Entity Alignment for Cross-lingual Knowledge Graph with Graph Convolutional Networks

Fan Xiong and **Jianliang Gao***
School of Computer Science and Engineering, Central South University, China
csufanxiong@gmail.com, gaojianliang@csu.edu.cn

Abstract

Graph convolutional network (GCN) is a promising approach that has recently been used to resolve knowledge graph alignment. In this paper, we propose a new method to entity alignment for cross-lingual knowledge graph. In the method, we design a scheme of attribute embedding for GCN training. Furthermore, GCN model utilizes the attribute embedding and structure embedding to abstract graph features simultaneously. Our preliminary experiments show that the proposed method outperforms the state-of-the-art GCN-based method.

1 Introduction

Knowledge graphs (KGs), aiming to represent human knowledge in structural forms, are playing an increasingly important role in AI-related and NLP-related applications, such as question answering [Aditya *et al.*, 2018]. Typically, KGs represent a collection of knowledge facts and are quite popular in the real world [Fang *et al.*, 2017]. Each fact is represented as a triplet (h, r, t), meaning that the head entity h has the relation r with the tail entity t. However, the complex structure and a large number of attributes of KGs often prevent us from getting the hidden information in the graphs. Therefore, more and more studies focus on the cross-lingual KG alignment.

Recently, graph convolutional network (GCN) has emerged for bearing on a large class of graph-based learning problems. Wang *et al.* put forward cross-lingual KG alignment method through GCN-based on pre-aligned entities [Wang *et al.*, 2018]. However, many attribute information does not play a role in the alignment process, but brings a negative impact on the overall alignment work, because the attributes of the same entity in different languages are quite different. Wang et al. propose heterogeneous graph attention network, including node-level and semantic-level attentions, which could not only learn the importance between nodes, but also learn the importance of different meta-paths [Wang *et al.*, 2019]. However, this method only considers the influence of different nodes, specifying different weights to nodes in a neighborhood, but not considers the impact of different attributes.

*Contact Author

Having the above observations, we propose the approach of cross-lingual KG alignment. It utilizes attribute embedding by using aggregation function of GCN.

2 The Proposed Method

Our method takes full advantage of the characteristics of attribute information in alignment and makes the attribute embedding matrix more suitable for GCN training.

2.1 Overview the Method with Graph Convolutional Networks

As shown in Fig. 1, a GCN-model consists of multiple stacked GCN layers. The $(l + 1)$th layer of the GCN model is based on the output of previous layer. The convolutional computation is as follows:

$$H^{(l+1)} = \sigma(\widehat{Q}^{-\frac{1}{2}}\widehat{P}\widehat{Q}^{-\frac{1}{2}}H(l)W(l)) \quad (1)$$

where P is a $n \times n$ adjacency matrix, n is the number of nodes, $\widehat{P} = P + I$, where I is the identity matrix and $\widehat{Q}$ is the diagonal node degree matrix of $\widehat{P}$, $H(l)$ is a vertex feature matrix which input to the l-th layer of the GCN-model. $W(l)$ is a weight matrix for the l-th neural network layer and σ is a non-linear activation function like the $RELU$ [Wang *et al.*, 2018].

We embed the entity attribute information of different languages into a unified vector space. In order to improve the accuracy, we design the following method of enhancing attribute (EA) embedding, which reduces the difference between equivalent entities.

2.2 Enhance Attribute Embedding

We enhance attribute embedding to make the AE (attribute embedding) vector more suitable for GCN training. The attributes of the aligned entities in different languages may be quite different due to the data characteristics of KGs, which misleads GCN training. EA embedding obtains the embedding of attribute features in the following steps:

Choose Attributes

In the training of GCN, we observe that the numbers of attributes are critical to the results. Therefore, we propose a method to choose attributes for GCN training. Firstly, the attributes are ordered descending by the numbers they appears

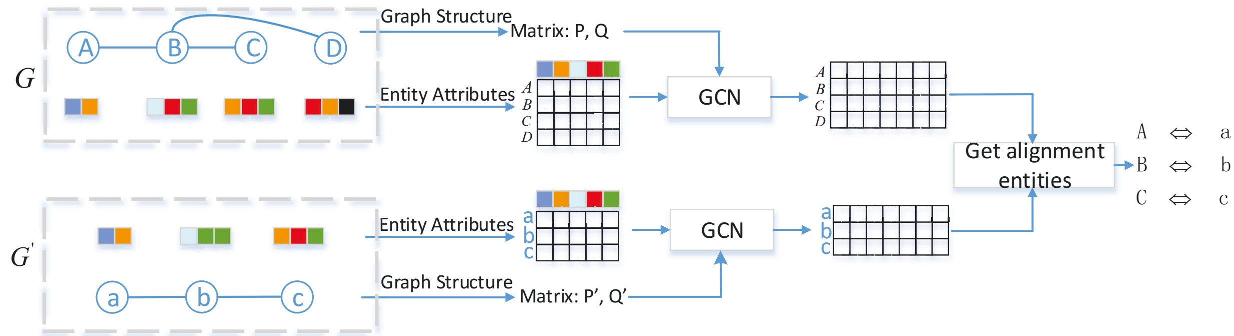

Figure 1: The process of cross-lingual knowledge graph alignment

Language		ZH → EN			EN → ZH		
Metric		Hits@1	Hits@10	Hits@50	Hits@1	Hits@10	Hits@50
GCN [Wang *et al.*, 2018]	AE	14.16	44.35	73.05	13.62	43.32	71.77
	SE+AE	41.25	74.38	86.23	36.49	69.94	82.45
The proposed method	AE	19.51	51.68	77.41	18.03	49.37	75.37
	SE+AE	45.16	77.93	88.64	37.90	71.71	83.63

Table 1: Result comparison of cross-lingual KG alignment

in entities. Then, we get the intersect of the ordered attributes belonging to two lingual knowledge graphs. Finally, the top-k attributes of the intersect are chosen as the attributes for embedding.

Weighting Attributes

In order to distinguish the different importance of attributes, Further, we weight the selected attributes. Eq. 2 is the method for weighting attribute β:

$$w_\beta = \alpha\left(1 - \frac{\left|n_\beta - n'_\beta\right|}{n_\beta + n'_\beta}\right) \tag{2}$$

wherew_β is the weight of attribute β, n_β and n'_β are the number of attributes β in different knowledge graphs respectively. α is the weight coefficient, which could enhance the role of high weight attributes in alignment.

3 Preliminary Results

We used the dataset DBP15K, which were generated from DBpedia [Sun *et al.*, 2017].

The preliminary results are shown in Table 1. We evaluated the alignment results of the proposed method and compared with that of the recent related work [Wang *et al.*, 2018]. In Table 1, AE means only attribute information is used for embedding, and $SE + AE$ means both structure and attribute information are used for embedding (AE: attribute embedding; SE: structure embedding).

$Hits@k$ measures the proportion of correctly aligned entities ranked in the top k candidates. It can be seen that, our approach outperforms the baseline method with respect to the metrics of $Hits@1$, $Hits@10$ and $Hits@50$.

4 Future Work

Till the current stage, the approach of enhance attribute embedding pays attention to the importance of different attributes for GCN. And the result of performance improvement shows the weights of attributes play an important role in the training process. In the future, we hope to train the attributes of networks, which could select important attributes and weight attributes according to their importance automatically.

References

[Aditya *et al.*, 2018] Somak Aditya, Yezhou Yang, and Chitta Baral. Explicit reasoning over end-to-end neural architectures for visual question answering. In *AAAI*, pages 629–637, 2018.

[Fang *et al.*, 2017] Yuan Fang, Kingsley Kuan, Jie Lin, Cheston Tan, and Vijay Chandrasekhar. Object detection meets knowledge graphs. In *IJCAI*, pages 1661–1667, 2017.

[Sun *et al.*, 2017] Zequn Sun, Wei Hu, and Chengkai Li. Cross-lingual entity alignment via joint attribute-preserving embedding. In *International Semantic Web Conference*, pages 628–644. Springer, 2017.

[Wang *et al.*, 2018] Zhichun Wang, Qingsong Lv, Xiaohan Lan, and Yu Zhang. Cross-lingual knowledge graph alignment via graph convolutional networks. In *the Conference on Empirical Methods in Natural Language Processing*, pages 349–357, 2018.

[Wang *et al.*, 2019] Xiao Wang, Houye Ji, Chuan Shi, Bai Wang, Peng Cui, P. Yu, and Yanfang Ye. Heterogeneous graph attention network. *arXiv preprint arXiv:1903.07293*, 2019.