Modeling Multi-mapping Relations for Precise Cross-lingual Entity Alignment

Xiaofei Shi[1]**, Yanghua Xiao**[1,2,3]
[1]School of Computer Science, Fudan University, China
[2]CETC Big Data Research Institute Co., Ltd., Guizhou, China
[3]Big Data Application on lmproving Government Governance Capabilities National
Engineering Laboratory, Guizhou, China
{shixf18, shawyh}@fudan.edu.cn

Abstract

Entity alignment aims to find entities in different knowledge graphs (KGs) that refer to the same real-world object. An effective solution for cross-lingual entity alignment is crucial for many cross-lingual AI and NLP applications. Recently many embedding-based approaches were proposed for cross-lingual entity alignment. However, almost all of them are based on TransE or its variants, which have been demonstrated by many studies to be unsuitable for encoding multi-mapping relations such as 1-N, N-1 and N-N relations, thus these methods obtain low alignment precision. To solve this issue, we propose a new embedding-based framework. Through defining dot product-based functions over embeddings, our model can better capture the semantics of both 1-1 and multi-mapping relations. We calibrate embeddings of different KGs via a small set of pre-aligned seeds. We also propose a weighted negative sampling strategy to generate valuable negative samples during training and we regard prediction as a bidirectional problem in the end. Experimental results (especially with the metric *Hits@1*) on real-world multilingual datasets show that our approach significantly outperforms many other embedding-based approaches with state-of-the-art performance.

1 Introduction

Knowledge bases in the form of knowledge graphs (KGs) contain lots of facts of the real world. In recent years, KGs have been successfully explored to serve many AI and NLP applications such as question answering, information extraction, semantic search and recommender systems. Multilingual KGs such as BabelNet (Navigli and Ponzetto, 2012), YAGO3 (Mahdisoltani et al., 2013) and DBpedia (Lehmann et al., 2015) play essential roles in many cross-lingual applications. However, for multilingual KGs, each language-specific part is constructed by different parties with different data sources. Thus these language-specific KGs contain different but complementary facts. To support cross-lingual applications, it is essential for us to integrate these language-specific KGs into a unified KG. Fortunately, existing inter-lingual links (ILLs) that link equivalent entities or relations can help us to achieve this goal. However, as Chen et al. (2017) and Sun et al. (2017) pointed out, the existing cross-lingual alignments usually account for a small proportion of the total. Therefore, it is valuable to find more cross-lingual entity alignment pairs to bridge the language gap in multilingual KGs.

Entity alignment is the task of finding entity pairs in different KGs that refer to the same real-world object. An effective solution for entity alignment is vital for integrating multiple KGs. In this paper, we focus on cross-lingual entity alignment. Our goal is to find more cross-lingual entity alignments based on existing alignment seeds.

Various methods have been explored for cross-lingual entity alignment. Traditional approaches rely on machine translation or feature engineering to find cross-lingual matching pairs. The effectiveness of these methods heavily depends on the quality of translations and defined features. Recently, many embedding-based approaches were proposed for cross-lingual entity alignment. Embedding-based approaches embed entities and relations as low-dimensional vectors, such that the similarity between entities or relations can be calculated via their vectors. For example, MTransE (Chen et al., 2017) embeds entities and relations separately in each KG and then learns the cross-lingual transitions between different language-specific embedding spaces. JAPE (Sun et al., 2017) encodes both structures of KGs and types of attribute values. GCN-EA (Wang et al., 2018) employs GCNs as its embedding model to encode entities.

Though these methods have made great progress in cross-lingual entity alignment, there are still several key issues worth study. To the best of our knowledge, almost all of the existing methods encode knowledge in KGs based on the well-known TransE (Bordes et al., 2013) and its variants (Sun et al., 2017; Zhu et al., 2017). Given a relational triple $\tau = (h, r, t)$, where r represents a specific relation between the head entity h and the tail entity t, TransE and its variants expect $\vec{v_h} + \vec{v_r} - \vec{v_t} \approx \vec{0}$, where $\vec{v_h}$, $\vec{v_r}$ and $\vec{v_t}$ denote the embeddings for h, r and t, respectively. However, as Wang et al. (2014) and Ebisu and Ichise (2018) pointed out: 1) TransE and its variants neglect some mapping properties of relations; and 2) TransE and its variants force entity embeddings to be on a sphere in the embedding space, which conflicts with the expectation of TransE and warps embeddings obtained by TransE. Thus TransE and its variants don't do well in modeling multi-mapping relations, such as 1-N, N-1 and N-N relations.

As the majority of relations in real KGs are multi-mapping relations, previous entity alignment approaches obtain low accuracy, especially with the metric *Hits@1*. And the gap between alignment results with the metric *Hits@1* and *Hits@10* in these methods is large. This means previous methods can cluster similar entities together, however, they don't have enough ability to distinguish which is the true counterpart. Although following works such as TransH (Wang et al., 2014), TransR (Lin et al., 2015b) and TransD (Ji et al., 2015) have extended TransE on modeling multi-mapping relations, we experimentally found these methods also have low performances for cross-lingual entity alignment. And though the GCN-based approach avoids the flaw in TransE and its variants, GCN-EA (Wang et al., 2018) only encodes entities in KGs. It doesn't distinguish among relations and the absence of relation embeddings results in low alignment precision, too.

Besides, most previous works use uniform sampling to generate negative samples during training, which often generates too many easy examples that contribute little to learning embeddings. Although Boot-EA (Sun et al., 2018) proposes a truncated method to select hard negative samples based on the semantic similarity, it only considers hard examples (easy ones are truncated) that

only account for a small proportion of all examples, which can lead to overfitting and low accuracy in practice (since easy examples are also useful).

To address the above issues, we propose a new embedding-based framework to do cross-lingual entity alignment in this paper. Motivated by the success of multiplicative approaches (Yang et al., 2015; Trouillon et al., 2016) in knowledge representation learning, we propose a new embedding approach for multilingual KGs. The score function of our method is based on multiplication instead of subtraction. Unlike TransE and its variants giving a strong constraint to embeddings, dot products of embeddings in our method scale well and can naturally handle both 1-1 and multi-mapping relations. Besides, we propose a weighted sampling strategy to pay more attention to harder examples than easier ones during the process of negative samples generating. We summarize the main contributions of this paper as follows:

- We propose a new KG embedding framework to jointly embed entities and relations from different KGs into a unified embedding space using a small set of pre-aligned alignment pairs obtained from inter-lingual links.

- We discuss the problem that generating too many easy negative examples during training hurts alignment performance and then we propose a weighted sampling strategy for hard negative example generation.

- We regard the alignment prediction as a bidirectional selection problem, and we propose to combine ranking matrices from both directions to align entities.

- We evaluate the proposed approach on three real-world cross-lingual datasets. Experimental results (especially with the metric *Hits@1*) show that our approach significantly outperforms four representative embedding-based methods for cross-lingual entity alignment.

2 Related Work

2.1 KG Embedding

In the past few years, many KG embedding methods have been proposed for modeling the semantics of KGs. TransE (Bordes et al., 2013)

interprets a relation as the translation from its head entity to its tail entity and achieves great success in many AI-related tasks. Following TransE, TransH (Wang et al., 2014), TransR (Lin et al., 2015b) and TransD (Ji et al., 2015) were proposed to extend TransE on modeling multi-mapping relations. PTransE (Lin et al., 2015a) and RTransE (García-Durán et al., 2015) extend TransE on modeling multi-step relation paths. There are also non-translation-based methods. ComplEx (Trouillon et al., 2016) learns embeddings via defining product-based functions over embeddings. And some methods utilizing extra resources, i.e. visual information (Xie et al., 2017) and entity descriptions (Xie et al., 2016; Xiao et al., 2017), were also proposed to improve embedding performance.

2.2 Cross-lingual Entity Alignment

Conventional methods for cross-lingual entity alignment rely on machine translation and hand-crafted features, which often need extra resources as input and are difficult to reuse. By contrast, embedding-based approaches are simple and easy to reuse. In recent years, many embedding-based methods have been proposed and achieved promising results in cross-lingual entity alignment.

Due to the simplicity and effectiveness of TransE, almost all of the existing cross-lingual entity alignment works employ TransE or its variants to encode knowledge. JE (Hao et al., 2016) employs a modified TransE to encode entities and relations. It bridges different KGs by adding a loss of alignment seeds in its objective function. MTransE (Chen et al., 2017) encodes entities and relations for each language-specific KG separately based on TransE, and then employs five strategies to learn cross-lingual transitions through pre-aligned triples. JAPE (Sun et al., 2017) learns structure embedding using a variant of TransE which adds weight to negative samples in its objective function. It also borrows the idea from Skip-gram model to learn attribute correlations and then leverages attribute correlations to refine structure embedding. To overcome the heterogeneity and language gap between different KGs, JAPE only considers the range types of attribute values, with the specific attribute values discarded.

There are also some efforts have been devoted to solving the problem of lacking enough training data. IPTransE (Zhu et al., 2017) and Boot-

EA (Sun et al., 2018) propose to expand training set from alignment results iteratively. KD-CoE (Chen et al., 2018) performs co-training of a KG embedding model and an entity description embedding model.

Besides, GCN-EA (Wang et al., 2018) employs GCNs to embed entities and attributes as low-dimensional vectors.

Different from the above methods, in this paper, we propose a new approach to jointly learn embeddings for both entities and relations from different KGs.

3 Problem Formulation

We begin with the problem definition. A knowledge graph (KG) usually contains a large number of triples in the form of (h, r, t), where r represents a relation between the head entity h and the tail entity t. Formally, a KG can be represented as $KG = (E, R, T)$, where E, R, T are the set of entities, relations and triples respectively.

Now suppose we have two heterogenous and language-specific KGs to be aligned, where $KG_1 = (E_1, R_1, T_1)$ and $KG_2 = (E_2, R_2, T_2)$. And in many cases, we already have a set of pre-aligned entities $S_e = \{(e_1, e_2) | e_1 \in E_1, e_2 \in E_2\}$ and relations $S_r = \{(r_1, r_2) | r_1 \in R_1, r_2 \in R_2\}$ as alignment seeds. The task of entity alignment is to automatically find new entity alignments based on the existing alignment seeds.

4 The Proposed Approach

Our approach consists of two steps. The first step is to learn embeddings for entities and relations, and we will introduce a new method to do this task. The second step is to align entities between the two input KGs based on the learned embeddings. In this section, we will introduce the two parts in details.

4.1 Multilingual KG Embedding

DistMA: The goal of KG embedding is to embed entities and relations into a unified vector space such that low-dimensional vectors of entities or relations can represent their semantics. A powerful KG embedding approach should embed similar entities or relations to have similar embeddings, such that the semantics of knowledge graphs can be well captured.

The most widely used approaches to encode entities and relations by previous works are

TransE (Bordes et al., 2013) and its variants. Given one triple $\tau = (h, r, t)$ in a knowledge graph, TransE and its variants define the energy function for this triple as $E(\tau) = \|\vec{v_h} + \vec{v_r} - \vec{v_t}\|_{1/2}$, where $\vec{v_h}, \vec{v_r}, \vec{v_t} \in \mathbb{R}^d$ represent the d-length vectors for h, r and t respectively. Then they optimize margin-based ranking losses so that the energy for positive triples is lower than negative ones. As discussed in Section 1, they don't do well in modeling multi-mapping relations, we need to explore new approaches to better capture the semantics of multi-mapping relations.

In this paper, we propose a new approach to encode both entities and relations in knowledge graphs. Our approach defines the energy function for a triple τ as:

$$E_1(h, r, t) = \langle \vec{v_h}, \vec{v_r} \rangle + \langle \vec{v_r}, \vec{v_t} \rangle + \langle \vec{v_h}, \vec{v_t} \rangle \quad (1)$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product.

Besides, instead of using a margin-based loss function, we employ the logistic loss with L_2 regularization on model parameters Θ. The objective function is defined as:

$$O_1 = - \sum_{(h,r,t) \in \mathbb{T}^+} \log \sigma(E_1(h, r, t))$$
$$- \sum_{(h',r,t') \in \mathbb{T}^-} \log \sigma(-1 \cdot E_1(h', r, t')) + \lambda \|\Theta\|_2^2 \quad (2)$$

where σ is the sigmoid function. $\mathbb{T}^+$ and $\mathbb{T}^-$ are positive and negative triples set, respectively. λ is a hyper-parameter that controls the influence of regularization. The value of $\sigma(E(h, r, t))$ measures the plausibility of the triple (h, r, t). This means our method expects positive triples to have high energy scores and negative ones to have low scores. Compared with TransE and its variants, dot products of embeddings scale well and can naturally handle both 1-1 and multi-mapping relations, which is helpful for capturing the semantics of KGs. We define the negative triples set $\mathbb{T}^-$ as:

$$\mathbb{T}^- = \{(h', r, t) | h' \in E\} \cup$$
$$\{(h, r, t') | t' \in E\}, (h, r, t) \in \mathbb{T}^+ \quad (3)$$

where E denotes the set of entities.

ComplEx: The energy function of DistMA doesn't distinguish between head entities and tail entities. This means DistMA can model symmetric relations like $similar\_to$, has_friend well but maybe unfriendly to some antisymmetric relations

like is_father_of, $part\_of$. To improve the ability of our model to encode such relations, we utilize the ComplEx (Trouillon et al., 2016). Instead of real-valued vectors, ComplEx gives complex-valued vectors to entities and relations. That is, for each entity e or relation r, let $\vec{w_e} \in \mathbb{C}^k$ or $\vec{w_r} \in \mathbb{C}^k$ be the corresponding complex-valued vector, i.e. $\vec{w_e} = Re(\vec{w_e}) + iIm(\vec{w_e})$ with $Re(\vec{w_e}) \in \mathbb{R}^k$ and $Im(\vec{w_e}) \in \mathbb{R}^k$ being the real and imaginary parts of $\vec{w_e}$, where $\mathbb{R}$, $\mathbb{C}$ represent the real and complex vector spaces respectively, and i denotes the square root of -1. Then the energy for a triple (h, r, t) is defined as follows:

$$E_2(h, r, t) = Re(\langle \vec{w_h}, \vec{w_r}, \vec{\bar{w}_t} \rangle) \quad (4)$$

where $\langle \cdot, \cdot, \cdot \rangle$ denotes the generalized inner product and $\vec{\bar{w}_t}$ represents the conjugate of $\vec{w_t}$: $\vec{\bar{w}_t} = Re(\vec{w_t}) - iIm(\vec{w_t})$. One can easily verify that Equation (4) can be expanded and written as:

$$\begin{aligned} E_2(h, r, t) =& \langle Re(\vec{w_h}), Re(\vec{w_r}), Re(\vec{w_t}) \rangle \\ &+ \langle Re(\vec{w_h}), Im(\vec{w_r}), Im(\vec{w_t}) \rangle \\ &+ \langle Im(\vec{w_h}), Re(\vec{w_r}), Im(\vec{w_t}) \rangle \\ &- \langle Im(\vec{w_h}), Im(\vec{w_r}), Re(\vec{w_t}) \rangle \end{aligned} \quad (5)$$

With this energy function, we use the same Equation (2) to calculate loss O_2 as O_1.

Joint Embedding: With ComplEx, we refine the final energy function for each triple as:

$$E(h, r, t) = E_1(h, r, t) + E_2(h, r, t) \quad (6)$$

Then the final loss O to optimize can be defined by Equation (2) using Equation (6) as the energy function.

We will evaluate each part of our method in the experiments.

Parameter Sharing: The aforementioned embedding model is trained separately in each KG. As our final goal is to align entities between two KGs, we need to calibrate the embeddings of the two KGs into a unified embedding space based on the existing seed alignment. A natural and effective assumption is: the pre-aligned pair should share embedding to bridge different KGs. Under this assumption, the semantic loss between equivalent entity or relation pair is zero. This idea has also been mentioned in Sun et al. (2017) and Zhu et al. (2017).

4.2 Weighted Negative Sampling

Most previous works (Bordes et al., 2013; Sun et al., 2017; Zhu et al., 2017; Wang et al., 2018) use uniform sampling to generate negative triples $\mathbb{T}^-$, that is, replacing the head or tail entity of a positive triple with a random entity. Despite the simplicity of this method, it can be blind in many cases. For example, suppose we have a positive triple $(David\ Hilbert, nationality, Germany)$ and randomly replace the tail entity with an arbitrary entity. Maybe we can get one negative triple $(David\ Hilbert, nationality, Barack\ Obama)$. This is not a good choice because such a ridiculous triple contributes little to learning embeddings. Since such an obviously false triple is "too easy" and can easily obtain low plausibility during training. In contrast, a negative triple like $(David\ Hilbert, nationality, France)$ is more valuable, since $France$ has similar semantics with $Germany$, but $France$ can't replace $Germany$. Such a negative triple can help the model to learn the underlying semantics difference between entities rather than types solely.

We propose a weighted sampling strategy to generate valuable negative samples. Take the tail entity for example, given a positive triple (h, r, t), we replace t with t' obtained from the following probability distribution:

$$P((h, r, t')|(h, r, t)) = \frac{exp(sim(t, t'))}{\sum_{e \in E} exp(sim(t, e))} \quad (7)$$

where E denotes the entity set and $sim(\cdot, \cdot)$ calculates the similarity between two entities. Here we employ the cosine similarity measure. This means our sampling method is more likely to replace one entity with its similar entities. Since these entities have strong semantic correlations with the original entity but don't express the original entity, which helps the model learn to distinguish. Note that our method also generates a few easy examples, which are also useful for learning embeddings.

4.3 Bidirectional Alignment

We regard alignment prediction as a ranking problem. That is, suppose we need to align the entities in KG_1 to KG_2, for each entity e_i in KG_1, we rank each entity e_j in KG_2 based on the similarity between e_i and e_j. The similarity can be calculated via their embeddings, which we define by cosine similarity measure as follows:

$$sim(e_i, e_j) = \langle [\vec{v_{e_i}}]_1, [\vec{v_{e_j}}]_1 \rangle \quad (8)$$

where $[\cdot]_1$ denotes the normalized vector. Finally, we will get a similarity matrix $\mathbf{S_{12}}$ between the two KGs. Then we can get a ranking matrix $\mathbf{M_{12}}$ for KG_1 with $\mathbf{M_{12}}(i, j)$ denoting the index of e_j in the ranking list of e_i.

One thing that hurts the alignment precision is the difference of knowledge distribution between different KGs. Considering the following fact during the entity alignment process. Suppose we select the first ranked entity e_j in the ranking list of e_i to be the counterpart of e_i. It is reasonable because, from the view of e_i, e_j is the most similar entity in KG_2. However, this can be misleading sometimes. For example, from the view of e_j, e_i may not rank first in the ranking list of e_j, and even rank at dozens. In such a case, we don't think e_j is the desired candidate for e_i. This means, instead of single direction in previous works, that the alignment prediction process should consider both directions of KGs. We combine ranking matrices from both directions to accomplish this goal.

From both views of KG_1 and KG_2, we refine the ranking matrix $\mathbf{M_{12}}$ as:

$$\mathbf{M_{12}} = \mathbf{M_{12}} + \mathbf{M_{21}^T} \quad (9)$$

where $\mathbf{M_{21}^T}$ denotes the transpose of the ranking matrix of KG_2.

4.4 Implementation Details

We initialize all embeddings for both DistMA and ComplEx based on the uniform distribution. We set $d = 150$ for DistMA and $k = 75$ for ComplEx so that the length of embeddings in DistMA and ComplEx are all equal to 150. Before predicting, we concatenate the embeddings in DistMA and ComplEx such that each entity is represented as a 300-dimensional vector in the end. Besides, during training, we sample ten negative triples for each positive one. Since the weighted negative sampling is time-consuming, to speed it up, we calculate Equation (7) every five epochs. We use the self-adaptive optimization method Adam (Kingma and Ba, 2015) for all trainings and we implement our model with Tensorflow [1].

5 Evaluation

5.1 Datasets

We evaluate our proposed method on DBP15K datasets (Sun et al., 2017; Wang et al., 2018).

---

[1] https://www.tensorflow.org/

The datasets contain three cross-lingual real-world datasets: DBP$_{ZH-EN}$ (Chinese to English), DBP$_{FR-EN}$ (French to English), DBP$_{JA-EN}$ (Japanese to English). In which each was built with 15 thousand ILLs from multilingual versions of DBpedia. The average number of entities, relations and relational triples of the three datasets are 166,255/4,291/420,025 respectively.

5.2 Baseline Methods

We select four representative methods as baselines for cross-lingual entity alignment. These methods can be categorized as:

- Translation-based methods where relations are modeled as translation operators in the embedding space, including JE (Hao et al., 2016), MTransE (Chen et al., 2017) and JAPE (Sun et al., 2017). For MTransE, it develops five variants in its alignment model, where the fourth obtains the best results according to the experiments of its authors. Thus, we choose this variant to represent MTransE. For JAPE, it learns both structure and attribute embeddings. We report the results of its full model.

- To the best of our knowledge, GCN-EA (Wang et al., 2018) is the only one method employing non-translation-based embedding technique. GCN-EA uses GCNs as its encoding network to encode entities and attributes. We report the results of its full model.

5.3 Evaluation Metrics

By convention, we use *Hits@k* and *Mean Reciprocal Rank (MRR)* as our metrics. *Hits@k* measures the percentage of correctly aligned entities ranked in the top-k. Generally, *Hits@1* indicates precision. *MRR* is the average of the reciprocal ranks of all test instances. Higher *Hits@k* and *MRR* indicate better performance.

5.4 Model Selection

We select hyper-parameters with the following values: learning rate $\eta \in \{0.001, 0.003, 0.005, 0.01, 0.1\}$, L_2 regularization factor $\lambda \in \{1e-5, 1e-4, 5e-4, 1e-3\}$, batch size $B \in \{2000, 5000, 10000, 20000, 30000\}$. The optimal hyper-parameters are: $\eta = 0.005$, $\lambda = 1e-4$ and $B = 20000$. We use these values in all experiments.

The total number of epochs is set to 3000, within which we test alignment every ten epochs and save embeddings when obtaining better *Hits@1* performances.

5.5 Results and Analysis

Following Sun et al. (2017) and Wang et al. (2018), we use 30% of the gold standards for training and the rest 70% for testing. The results of four compared baselines and our approach are shown in Table 1. Among the four baselines, JAPE and GCN-EA obtain better results, since semantic loss happens when learning the translations between embedding spaces in the other two methods. GCN-EA is the most powerful method. It doesn't encode relations but it achieves a better performance except for the dataset DBP15K$_{ZH-EN}$.

However, we can easily see our method significantly outperforms all the four baselines, especially with the metric *Hist@1*, which indicates our method can better capture the semantics of different KGs. As discussed in Section 1, baseline methods all obtain low performance. And the *Hits@3* results of our method are comparable to the *Hits@10* results of all the baselines. Besides, we observe that the gap between the results with metrics *Hits@1* and *Hits@10* of our method is much smaller than the other methods (average 20.24% *vs.* 29.34%), which means our method has stronger ability to distinguish between target counterparts and error ones which have strong semantic similarity with target counterparts.

We also find the results after applying bidirectional alignment have a large improvement compared with single directional alignment under all the metrics, especially for *Hits@1*. This is due to the fact that entity alignment is essentially a bidirectional selection process. Bidirectional selection can soften the impact of unbalanced data distribution between different KGs, making the alignment prediction more reasonable.

Evaluation of Weighted Negative Sampling: Here we come to evaluate the effectiveness of our proposed weighted negative sampling strategy. We compare with uniform sampling with the negative sampling rate $neg = 1$ and $neg = 10$ (1 and 10 negative triples per positive triple), respectively. Table 2 shows the testing results with metrics *Hits@1* and *MRR*.

We summarize Table 2 in three aspects. Firstly, we can see that both sampling methods obtain

DBP15K$_{ZH-EN}$	ZH − EN				EN − ZH			
	Hits@1	Hits@3	Hits@10	Hits@50	Hits@1	Hits@3	Hits@10	Hits@50
*JE	21.27	-	42.77	56.74	19.52	-	39.36	53.25
*MTransE	30.83	-	61.41	79.12	24.78	-	52.42	70.45
*JAPE	41.18	-	**74.46**	**88.90**	**40.15**	-	**71.05**	**86.18**
†GCN-EA	**41.25**	-	74.38	86.23	36.49	-	69.94	82.45
Ours	65.46	78.13	86.11	91.56	64.73	77.56	85.79	91.20
Ours(*Bid.*)	**68.07**	**79.33**	**86.74**	**91.75**	**67.72**	**79.38**	**86.57**	**91.51**

DBP15K$_{JA-EN}$	JA − EN				EN − JA			
	Hits@1	Hits@3	Hits@10	Hits@50	Hits@1	Hits@3	Hits@10	Hits@50
*JE	18.92	-	39.97	54.24	17.80	-	38.44	52.48
*MTransE	27.86	-	57.45	75.94	23.72	-	49.92	67.93
*JAPE	36.25	-	68.50	85.35	38.37	-	67.27	82.65
†GCN-EA	**39.91**	-	**74.46**	**86.10**	**38.42**	-	**71.81**	**83.72**
Ours	62.84	75.94	84.90	91.35	62.27	75.33	84.71	91.32
Ours(*Bid.*)	**65.53**	**77.50**	**85.90**	**91.60**	**65.22**	**77.35**	**85.77**	**91.61**

DBP15K$_{FR-EN}$	FR − EN				EN − FR			
	Hits@1	Hits@3	Hits@10	Hits@50	Hits@1	Hits@3	Hits@10	Hits@50
*JE	15.38	-	38.84	56.50	14.61	-	37.25	54.01
*MTransE	24.41	-	55.55	74.41	21.26	-	50.60	69.93
*JAPE	32.39	-	66.68	83.19	32.97	-	65.91	82.38
†GCN-EA	**37.29**	-	**74.49**	**86.73**	**36.77**	-	**73.06**	**86.39**
Ours	64.30	78.76	87.92	93.90	63.50	78.39	87.83	93.94
Ours(*Bid.*)	**67.70**	**80.84**	**89.01**	**94.32**	**67.34**	**80.99**	**89.05**	**94.21**

Table 1: Results comparison. (* marks the results copied from Sun et al. (2017). † marks the results reported by its authors. - marks the unreported results. *Bid.* indicates bidirectional alignment.)

better performance with the increase of the negative sampling rate. This shows that more negative samples can assist model to better capture the semantics of KGs. Secondly, we can find that, compared with uniform negative sampling, our weighted sampling strategy gets better performance for all the three datasets. This is because our sampling method can generate more valuable negative samples to assist model to learn correct embeddings. Besides, we observe our method is more effective with a smaller negative sampling rate (neg = 1). Because when neg is large, it is likely to generate a few valuable negative samples though using uniform sampling. But when neg is small, the uniform sampling method can't generate enough valuable samples, which seriously hurts the final performance. However, our sampling strategy can pay more attention to valuable samples, generating many valuable negative samples even when the amount of total samples is small.

Ablation Study: For ablation study, we separate two variants from our approach. The first one only optimizes the objective function O_1, called DistMA. And the second one only optimizes O_2, called ComplEx. Table 3 shows the results of the two variants and the overall method.

As expected, Table 3 shows that each variant gets better results compared with four baselines. The DistMA can model symmetric relations well and the ComplEx can improve the ability in modeling antisymmetric relations. The overall method gets the best results due to its strong ability in modeling both relations.

Training with Different Sizes of Seed Alignment: We further investigate how the size of the training data affects the performance of our proposed approach. Here we use different proportions of seed alignment, which ranges from 20% to 50% with step 10%. And we choose two strong baselines JAPE and GCN-EA for comparison. Figure 1 illustrates the *Hits@1* results of the three approaches.

From Figure 1 we can see that the results for

Variants	$ZH-EN$		$EN-ZH$		$JA-EN$		$EN-JA$		$FR-EN$		$EN-FR$	
	Hits@1	MRR	Hits@1	MRR	Hits@1	MRR	Hits@1	MRR	Hits@1	MRR	Hits@1	MRR
neg=1, $unif.$	62.36	0.709	62.14	0.707	59.79	0.686	59.78	0.686	59.69	0.697	59.66	0.697
neg=1, $weig.$	**63.85**	**0.716**	**63.96**	**0.716**	**62.23**	**0.701**	**61.91**	**0.698**	**61.17**	**0.705**	**61.10**	**0.704**
neg=10, $unif.$	66.86	0.742	66.41	0.740	64.98	0.726	64.60	0.724	66.11	0.747	65.56	0.743
neg=10, $weig.$	**68.07**	**0.748**	**67.72**	**0.746**	**65.53**	**0.727**	**65.22**	**0.725**	**67.70**	**0.755**	**67.34**	**0.752**

Table 2: Evaluation of weighted negative sampling. (*unif.* indicates the uniform negative sampling and *weig.* indicates our weighted sampling strategy)

Variants	$ZH-EN$			$JA-EN$			$FR-EN$		
	Hits@3	Hits@10	MRR	Hits@3	Hits@10	MRR	Hits@3	Hits@10	MRR
DistMA	73.50	**83.66**	0.672	**71.09**	**82.09**	0.646	**72.57**	**83.97**	**0.656**
ComplEx	**73.82**	80.82	**0.692**	69.29	78.10	**0.653**	70.85	80.81	0.652
Overall	**79.33**	**86.74**	**0.748**	**77.50**	**85.90**	**0.727**	**80.84**	**89.01**	**0.755**

Table 3: Ablation study.

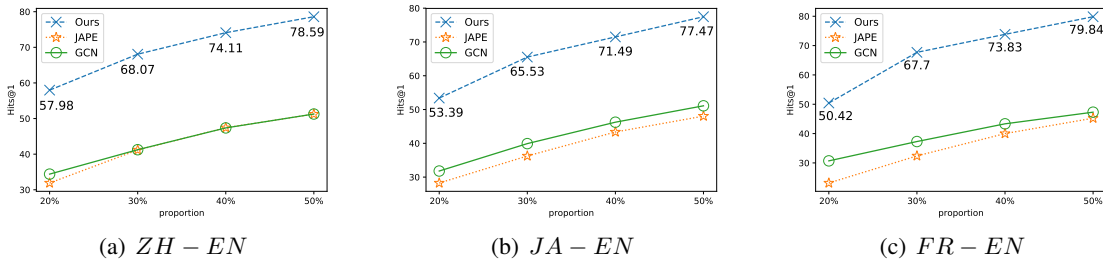(a) $ZH-EN$ (b) $JA-EN$ (c) $FR-EN$

Figure 1: *Hits@1* w.r.t. proportion of seed alignment.

all the three approaches become better with the increase of proportion of training data, because more seed alignment can provide more information to bridge different KGs. And we can also see that our approach consistently outperforms the other two methods under different proportions. Moreover, only with 20% seed alignment, our approach achieves comparable performances for all the three datasets with the other two methods using 50% seed alignment. When using 30% seed alignment, our approach largely outperforms the other two methods using 50% seed alignment. All these results show the robustness and effectiveness of our method.

Hits@1 Results by Mapping Properties of Relations: To better know how effective our method is for different relations, compared with JAPE (Sun et al., 2017), we evaluate the improvement of *Hits@1* results by mapping properties of relations. To do this, we first divide all relations into 1-N, N-1, 1-1 and N-N relations [2]. Then we

evaluate the *Hits@1* alignment results for head entities and tail entities of the four kinds of relations respectively.

The results are shown in Table 4. We can see that the result of N-N relations has the greatest improvement (with 111.4% average relative increase for head entities and 93.6% for tail entities). In addition, there are two points are also remarkable: the head entities of N-1 relations (96.2%) and the tail entities of 1-N relations (79.3%), which suggests the superiority of our method in modeling multi-mapping relations.

6 Conclusion and Future Work

This paper presents a simple and effective embedding framework for multilingual KGs which successfully improves the performance of cross-

---

[2]Following Wang et al. (2014), for each relation r, we

compute the average number of tails per head (tph_r) and the average number of head per tail (hpt_r). If $tph_r \geq 1.5$ and $hpt_r < 1.5$, r is treated as 1-N; if $tph_r < 1.5$ and $hpt_r \geq 1.5$, r is treated as N-1; if $tph_r < 1.5$ and $hpt_r < 1.5$, r is treated as 1-1; if $tph_r \geq 1.5$ and $hpt_r \geq 1.5$, r is treated as a N-N.

Position	Rel.	ZH − EN			JA − EN			FR − EN			Improv.(Avg.)
		JAPE	Ours	Improv.	JAPE	Ours	Improv.	JAPE	Ours	Improv.	
Head	1-N	49.25	69.47	41.1%	37.03	64.91	75.3%	39.03	75.82	94.3%	70.2%
	N-1	41.50	71.44	72.1%	32.57	63.32	94.4%	30.57	67.86	122.0%	96.2%
	1-1	46.77	70.67	51.1%	42.15	69.36	64.6%	36.19	72.72	100.9%	72.2%
	N-N	34.73	67.34	93.9%	31.03	62.58	101.7%	26.80	63.97	138.7%	**111.4%**
Tail	1-N	45.44	73.58	61.9%	36.19	66.81	84.6%	38.59	73.83	91.3%	79.3%
	N-1	59.68	79.84	33.8%	49.38	74.34	50.6%	46.85	75.65	61.5%	48.6%
	1-1	46.92	72.44	54.4%	47.40	72.77	53.5%	36.17	73.62	103.5%	70.5%
	N-N	41.42	69.03	66.6%	34.51	67.86	96.7%	31.48	68.49	117.5%	**93.6%**

Table 4: Comparison of JAPE and our model in *Hits@1* results by mapping properties of relations.

lingual entity alignment. We also propose a weighted sampling strategy to generate hard negative examples during training. Moreover, we propose to align entities from both directions. Experimental results on real-world datasets show that our method significantly outperforms four competitors, especially with the metric *Hits@1*.

Our proposed embedding framework can be easily applied to other tasks, i.e. link prediction and triple classification. And the idea of weighted negative sampling can be helpful for many AI and NLP tasks such as relation extraction and classification. For future work, we plan to explore more powerful KG embedding methods. And we also have the idea of using categorical attributes or hierarchical types to guide the negative sampling process.

Acknowledgments

References

Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *Advances in neural information processing systems*, pages 2787–2795.

Muhao Chen, Yingtao Tian, Kai-Wei Chang, Steven Skiena, and Carlo Zaniolo. 2018. Co-training embeddings of knowledge graphs and entity descriptions for cross-lingual entity alignment. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, IJCAI'18, pages 3998–4004. AAAI Press.

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

Takuma Ebisu and Ryutaro Ichise. 2018. Toruse: Knowledge graph embedding on a lie group. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

Alberto García-Durán, Antoine Bordes, and Nicolas Usunier. 2015. Composing relationships with translations. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 286–290.

Yanchao Hao, Yuanzhe Zhang, Shizhu He, Kang Liu, and Jun Zhao. 2016. A joint embedding method for entity alignment of knowledge bases. In *China Conference on Knowledge Graph and Semantic Computing*, pages 3–14. Springer.

Guoliang Ji, Shizhu He, Liheng Xu, Kang Liu, and Jun Zhao. 2015. 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)*, volume 1, pages 687–696.

Diederik P Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *The 3rd International Conference on Learning Representations*.

Jens Lehmann, Robert Isele, Max Jakob, Anja Jentzsch, Dimitris Kontokostas, Pablo N Mendes, Sebastian Hellmann, Mohamed Morsey, Patrick Van Kleef, Sören Auer, et al. 2015. Dbpedia–a large-scale, multilingual knowledge base extracted from wikipedia. *Semantic Web*, 6(2):167–195.

Yankai Lin, Zhiyuan Liu, Huanbo Luan, Maosong Sun, Siwei Rao, and Song Liu. 2015a. Modeling relation paths for representation learning of knowledge bases. In *Proceedings of EMNLP*, pages 705–714.

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

Farzaneh Mahdisoltani, Joanna Biega, and Fabian M Suchanek. 2013. Yago3: A knowledge base from multilingual wikipedias. In *CIDR*.

Roberto Navigli and Simone Paolo Ponzetto. 2012. Babelnet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, 193:217–250.

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

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

Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. 2016. Complex embeddings for simple link prediction. In *International Conference on Machine Learning*, pages 2071–2080.

Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. 2014. Knowledge graph embedding by translating on hyperplanes. In *Twenty-Eighth AAAI conference on artificial intelligence*.

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

Han Xiao, Minlie Huang, Lian Meng, and Xiaoyan Zhu. 2017. Ssp: semantic space projection for knowledge graph embedding with text descriptions. In *Thirty-First AAAI Conference on Artificial Intelligence*.

Ruobing Xie, Zhiyuan Liu, Jia Jia, Huanbo Luan, and Maosong Sun. 2016. Representation learning of knowledge graphs with entity descriptions. In *Thirtieth AAAI Conference on Artificial Intelligence*.

Ruobing Xie, Zhiyuan Liu, Huanbo Luan, and Maosong Sun. 2017. Image-embodied knowledge representation learning. In *IJCAI*.

Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. 2015. Embedding entities and relations for learning and inference in knowledge bases. In *International Conference on Learning Representations*.

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

822