Neighborhood-Aware Attentional Representation for Multilingual Knowledge Graphs

Qiannan Zhu[1,2] , **Xiaofei Zhou***[1,2] , **Jia Wu**[3] , **Jianlong Tan**[1,2] and **Li Guo**[1,2]
[1]Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China
[2]School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China
[3]Department of Computing, Macquarie University, Sydney, Australia
{zhouxiaofei,zhuqiannan}@iie.ac.cn,Jia.Wu@mq.edu.au

Abstract

Multilingual knowledge graphs constructed by entity alignment are the indispensable resources for numerous AI-related applications. Most existing entity alignment methods only use the triplet-based knowledge to find the aligned entities across multilingual knowledge graphs, they usually ignore the neighborhood subgraph knowledge of entities that implies more richer alignment information for aligning entities. In this paper, we incorporate neighborhood subgraph-level information of entities, and propose a neighborhood-aware attentional representation method NAEA for multilingual knowledge graphs. NAEA devises an attention mechanism to learn neighbor-level representation by aggregating neighbors' representations with a weighted combination. The attention mechanism enables entities not only capture different impacts of their neighbors on themselves, but also attend over their neighbors' feature representations with different importance. We evaluate our model on two real-world datasets DBP15K and DWY100K, and the experimental results show that the proposed model NAEA significantly and consistently outperforms state-of-the-art entity alignment models.

1 Introduction

The multilingual knowledge graphs (KGs) like YAGO [Suchanek *et al.*, 2008] and DBpedia [Bizer *et al.*, 2009] increasingly play an significant role in supporting various knowledge-driven tasks. Those multilingual knowledge graphs consist of monolingual knowledge, in forms of directed graphs, where entities are represented as nodes and relations as edges. The monolingual knowledge are stored as triplets (e_h, r, e_t), representing that the head entity e_h and tail entity e_t are linked by relation r. Besides monolingual knowledge, the multilingual KGs also embody cross-lingual knowledge $(e_h, align(), e'_h)$ that matches the same real-world entities e_h and e'_h among different human languages L and L' by alignment operation $align()$, see Figure 1. A great deal of methods focus on exploiting monolingual knowledge

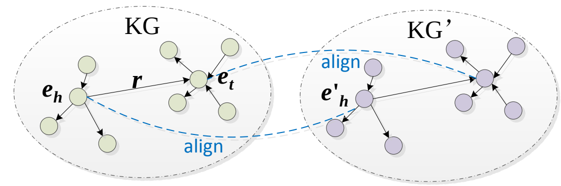

Figure 1: Multilingual Knowledge Graphs. KG and KG$'$ are the knowledge graphs of languages L and L'.

graphs in recent years. Particularly, the embedding-based methods that encode entities and relations into continue low-dimensional vector spaces, have achieved promising performance. Exemplarily, given a triplet (e_h, r, e_t), TransE [Bordes *et al.*, 2013] regards the relation embedding $\mathbf{r}$ as the translation vector between the head and tail entity embedding $\mathbf{e}_h$ and $\mathbf{e}_t$, and expects $\mathbf{e}_h + \mathbf{r} \approx \mathbf{e}_t$ when (e_h, r, e_t) holds. Other extended works such as TransH [Wang *et al.*, 2014], TransR [Lin *et al.*, 2015a] and TransD [Ji *et al.*, 2015] also emerged with different translation forms in characterizing relation r. However, a few methods have been done for modeling multilingual knowledge graphs.

Entity alignment is an effective way to integrate the multilingual KGs, which is the task of finding the same real-world entities in different KGs. The traditional multilingual entity alignment methods mainly based on machine translation, have low accuracy due to the poor performance in translation between multiple languages. Most recently, following above popular embedding-based models, MTransE [Chen *et al.*, 2017] provides the cross-lingual transitions for both entities and relations across different knowledge graph embeddings. IPTransE [Zhu *et al.*, 2017] jointly encodes both entities and relations of various KGs into an unified low-dimensional semantic space via sharing parameters on a seed set of aligned entities, JAPE [Sun *et al.*, 2017] further incorporates attribute triplets as additional information for learning KGs' embeddings in an unified space. BootEA [Sun *et al.*, 2018] adopts bootstrapping [Yarowsky, 1995; Abney, 2004] approach to iteratively label likely entity alignment as training data and leverage it for learning alignment-oriented embeddings. Existing entity alignment methods only use the triplet-based information, but ignore the inherent neighborhood information of entities for aligning entities.

*Corresponding Author

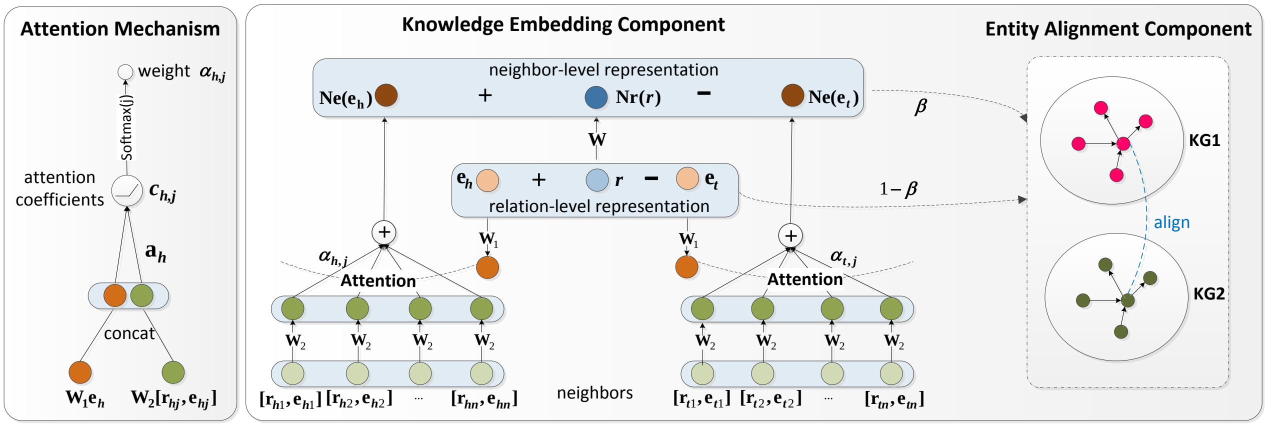

Figure 2: Simple Visualization of NAEA Model.

Compared with the triplet-based information, the neighbor-level information factually implies more richer alignment properties because arbitrary two aligned (equivalent) entities in different KGs usually contain equivalent neighborhood information.

This paper proposes a neighborhood-aware attentional representation method NAEA for multilingual knowledge graph. As Figure 2 shown, NAEA are composed of knowledge embedding component KE and entity alignment component EA. Both KE and EA components incorporate neighbor-level and relation-level information with different weights for learning KGs' embedding representation and aligning entities respectively. KE component devises an attention mechanism to obtain neighbor-level representation by aggregating neighbors with a weighted combination, and uses the triplet-based characteristic information to learn relation-level representation. EA component determines alignments between entities across multilingual KGs by measuring the similarity of their integrated representations, which are also fused from neighbor-level and relation-level representations with different weights.

Our contributions. We propose a neighborhood-aware attentional representation model NAEA for multilingual knowledge graphs, which can better capture the entity and relation features for targeting aligned entities. Two components, knowledge embedding component KE and entity alignment component EA, are included in NAEA.

- NAEA incorporates neighbor-level with relation-level feature information in knowledge graphs to perform multilingual entity alignment task.
- KE component devises a neighborhood-aware attention mechanism to learn neighbor-level representation by aggregating neighbors' embeddings with different impacts.
- EA component performs entity alignment by the similarity of entity integrated representations, which integrates neighbor-level and relation-level feature representation.

We evaluate our NAEA model on two real-world datasets DBP15K and DWY100K. Experimental results show that NAEA significantly achieves state-of-the-art performance on entity alignment task.

2 Problem Formulation

We describe an entity set as E, a relation set as R, a knowledge graph as $G = \{\tau | \tau = (e_h, r, e_t)\}$ with $e_h, e_t \in E, r \in R$. A KG pair specified by languages L_1 and L_2 is $G' = (G_{L_1}, G_{L_2})$, and its known aligned entities set is $A = \{(e_I, e_J) | e_I \in E_{L_1}, e_J \in E_{L_2}\}$, where E_{L_1} and E_{L_2} are respectively the entity set of G_{L_1} and G_{L_2}. Each aligned entity pair (e_I, e_J) represents an entity e_I in G_{L_1} has its synonymous counterpart e_J in G_{L_2} with language-specific surface names.

The entity alignment task in this paper is as follows: given the language-specific knowledge graph pair $G' = (G_{L_1}, G_{L_2})$, and the known alignment seeds $A = \{(e_I, e_J)\}$, entity alignment aims to automatically find and align more *unaligned* entities.

3 Model

This paper proposes a neighborhood-aware attentional representation method NAEA for multilingual knowledge graphs. As illustrated in Figure 2, NAEA consists of knowledge embedding component KE and entity alignment component EA: (1) KE in section 3.1 aims to learn alignment-oriented knowledge embedding representation, and (2) EA in section 3.2 aims to perform alignment in various KGs.

3.1 Neighborhood-aware Knowledge Embedding

KE fuses neighbor-level and relation-level information, and introduces an attention mechanism to learn neighbor-level representation of heterogeneous KGs.

Neighbor-level Attentional Representation

For obtaining neighbor-level entity representation, we introduce an attention mechanism to aggregate entities' neighbors with different importance. As shown in Figure 2, the attention mechanism takes as input an entity e_h with its neighborhood set $\{(r_{h1}, e_{h1}), (r_{h2}, e_{h2}), ..., (r_{hn}, e_{hn})\}$, and outputs its corresponding neighbor-level feature representation $\mathrm{Ne}(e_h)$. Here r_{hj} with $j = 1, ..., n$ is the relation that links from entity e_h to e_{hj} or vice versa, and n is the number of neighbors of entity e_h.

Specifically, for each entity-neighbor pair, we first use two shared linear transformations, parameterized by two weight matrices $\mathbf{W}_1 \in R^{m\times m}$ and $\mathbf{W}_2 \in R^{m\times 2m}$, to transform the input features into higher-level features, and then adopt *self-attention* on entities to compute attention coefficients between entity e_h and its neighborhood entity e_{hj} with relation r_{hj}, i.e.,

$$c_{h,j} = \mathbf{a}_h^T[\mathbf{W}_1\mathbf{e}_h, \mathbf{W}_2[\mathbf{r}_{hj}, \mathbf{e}_{hj}]]$$

where $c_{h,j}$ represents the importance of entity e_{hj}'s features to entity e_h with relation r_{hj}, $\mathbf{a}_h \in R^{2m}$ is an entity-specific weight vector, $\mathbf{e}_h$,$\mathbf{e}_{hj}$,$\mathbf{r}_{hj} \in R^m$ are the vector embeddings of e_h,e_{hj},r_{hj}, and $[,]$ indicates the concatenation operation. m is the dimension of entity and relation embedding space.

Obviously the above self-attention mechanism allows to calculate the importance of arbitrary two linked entities. To make coefficients easily comparable across different entities, we normalize them across all choices of $j = 1, 2, ..., n$ using the softmax function:

$$\alpha_{h,j} = \text{softmax}(c_{h,j}) = \frac{\exp(\text{LeakyReLU}(c_{h,j}))}{\sum_{i=1}^{n}\exp(\text{LeakyReLU}(c_{h,i}))}$$

To obtain the final output neighbor-level feature representation $\text{Ne}(e_h)$ of input entity e_h, we compute a linear combination of its neighbors' high-level features with different weight coefficients $\alpha_{h,j}$, and adopt multi-head attention [Vaswani *et al.*, 2017] to stabilize the process of self-attention, i.e.,

$$\text{Ne}(e_h) = \phi(\frac{1}{K}\sum_{k=1}^{K}\sum_{j=1}^{n}\alpha_{h,j}^k\mathbf{W}_2^k[\mathbf{r}_{hj}, \mathbf{e}_{hj}])$$

where $\phi = sigmoid$ is the activation function, $\alpha_{h,j}^k$ are the normalized attention coefficients computed by the k-th attention mechanism, and K is the number of head. Although the multi-head attention mechanism expands the parameter requirements by a factor of K, the individual heads' computations are fully independent and can be parallelized.

We mainly perform such attention mechanism to get neighbor-level entity representations. For the sake of fairness, we use $\text{Nr}(r) = \mathbf{Wr}$ to represent the neighbor-level relation representation of relation r with $\mathbf{W} \in R^{m\times m}$.

Relation-level Representation

Besides utilizing the neighbor-level information, we also consider the triplet-based information for our model NAEA. In order to fully preserve the inherent structural property of triplets, we use $(\mathbf{e}_h, \mathbf{r}, \mathbf{e}_t)$ to describe the relation-level embedding representation of a triplet (e_h, r, e_t).

Loss Function

For capturing neighbor-level information of knowledge graphs, similar to the widely-used embedding based model TransE [Bordes *et al.*, 2013], we define the neighbor-level score function $f_1(\tau) = \|\text{Ne}(e_h) + \text{Nr}(r) - \text{Ne}(e_t)\|_2^2$ for a triplet $\tau = (e_h, r, e_t)$. Different from the margin-based ranking loss function used in TransE, we adopt following limit-based scoring loss function [Zhou *et al.*, 2017] since it not only ensures the discrimination between the scores of positive and negative triplets, but also ensures the lower scores for positive triplets, i.e.,

$$\begin{aligned} L_1 = &\sum_{\tau\in T}\sum_{\tau'\in T'}[f_1(\tau) + \mu_1 - f_1(\tau')]_+ \\ &+ \gamma\sum_{\tau\in T}[f_1(\tau) - \mu_2]_+ \end{aligned} \tag{1}$$

where $[\cdot]_+$ means that $max(\cdot, 0)$, τ and τ' are respectively the positive and negative triplet from positive and negative triplet set T and T', $\gamma > 0$ is a balance hyper-parameter, and μ_1, μ_2 are two hyper-parameters with the constraint $\mu_1, \mu_2 > 0$.

For capturing relation-level information of knowledge graphs, we also define the relation-level score function $f_2(\tau) = \|\mathbf{e}_h + \mathbf{r} - \mathbf{e}_t\|_2^2$ for measuring the plausibility of the triplet $\tau = (e_h, r, e_t)$, and use the same loss framework as L_1 to define relation-level scoring loss function L_2.

We aim to learn embedding representations of KGs by fusing neighbor-level and relation-level information of knowledge graphs with different weights. Therefore, we propose following combined loss function for optimization:

$$L_E = \beta L_1 + (1-\beta)L_2 \tag{2}$$

where β is the weight hyper-parameter that balances the importance of neighbor-level of information.

3.2 Entity Alignment

Entity alignment aims to find $A' = \{(e_i, e_j) \in E_{L_1} \times E_{L_2} | e_i \equiv_r e_j\}$, where $e_i \equiv_r e_j$ means that an equivalence relation $\equiv_r$ holds between e_i and e_j. Usually the already known aligned pairs subset $A = \{(e_I, e_J)|e_I \in E_{L_1}, e_J \in E_{L_2}\}$ of A' is used as training data.

In this paper, we consider the alignment task as the classification problem that labels entity e_i in G_{L_1} using entity e_j in G_{L_2}. As said in [Lacoste-Julien *et al.*, 2013; Zhang *et al.*, 2015; Sun *et al.*, 2018], this entity alignment task has the one-to-one alignment constraint: an entity can be aligned with at most one label, and a label can be assigned to at most one entity. This constraint makes difference between our problem and the common classification problem [Liu *et al.*, 2019; Liu *et al.*, 2017]. We define the alignment probability of using entity e_j to label e_i as

$$\begin{aligned} \pi(e_j|e_i) = &\beta\ \phi(\text{sim}(\text{Ne}(e_i), \text{Ne}(e_j))) \\ &+ (1-\beta)\ \phi(\text{sim}(\mathbf{e}_i, \mathbf{e}_j)) \end{aligned}$$

where $\text{sim}(\cdot)$ is the cosine similarity. Based on neighbor-level and relation-level knowledge information, we minimize the following negative log-likelihood function to obtain the optimal parameters with the highest alignment likelihood:

$$L_A = -\sum_{e_i\in E_{L_1}}\sum_{e_j\in E_{L_2}}\mathbf{1}\ log\ \pi(e_j|e_i)$$

where $\mathbf{1}$ is the indicator function that equals to 1 when e_j is the true label of entity e_i, otherwise 0.

3.3 Training

To enable parameters of our model not only to capture the alignment likelihood information, but also model the inherent semantic knowledge of KGs. We minimize following joint objective function to perform model training:

$$L = L_E + \lambda L_A \tag{3}$$

where λ is the positive hyper-parameter that measures the importance of L_A.

In training stage, for regularization and avoiding overfitting, we apply dropout to the weight parameters $\mathbf{W}, \mathbf{W}_1, \mathbf{W}_2$ and L2 regularization to embeddings of KGs. For addressing the inadequate prior alignment, we use *bootstrapping* strategy used in [Sun *et al.*, 2018] that iteratively label likely alignment as training data to train our model.

Construct T and T'. For constructing the positive triplets set T, we deliberately exchange the entity with its counterpart in triplets to calibrate different KGs in the unified embedding space. Given an aligned entity pair $(e_I, e_J) \in A$, following positive triplets are generated:

$$\begin{aligned} T_a(e_I, e_J) = &\{(e_J, r, e_t)|(e_I, r, e_t) \in T_{L_1}\} \\ &\cup \{(e_h, r, e_J)|(e_h, r, e_I) \in T_{L_1}\} \\ &\cup \{(e_I, r, e_t)|(e_J, r, e_t) \in T_{L_2}\} \\ &\cup \{(e_h, r, e_I)|(e_h, r, e_J) \in T_{L_2}\} \end{aligned}$$

where T_{L_1} and T_{L_2} are respectively the positive triplets set of knowledge graphs G_{L_1} and G_{L_2}. Thus we have the general positive triplets set in Eq. (1) as $T = T_{L_1} \cup T_{L_2} \cup T_A$ with $T_A = \bigcup_{(e_I,e_J)\in A} T_a(e_I, e_J)$. For constructing the negative triplets set T', instead of the traditional uniform negative sampling in [Bordes *et al.*, 2013] that samples the replacer of entity from entire entity set, we adopt *Nearest Neighbor Sampling* strategy to limit the scope of sampling candidates. Specifically, for an entity e that will be replaced, we choose the top-s nearest neighbors of entity e in the embedding space as candidates. In this way, those entities that have low embedding similarities with entity e, are truncated and would not be sampled. Here, we use cosine similarity between *integrated* representation of entities to determine the top-s nearest neighbors of entities. The integrated representation of entity e is defined as $\beta\, \mathrm{Ne}(e) + (1-\beta)\mathbf{e}$.

4 Experiment

4.1 Experiment Setup

Datasets. We conduct experiments on two real-world datasets DBP15K and DWY100K. DBP15K [Sun *et al.*, 2017] is selected from the multilingual versions of DBpedia that includes entity alignment links from entities of English version to those in other languages. Three multilingual datasets $\mathrm{DBP}_{\text{ZH-EN}}$ (Chinese to English), $\mathrm{DBP}_{\text{JA-EN}}$ (Japanese to English) and $\mathrm{DBP}_{\text{FR-EN}}$ (French to English) are built in DBP15K. Each dataset contains 15 thousand reference alignment links with popular entities from English to Chinese, Japanese and French respectively. DWY100K [Sun *et al.*, 2018] is built from three large-scale multi-lingual knowledge graph DBpedia, Wikidata and YAGO3. Two large-scale datasets, DBP-WD and DBP-YG, are respectively extracted with 100 thousand reference alignment links from the English version of DBpedia to that of Wikidata and YAGO3. Table 1 illustrates the statistics of those data sets.

DataSets		#Ent	#Rel	#Att	#Rel.tri	#Att.tri
$\mathrm{DBP15K}_{\text{ZH-EN}}$	Chinese	66,469	2,830	8,113	153,929	379,684
	English	98,125	2,317	7,173	237,674	567,755
$\mathrm{DBP15K}_{\text{JA-EN}}$	Japanese	65,744	2,043	5,882	164,373	354,619
	English	95,680	2,096	6,066	233,319	497,230
$\mathrm{DBP15K}_{\text{FR-EN}}$	French	66,858	1,379	4,547	192,191	528,665
	English	105,889	2,209	6,422	278,590	576,543
DBP-WD	DBpedia	100,000	330	351	463,294	381,166
	Wikidata	100,000	220	729	448,774	789,815
DBP-YG	DBpedia	100,000	302	334	428,952	451,646
	YAGO3	100,000	31	23	502,563	118,376

Table 1: Statistics of the Datasets.

Parameter settings. In our model, we set the maximum number of neighbors n as 200, and select the dimension of entity(relation) embeddings m from $\{50, 75, 100, 150, 200\}$, the learning rate η from $\{0.001, 0.01, 0.1\}$, β from $\{0, 0.2, 0.4, 0.6, 0.8, 1\}$, λ from $\{0.1, 0.5, 1, 1.5, 2\}$, μ_1 from $\{0.5, 1, 2, 3, 4\}$, μ_2 from $\{0.01, 0.1, 0.5, 0.8, 1, 1.5, 2\}$, γ from $\{0.1, 0.5, 1, 1.5, 2, 2.5\}$, the number of head K from $\{1, 2, 4, 6, 8\}$. For our model, the best optimal parameter configurations are $m = 75$, $\beta = 0.8$, $\lambda = 1$, $\mu_1 = 1$, $\mu_2 = 0.1$, $\gamma = 2$, $K = 4$, $\eta = 0.01$. For each positive triplet, we select 10 negative triples for training, and set the training epochs as 1000. Following BootEA [Sun *et al.*, 2018], we used 30% of the gold standards as seed alignment while left the remaining as testing data, i.e., the latent aligned entities to discover.

4.2 Baselines

For comparative models, we select current four state-of-the-art methods as baselines in our experiments.

- MTransE [Chen *et al.*, 2017] provides the transformation for entities and relations in different language-specific knowledge graphs.
- IPTransE [Zhu *et al.*, 2017] incorporates relational paths and learns joint embeddings by sharing parameters based on iteratively aligning entities according to their semantic distance.
- JAPE [Sun *et al.*, 2017] is an attribute-preserving embedding model that incorporates the relation and attribute embeddings for entity alignment.
- BootEA [Sun *et al.*, 2018] adopts the *bootstrapping* process to label likely alignment as training data and edit alignment during iterations.

In this experiment, we directly copy the experiment results of baseline models reported in their papers since the same datasets are used.

4.3 Multilingual Entity Alignment

The objective of this task is to find the same semantic entities from different languages in knowledge graphs. Comparing with baseline models, we consider two measures as evaluation metrics: (1)Hits@k: the proportion of correct alignment ranked in top-k. Here Hits@1 and Hits@10 are adopted, (2) Mean Reciprocal Rank(MRR): the average of the reciprocal ranks of results. Higher Hits@k and MRR are expected, which indicates better alignment performance.

Model	DBP15K$_{ZH-EN}$			DBP15K$_{JA-EN}$			DBP15K$_{FR-EN}$			DBP-WD			DBP-YG		
	Hits@1	Hits@10	MRR	Hits@1	Hits@10	MRR	Hits@1	Hits@10	MRR	Hits@1	Hits@10	MRR	Hits@1	Hits@10	MRR
MTransE	30.83	61.41	0.364	27.86	57.45	0.349	24.41	55.55	0.335	28.12	51.95	0.363	25.15	49.29	0.334
IPTransE	40.59	73.47	0.516	36.69	69.26	0.474	33.30	68.54	0.451	34.85	63.84	0.447	29.74	55.76	0.386
JAPE	41.18	74.46	0.490	36.25	68.50	0.476	32.39	66.68	0.430	31.84	58.88	0.411	23.57	48.41	0.320
BootEA	62.94	84.75	0.703	62.23	85.39	0.701	65.30	87.44	0.731	74.79	89.84	0.801	76.10	89.44	0.808
NAEA	**65.01**	**86.73**	**0.720**	**64.14**	**87.27**	**0.718**	**67.32**	**89.43**	**0.752**	**76.70**	**91.79**	**0.817**	**77.86**	**91.25**	**0.821**

Table 2: Results on DBP15K and DWY100K.

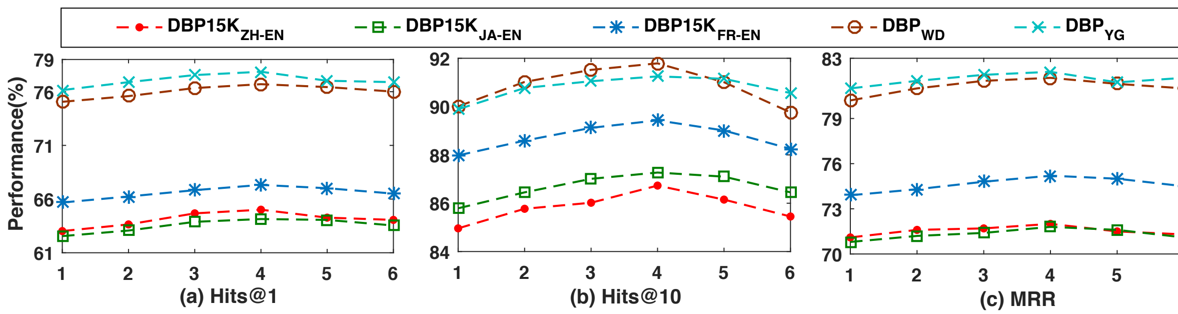

Figure 3: Performance of NAEA on Different Number of Head K.

Table 2 gives the convinced results of entity alignment on DBP15K and DWY100K. From Table 2, we can observe that our model NAEA consistently outperforms all baselines on five data sets. More specifically, NAEA achieves at least 1.76% on Hits@1, 1.81% on Hits@10 and 1.3% on MR-R higher performance than other models. We attribute the superiority of our model to its three advantages: (1) Our model devises knowledge embedding component for learning knowledge embeddings and entity alignment component for aligning entities ingeniously. (2) Our model incorporates the neighbor-level and relation-level information of KGs with different weights for multilingual entity alignment. (3) Our model introduces a multi-head attention mechanism to learn the neighbor-level representations by integrating neighbors' embeddings with a weighted combination.

MTransE [Chen *et al.*, 2017] obtains the worst alignment performance because it learns the embeddings of KGs in different vector spaces, and losses information when modeling the translation between different embedding spaces. IPTransE [Zhu *et al.*, 2017] and JAPE [Sun *et al.*, 2017] achieve better performance than MTransE due to relational path and entity attribute information of KGs are respectively leveraged. Expect for our model NAEA, BootEA has the highest results among baselines since it uses bootstrapping process to accurately label likely alignment as training data, and address the problem caused by the small proportion of prior alignment.

Link prediction in monolingual KG. Effective link prediction can help to improve the completeness of a knowledge graph, and further helps to improve the alignment performance. Link prediction task is to predict the missing head/tail entity given a triplet $(h, r, ?)/(?, r, t)$. Usually two evaluation metrics are used: (1) mean rank of correct entities (Mean Rank), (2) proportion of correct answers ranked in top-10 (Hit@10). In above settings, lower Mean and higher Hit@10 are expected. To testify the effectiveness of neighbor-level information for improving link prediction performance, we compare NAEA with TransE, BootEA on a monolingual KG with the representations obtained by only using individual embedding objective, such as Eq. (2). In this experiment, we randomly select monolingual triplets from DBP15K$_{ZH-EN}$ and DBP-WD to organize the training, valid and test set according to ratio $8:1:1$. The results of link prediction on datasets are shown in Table 3, which are that NAEA achieves best performance among baselines on all metrics since NAEA successfully alleviates the neighbor-level information of entities in knowledge graphs. It suggests that the neighbor-level information in KGs can provide indispensable characteristic information of knowledge graphs, and is indeed beneficial to enhance link prediction performance.

4.4 Discussion

In this section, we evaluate how different choices of parameters affect our model's performance. In the following experiments, except for the parameter being tested, the rest parameters are set as the optimal configurations.

Performance on different K. The above experimental results show that the usage of neighbor-level information through the multi-head attention mechanism is beneficial for improving alignment performance. Thus this subsection mainly explores how the performance of our model changes with different number of head K. In this experiment, we test our model with selecting K from $\{1, 2, 3, 4, 5, 6\}$. Figure 3 gives the convinced results. From this figure, we can see that:(1) NAEA has the best performance when $K = 4$, indicating that $K = 4$ setting best expresses the neighbor-level information of entities, and delivers alignment characteristics of entities in multilingual knowledge graphs. (2) The performance on all datasets begins to gradually rise to the highest

Model	DBP15K$_{ZH-EN}$				DBP-WD			
	ZH		EN		DBpedia		Wiki	
	Mean	Hits@10	Mean	Hits@10	Mean	Hits@10	Mean	Hits@10
TransE	471	57.3	402	60.9	360	62.7	298	65.8
BootEA	286	72.4	234	75.2	221	76.4	191	78.9
NAEA	**209**	**77.2**	**183**	**79.3**	**165**	**80.6**	**152**	**81.3**

Table 3: Results of Link Prediction.

point and then declines as the number of head K grows. This mainly because that too small K can not capture the richer neighbor-level information, and too large K may introduce noisy and lead to over-fitting problem.

Performance on different β. In our model, β is to weight the importance of neighbor-level information for multilingual entity alignment. To evaluate the influences of different β on alignment performance, we test our model with different β selected from $\{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7. \ 0.8, 0.9, 1\}$ on Hit@10 metric. We report the experimental results in Figure 4. From Figure 4, we find that:(1) Under $\beta = 0.8$ setting, NAEA can achieve the best alignment performance on all datasets, which indicates that such setting can sufficiently reflect the significance of neighbor-level information in aligning entities. (2) The performance of our model first is increasing with the growth of β and then drops with β further increases, and has the lowest accuracy when $\beta = 0$ or $\beta = 1$. It suggests that the neighbor-level and relation-level information in KGs are both necessary.

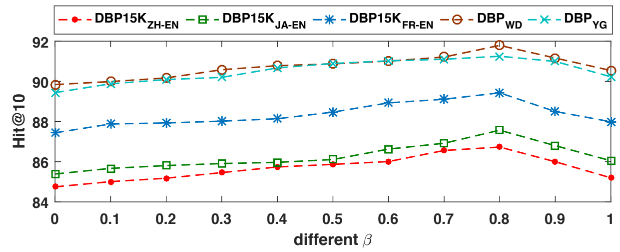

Figure 4: Performance of NAEA on Different β.

5 Related Work

5.1 Knowledge Representation

In recent years, a series of embedding-based methods such as [Bordes *et al.*, 2013; Wang *et al.*, 2014; Zhu *et al.*, 2019] have been quickly developed for knowledge representation learning. Those methods attempt to encode entities and relations into low-dimension vector spaces while preserving KGs' properties. Among those methods, the most widely used translation-based method TransE [Bordes *et al.*, 2013] projects entities and relations into a continuous low-dimensional vector space, and treats relation vector as the translation between head and tail entity vectors, i.e., expects $\mathbf{e}_h + \mathbf{r} \approx \mathbf{e}_t$ for a triplet (e_h, r, e_t). TransE is effective and promising for knowledge graph completion, while it has issues in modeling complex relations. Therefore, later works such as [Wang *et al.*, 2014; Lin *et al.*, 2015a; Ji *et al.*, 2015; Yang *et al.*, 2015] are proposed to address the issues of TransE. Also, there exists some non-translation based models, such as [Socher *et al.*, 2013; Nickel *et al.*, 2016; Shi and Weninger, 2017; Dettmers *et al.*, 2018], for learning knowledge graph embedding representations.

Besides above models that use one-step relational path to complete knowledge graphs, many models such as [García-Durán *et al.*, 2015; Lin *et al.*, 2015b] based on multi-step relational paths are also presented and achieve better performance in knowledge graph embedding.

5.2 Knowledge Alignment

Traditional knowledge alignment methods heavily rely on human efforts [Vrandecic, 2012] or well-designed hand-crafted features [Mahdisoltani *et al.*, 2015]. Though achieving better alignment performance, they are time-consuming, labor-expensive and usually suffer from inflexible extension.

Automated knowledge alignment mainly leverages various heterogeneous information in different KGs for knowledge alignment. Some models such as [Wang *et al.*, 2013; Suchanek *et al.*, 2011] make use of external lexicons or Wikipedia links to address the heterogeneity between different KGs. Differing from these models, many embedding-based models only using internal triplet-based information, were proposed for entity alignment. MTransE [Chen *et al.*, 2017] utilizes TransE to embed language-specific KGs into separate embedding spaces, and learns the transition between different spaces. IPTransE [Zhu *et al.*, 2017] uses P-TransE [Lin *et al.*, 2015b] and parameter sharing module to jointly encode entities and relations into a unified semantic spaces. JAPE [Sun *et al.*, 2017] employs semantic structures and attribute correlations of KBs to embed entity and relation embedding representations into an unified embedding spaces. BootEA [Sun *et al.*, 2018] leverages bootstrapping idea to label likely alignment as training data and edit alignment during iterations, which is state-of-the-art entity alignment model.

6 Conclusion

This paper presents a neighborhood-aware attentional representation method NAEA for multilingual knowledge graphs, which incorporates the neighbor-level and relation-level information of KGs. NAEA devises knowledge embedding component KE for learning knowledge embeddings and entity alignment component EA for aligning entities. KE introduces an attention mechanism to obtain neighbor-level representation by assigning different importance to entities' neighbors. EA discovers aligned entities based on the integration of their neighbor-level and relation-level representation with different weights. We empirically conduct experiments on multilingual entity alignment task and monolingual link prediction task with two data sets DBP15K and DB100K. The experimental results show that NAEA significantly and consistently achieves state-of-the-art performance.

Acknowledgments

This work is supported by National Key R&D Program No.2017YFB0803003, the National Natural Science Foundation of China (No.61202226), MQEPS (No 96804590) and MQRSG (No. 95109718). We thank all anonymous reviewers for their constructive comments.

References

[Abney, 2004] Steven P. Abney. Understanding the yarowsky algorithm. *Computational Linguistics*, 30(3):365–395, 2004.

[Bizer *et al.*, 2009] Christian Bizer, Jens Lehmann, Georgi Kobilarov, Sören Auer, Christian Becker, Richard Cyganiak, and Sebastian Hellmann. Dbpedia A crystallization point for the web of data. *J. Web Sem.*, 7(3):154–165, 2009.

[Bordes *et al.*, 2013] A. Bordes, N. Usunier, and A. Garcia-Duran. Translating embeddings for modeling multi-relational data. In *Proceedinds of NIPS*, pages 2787–2795, 2013.

[Chen *et al.*, 2017] Muhao Chen, Yingtao Tian, Mohan Yang, and Carlo Zaniolo. Multilingual knowledge graph embeddings for cross-lingual knowledge alignment. In *Proceedinds of IJCAI*, pages 1511–1517, 2017.

[Dettmers *et al.*, 2018] Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. Convolutional 2d knowledge graph embeddings. In *AAAI*, pages 1811–1818, 2018.

[García-Durán *et al.*, 2015] Alberto García-Durán, Antoine Bordes, and Nicolas Usunier. Composing relationships with translations. In *EMNLP*, pages 286–290, 2015.

[Ji *et al.*, 2015] G. Ji, S. He, L. Xu, K. Liu, and J. Zhao. Knowledge graph embedding via dynamic mapping matrix. In *Proceedinds of ACL*, pages 687–696, 2015.

[Lacoste-Julien *et al.*, 2013] Simon Lacoste-Julien, Konstantina Palla, Alex Davies, Gjergji Kasneci, Thore Graepel, and Zoubin Ghahramani. Sigma: simple greedy matching for aligning large knowledge bases. In *KDD*, pages 572–580, 2013.

[Lin *et al.*, 2015a] Y. Lin, Z. Liu, M. Sun, Y. Liu, and X. Zhu. Learning entity and relation embeddings for knowledge graph completion. In *Proceedinds of AAAI*, pages 2181–2187, 2015.

[Lin *et al.*, 2015b] Yankai Lin, Zhiyuan Liu, Huan-Bo Luan, Maosong Sun, Siwei Rao, and Song Liu. Modeling relation paths for representation learning of knowledge bases. In *EMNLP*, pages 705–714, 2015.

[Liu *et al.*, 2017] Weiwei Liu, Ivor W. Tsang, and Klaus-Robert Müller. An easy-to-hard learning paradigm for multiple classes and multiple labels. *Journal of Machine Learning Research*, 18:94:1–94:38, 2017.

[Liu *et al.*, 2019] Weiwei Liu, Donna Xu, Ivor W. Tsang, and Wenjie Zhang. Metric learning for multi-output tasks. *IEEE TPAMI*, 41(2):408–422, 2019.

[Mahdisoltani *et al.*, 2015] Farzaneh Mahdisoltani, Joanna Biega, and Fabian Suchanek. Yago3: A knowledge base from multilingual wikipedias. In *CIDR*, 2015.

[Nickel *et al.*, 2016] Maximilian Nickel, Lorenzo Rosasco, and Tomaso Poggio. Holographic embeddings of knowledge graphs. *Computer Science*, 2016.

[Shi and Weninger, 2017] B. Shi and T. Weninger. Proje: Embedding projection for knowledge graph completion. In *AAAI*, pages 1236–1242, 2017.

[Socher *et al.*, 2013] Richard Socher, Danqi Chen, Christopher D. Manning, and Andrew Y. Ng. Reasoning with neural tensor networks for knowledge base completion. In *NIPS*, pages 926–934, 2013.

[Suchanek *et al.*, 2008] Fabian M. Suchanek, Gjergji Kasneci, and Gerhard Weikum. YAGO: A large ontology from wikipedia and wordnet. *J. Web Sem.*, 6(3):203–217, 2008.

[Suchanek *et al.*, 2011] Fabian M. Suchanek, Serge Abiteboul, and Pierre Senellart. PARIS: probabilistic alignment of relations, instances, and schema. *CoRR*, abs/1111.7164, 2011.

[Sun *et al.*, 2017] Zequn Sun, Wei Hu, and Chengkai Li. Cross-lingual entity alignment via joint attribute-preserving embedding. In *Proceedinds of ISWC*, pages 628–644, 2017.

[Sun *et al.*, 2018] Zequn Sun, Wei Hu, Qingheng Zhang, and Yuzhong Qu. Bootstrapping entity alignment with knowledge graph embedding. In *IJCAI*, pages 4396–4402, 2018.

[Vaswani *et al.*, 2017] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *NIPS*, pages 6000–6010, 2017.

[Vrandecic, 2012] Denny Vrandecic. Wikidata: A new platform for collaborative data collection. In *WWW*, 2012.

[Wang *et al.*, 2013] Zhichun Wang, Juanzi Li, and Jie Tang. Boosting cross-lingual knowledge linking via concept annotation. In *IJCAI*, pages 2733–2739, 2013.

[Wang *et al.*, 2014] Z. Wang, J. Zhang, J. Feng, and Z. Chen. Knowledge graph embedding by translating on hyperplanes. In *Proceedinds of AAAI*, pages 1112–1119, 2014.

[Yang *et al.*, 2015] B. Yang, W. t. Yih, X. He, J. Gao, and L. Deng. Learning multi-relational semantics using neural-embedding models. In *ICLR*, 2015.

[Yarowsky, 1995] David Yarowsky. Unsupervised word sense disambiguation rivaling supervised methods. In *A-CL*, pages 189–196, 1995.

[Zhang *et al.*, 2015] Duo Zhang, Benjamin I. P. Rubinstein, and Jim Gemmell. Principled graph matching algorithms for integrating multiple data sources. *IEEE TKDE.*, 27(10):2784–2796, 2015.

[Zhou *et al.*, 2017] Xiaofei Zhou, Qiannan Zhu, Ping Liu, and Li Guo. Learning knowledge embeddings by combining limit-based scoring loss. In *CIKM*, pages 1009–1018, 2017.

[Zhu *et al.*, 2017] Hao Zhu, Ruobing Xie, Zhiyuan Liu, and Maosong Sun. Iterative entity alignment via joint knowledge embeddings. In *IJCAI*, pages 4258–4264, 2017.

[Zhu *et al.*, 2019] Qiannan Zhu, Xiaofei Zhou, Peng Zhang, and Yong Shi. A neural translating general hyperplane for knowledge graph embedding. *J. Comput. Science*, 30:108–117, 2019.