,943	11	86,835	3,134	3,034

WN18RR in the experiment, as shown in Table 1. Among them, WN18RR and FB15K-237 respectively remove all inverse triples based on WN18 and FB15K. Therefore, the model's performance can be better evaluated by adding these datasets.

Besides, we set layer's number to 6, attention heads to 32, each head's hidden dimension to 16, hidden dimension d to 512, dropout to 0.1, batch size to 1024, warm-up step to 60K, max step to 1M, max epoch to 300, peak learning rate to 3e-4 and so on.

Baselines and Evaluation Metrics

The experiment regards the GCN-based methods as baselines, including ConvE (Dettmers, Minervini, Stenetorp & Riedel, 2017), R-GCN (Schlichtkrull et al., 2017), RotatE (Sun, Deng, Nie & Tang, 2018), KG-BERT (Yao, Mao & Luo, 2019), CompGCN (Vashishth, Sanyal, Nitin & Talukdar, 2019), RGHAT (Zhang et al., 2020), comparing our proposed architecture CogTrans to verify its validity thoroughly.

In the link prediction task, two kinds of standard metrics were used to evaluate the experimental performance, including Mean Reciprocal Ranking (MRR) and $Hits@K$. For each metric, a higher score indicates a better effect. The MRR is calculated as follows:

$$MRR = \frac{1}{|N|}\left(\frac{1}{rank_1} + \frac{1}{rank_2} + \cdots + \frac{1}{rank_{|N|}}\right) = \frac{1}{|N|}\sum \frac{1}{rank_i} \quad (15)$$

where N is the set of triples and $rank_i$ is the link prediction ranking of the i-th triple. In addition, $Hits@K$ $(K = 1,3,10)$ is the average proportion of triples that rank less than K, which is described as follows:

$$Hits@K = \frac{1}{|N|}\sum \mathbb{I}(rank_i \leq K) \quad (16)$$

where $\mathbb{I}(\cdot)$ is an indicator function that if the condition is true, the function value sets to 1, otherwise it sets to 0.

Results and Analysis

The link prediction overall results for the four standard datasets are shown in Table 2 and Table 3. Compared with baselines, the link prediction results of our proposed CogTrans architecture on all experimental datasets are obviously improved. This is because: (i) by establishing new associations for entities and relations at the same level during the layering process of the intuition stage, richer local information can be obtained to enhance inference at the scale of the initial KG; (ii) the fine-grained multi-head self-attention mechanism is proposed to encode in the reasoning stage, which can fully learn and train our model to realize more effective knowledge sharing; (iii) by taking the advantage of the similarity of KG hierarchical structure, transfer learning is designed to transfer source domain information into target domain information, which can learn new knowledge more effectively and improve the accuracy of link prediction.

Table 2 Overview of results on FB15K and WN18

		Metric			
		Hits@K			
Dataset	Model	@1	@3	@10	MRR
FB15K	ConvE	0.558	0.723	0.831	0.657
	R-GCN	0.601	0.760	0.842	0.696
	RotatE	0.746	0.830	0.884	0.797
	KG-BERT	0.761	0.840	0.902	0.811
	RGHAT	0.760	0.843	0.812	0.812
	CogTrans	**0.769**	**0.848**	**0.911**	**0.815**
WN18	ConvE	0.935	0.946	0.956	0.943
	R-GCN	0.697	0.929	0.964	0.819
	RotatE	0.944	0.952	0.959	0.949
	KG-BERT	0.944	0.951	0.952	0.958
	RGHAT	0.949	0.951	0.964	0.954
	CogTrans	**0.958**	**0.965**	**0.968**	**0.964**

Table 3 Overview of results on FB15K-237 and WN18RR

		Metric			
		Hits@K			
Dataset	Model	@1	@3	@10	MRR
FB15K -237	ConvE	0.237	0.356	0.501	0.325
	R-GCN	0.10	0.181	0.30	0.164
	RotatE	0.241	0.375	0.533	0.338
	KG-BERT	0.197	0.289	0.420	0.268
	CompGCN	0.264	0.390	0.535	0.355
	RGHAT	0.462	0.546	0.631	0.522
	CogTrans	**0.465**	**0.551**	**0.637**	**0.533**
WN18 RR	ConvE	0.40	0.44	0.52	0.43
	R-GCN	0.08	0.137	0.207	0.123
	RotatE	0.428	0.492	0.571	0.476
	KG-BERT	0.412	0.465	0.524	0.438
	CompGCN	0.443	0.494	0.546	0.479
	RGHAT	0.425	0.499	0.588	0.483
	CogTrans	**0.452**	**0.506**	**0.593**	**0.488**

Conclusion

In this paper, we proposed a cross-shaped multi-head self-attention mechanism architecture based on the cognitive transfer learning for KGR named CogTrans. Horizontally, CogTrans' intuitive stage completed pre-training in the global KG through a hierarchical random walk to obtain entity and relation characteristics that are more in line with human intuition. Then, in the reasoning stage, a fine-grained multi-head self-attention mechanism is designed to calculate the attention score of each entity layer and relation layer, which is aggregated to complete the encoding for more granular knowledge sharing, and then the existing KGR method is used to decode the reasoning to improve the scalability of the architecture. Longitudinally, CogTrans uses the similarity of the hierarchical structure of the KG to transfer the knowledge from the source domain to the target domain to learn new knowledge faster and more effectively, which so enables our CogTrans model to accomplish the ability to “draw inferences from one instance”. Extensive comparison experiments on classical benchmarks demonstrably certified our CogTrans architecture can achieve the SOTA performance against current GCN-based baselines.

Although this work can make accessibility improvements to current GCN-based models, the prediction scores need to be further advanced. In the future, we are interested in putting forward more novel graph attention mechanisms and cognitive-aware models.

Acknowledgment

This work is supported by the National Natural Science Foundation of China under Grant No.62162046, the Inner Mongolia Science and Technology Project under Grant No.2021GG0155, the Natural Science Foundation of Major Research Plan of Inner Mongolia under Grant No.2019ZD15, the Inner Mongolia Natural Science Foundation under Grant No.2019GG372, and the Special Fund for Graduate Innovation and Entrepreneurship of Inner Mongolia University under Grant No.11200-5223737.

References

Bollacker, K.D., Evans, C., Paritosh, P.K., Sturge, T., & Taylor, J. (2008). Freebase: a collaboratively created graph database for structuring human knowledge. SIGMOD Conference.

Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., & Ives, Z.G. (2007). DBpedia: A Nucleus for a Web of Open Data. ISWC/ASWC.

Miller, G.A. (1992). WordNet: A Lexical Database for English. Commun. ACM, 38, 39-41.

Ye, X., Yavuz, S., Hashimoto, K., Zhou, Y., & Xiong, C. (2021). RNG-KBQA: Generation Augmented Iterative Ranking for Knowledge Base Question Answering. ArXiv, abs/2109.08678.

Al-Obeidat, F.N., Adedugbe, O.A., Bani-Hani, A.I., Benkhelifa, E., & Majdalawieh, M. (2020). Cone-KG: A Semantic Knowledge Graph with News Content and Social Context for Studying Covid-19 News Articles on Social Media. 2020 Seventh International Conference on Social Networks Analysis, Management and Security (SNAMS), 1-7.

Ye, Q., Hsieh, C., Yang, Z., Kang, Y., Chen, J., Cao, D., He, S., & Hou, T. (2021). A unified drug–target interaction prediction framework based on knowledge graph and recommendation system. Nature Communications, 12.

Lin, X.V., Socher, R., & Xiong, C. (2018). Multi-Hop Knowledge Graph Reasoning with Reward Shaping. Conference on Empirical Methods in Natural Language Processing.

Zhang, M., & Chen, Y. (2018). Link Prediction Based on Graph Neural Networks. ArXiv, abs/1802.09691.

Wang, Z., Sun, Y., Hu, X., Zhao, J., Yang, Z., & Lin, H. (2022). A Semantic Network Encoder for Associated Fact Prediction. IEEE Transactions on Knowledge and Data Engineering, 34, 5114-5125.

Wickramarachchi, R., Henson, C.A., & Sheth, A. (2022). Knowledge-Based Entity Prediction for Improved Machine Perception in Autonomous Systems. IEEE Intelligent Systems, 37, 42-49.

Zhao, F., Li, Y., Hou, J., & Bai, L. (2022). Improving question answering over incomplete knowledge graphs with relation prediction. Neural Computing and Applications, 34, 6331 - 6348.

Pujara, J., Augustine, E., & Getoor, L. (2017). Sparsity and Noise: Where Knowledge Graph Embeddings Fall Short. Conference on Empirical Methods in Natural Language Processing.

Schlichtkrull, M., Kipf, T., Bloem, P., Berg, R.V., Titov, I., & Welling, M. (2017). Modeling Relational Data with Graph Convolutional Networks. Extended Semantic Web Conference.

Tian, A., Zhang, C., Rang, M., Yang, X., & Zhan, Z. (2020). RA-GCN: Relational Aggregation Graph Convolutional Network for Knowledge Graph Completion. Proceedings of the 2020 12th International Conference on Machine Learning and Computing.

Wang, S., Zhong, Y., & Wang, C. (2021). Attention Relational Graph Convolution Networks for Relation Prediction in Knowledge Graphs. Journal of Physics: Conference Series, 1848.

Zhang, Z., Zhuang, F., Zhu, H., Shi, Z., Xiong, H., & He, Q. (2020). Relational Graph Neural Network with Hierarchical Attention for Knowledge Graph Completion. AAAI Conference on Artificial Intelligence.

Vashishth, S., Sanyal, S., Nitin, V., & Talukdar, P.P. (2019). Composition-based Multi-Relational Graph Convolutional Networks. ArXiv, abs/1911.03082.

Yao, L., Mao, C., & Luo, Y. (2019). KG-BERT: BERT for Knowledge Graph Completion. ArXiv, abs/1909.03193.

Ding, M., Zhou, C., Chen, Q., Yang, H., & Tang, J. (2019). Cognitive Graph for Multi-Hop Reading Comprehension at Scale. ArXiv, abs/1905.05460.

Sun, Z., Deng, Z., Nie, J., & Tang, J. (2018). RotatE: Knowledge Graph Embedding by Relational Rotation in Complex Space. *ArXiv, abs/1902.10197*.

Dettmers, T., Minervini, P., Stenetorp, P., & Riedel, S. (2017). Convolutional 2D Knowledge Graph Embeddings. AAAI Conference on Artificial Intelligence.