

Uncertainty-aware Pseudo Label Refinery for Entity Alignment

Jia Li, Dandan Song*

School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China
{lijia92,sdd}@bit.edu.cn

ABSTRACT

Entity alignment (EA), which aims to discover equivalent entities in knowledge graphs (KGs), bridges heterogeneous sources of information and facilitates the integration of knowledge. Recently, based on translational models, EA has achieved impressive performance in utilizing graph structures or by adopting auxiliary information. However, existing entity alignment methods mainly rely on manually labeled entity alignment seeds, limiting their applicability in real scenarios. In this paper, a simple but effective Uncertainty-aware Pseudo Label Refinery (UPLR) framework is proposed without manually labeling requirement and is capable of learning high-quality entity embeddings from pseudo-labeled data sets containing noisy data. Our proposed model relies on two key factors: First, a non-sampling calibration strategy is provided that does not require artificially designed thresholds to reduce the influence of noise labels. Second, the entity alignment model achieves goal-oriented uncertainty correction through a gradual enhancement strategy. Experimental results on benchmark datasets demonstrate that our proposed model outperforms the existing supervised methods in cross-lingual knowledge graph tasks. Our source code is available at: <https://github.com/Jia-Li2/UPLR/>.

CCS CONCEPTS

• **Computing methodologies** → **Knowledge representation**; *Unsupervised learning*.

KEYWORDS

Knowledge Graph, Entity Alignment, Unsupervised

ACM Reference Format:

Jia Li, Dandan Song*. 2022. Uncertainty-aware Pseudo Label Refinery for Entity Alignment. In *Proceedings of the ACM Web Conference 2022 (WWW '22)*, April 25–29, 2022, Virtual Event, Lyon, France. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3485447.3511926>

1 INTRODUCTION

Knowledge graphs (KGs), which store facts as triples in the form of (subject entity, relationship, object entity) or (subject entity, attribute, text value), play an essential role in today's knowledge graphs machine learning. Many tasks [6, 7, 9] rely on the quality of

the knowledge graph to successfully conduct their operations. As knowledge graph was initially proposed to enhance the searching results of the search engines, it is widely used in Web applications such as Google search. However, with different KGs usually extracted from different data sources or contributed by people with diverse expertise, KGs are incomplete, severely hindering knowledge graphs' practicality.

In order to use the information obtained in various knowledge graphs, the knowledge graphs are expected to be integrated. The entity alignment task [37] is defined to identify entities that refer to the same object in the real world from two knowledge graphs, which plays a vital role in automatically integrating multiple knowledge graphs.

Existing methods usually aim to solve the alignment problem in a supervised manner through the artificially designed features [46] or the entity representation learned from the KG embedding method [8, 22, 24]. In addition, some semi-supervised methods [18, 38, 47] are also proposed to improve the performance of supervised entity alignment, such as names, attributes, and text descriptions with auxiliary information data to help embed learning. Supervised or semi-supervised alignment methods have significantly discovered semantically related entities.

However, they rely on clean labeled entity pairs as the training set, and creating clean pre-aligned labeled entity pairs (seeds) requires many resources. Significantly, the lack of label sets will reduce the performance of traditional entity alignment models when applied in the real world since the knowledge graph usually contains millions of entities, relationships, and triples. Unsatisfactory alignment results will be obtained if the model does not have a sufficient ratio of labeled entity pairs. For example, the reduced seed seriously affects the model's optimization of the geometric arrangement of the vector space of the two knowledge graphs. In DBP15K_{ZH-EN}, the model [29] Hits@1 is about 50% when the proportion of seed alignment is 50%, whereas the model Hits@1 is about 25% when the proportion of entity alignment seed of the model is reduced to 10%.

In the process of entity alignment, in order to get rid of the constraint of labeling entity (seed), the idea of using entity semantic information to construct cross-lingual entity pseudo-label is proposed to realize unsupervised entity alignment.

Due to two main challenges, its implementation is not easy:

- The newly proposed pseudo-label pairs inevitably contain noise pairs. Through experiments (SE of Table 3), it is found that the accuracy of the constructed pseudo-labels is only 58.6% in the DBP15K_{ZH-EN} dataset. It is not clear where the noise is, and it is expensive and laborious to distinguish whether the marked entity is noise or not. Therefore, the designed model should reduce the impact of noise without any supervision.

* Corresponding Author.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

WWW '22, April 25–29, 2022, Virtual Event, Lyon, France

© 2022 Association for Computing Machinery.

ACM ISBN 978-1-4503-9096-5/22/04...\$15.00

<https://doi.org/10.1145/3485447.3511926>

- An unified model for joint training is designed to reduce the impact of noise and cross-lingual entity alignment. On the one hand, the noisy pseudo-label data set is beneficial to solve the entity alignment problem in an unsupervised method. On the other hand, a high-precision entity alignment model is conducive to constructing a high-quality pseudo-label data set. However, the difficulty in designing the expected unified model lies in building a high-precision entity alignment model because it requires clean labeled entity pairs as training data, and noisy labels are also challenging to detect. In addition, even if the influence of noise has been reduced, it is hard to design feedback from reducing the influence of noise to entity alignment. Therefore, designing a joint training strategy is a challenging issue.

This paper proposes an Uncertainty-aware Pseudo Label Refinery (UPLR) framework that aligns entities without using any cross-graph alignment label set. UPLR achieves higher performance than state-of-the-art contrastive methods without using negative pairs. It iteratively bootstraps the outputs of a network to serve as targets for an enhanced representation. Moreover, UPLR is more robust to the choice of label augmentations than contrastive methods; It is suspected that not relying on negative pairs is one of the leading reasons for its improved robustness. Though previous methods based on bootstrapping have used pseudo-labels [14], cluster indices [4], or a handful of labels [1, 13, 34], directly bootstrapping the representations is proposed. Specifically, a set of pseudo-labels denoting potential cross-graph entity pairs is created based on semantic embeddings of entities. When extracting entity characteristics, the GAT instead of the graph convolutional network is used. It is an advantage that GAT can implicitly specify different weights to different nodes in a neighborhood, avoiding the possible effects of noise propagation on nodes in the graph. The GAT model further enhances the validity of the representative feature vectors of the entity. Then, attention scores help us capture and filter the importance of the remote nodes related to the central node. Finally, a gate graph attention network (GateGAT) is trained with pseudo-labeling seeds to update semantic embeddings of entities and generate high-quality aligned entities. This process is performed iteratively to gradually improve the pseudo-labeling accuracy and obtain the final entity alignment result. Experiments on three benchmark datasets demonstrate that the UPLR model in a completely unsupervised environment has achieved better results than supervised methods trained with manually labeled alignment seeds.

Figure 1 illustrates the difference between traditional methods and our proposed model. Top (Traditional Method): The traditional method to implement the entity alignment model is to perform negative sampling based on the manual entity alignment seed. Bottom (UPLR): Our proposed model to implement the entity alignment model is to construct a pseudo-labeled data set by using semantic embedding and directly input the unlabeled data set without negative sampling.

The contributions of this paper are listed as follows:

- A new Uncertainty-aware Pseudo Label Refinery (UPLR) framework is proposed, which adaptively mines confident samples from unlabeled data that will be labeled into trusted

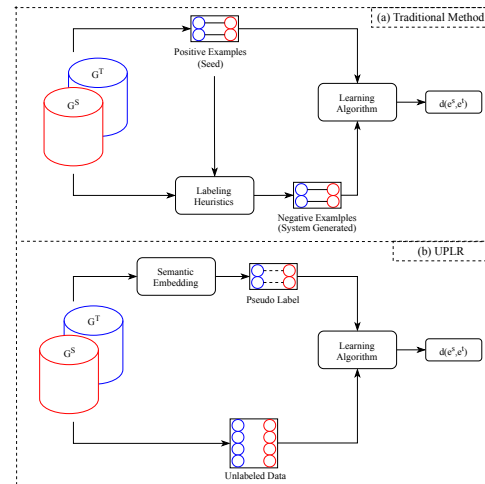


Figure 1: A comparison of different learning frameworks.

“positive” (pseudo label) classes. A hybrid loss is applied to both the augmented “labeled” examples and remaining unlabeled data for “supervision”.

- Considering the existence of errors in the pseudo labels, a novel non-sampling calibration strategy without any manually designed threshold is proposed to reduce the influence of noisy labels on entity representation learning.
- The drawback with pseudo-labels is that if the domains are not similar enough, it is not easy for us to obtain high-quality pseudo-labels since the labeling noise might be too high to hurt the performance. Therefore, a gradual enhancement strategy is proposed to enhance the similarity of the two domains.

2 RELATED WORK

2.1 EA based on translational models

The TransE-based [2] method represents the entity by modeling the triples involved, which provides fine-grained relational semantics of knowledge. Among the various translational methods, BootEA [31], and TransEdge [32] both utilize an iterative strategy to optimize model performance, which is a key technology to improve performance. NAEA [48] designs an attentional mechanism based on TransE to learn neighbor-level representations through weighted combinations of neighbor representations. KECG [15] combines the advantages of TransE and GCN, just like GCN, it convolves all the adjacent information of the entity while keeping the TransE translation between the head, relationship, and tail. PRASE [25] establishes corresponding filtering rules through the pre-training model. Furthermore, the adversarial learning of AKE [16] and SEA [23] based on TransE fails to achieve the desired effect, which is due to the high heterogeneity of KG that makes it difficult to convert from one KG to another through linear mapping functions similar to multilingual lexical space transformations.

2.2 EA based on graph structures

The GCN-based method GCN-Align [37] represents the entity by recursively aggregating the features of its neighbors, which utilizes the global KG structure. AliNet [33] shows the effectiveness of the multi-hop neighbors. HyperKA [28] maps the entity embedding obtained by GNN to the hyperbolic space for entity alignment. REA [24] has designed noise detection and noise perception to improve entity alignment performance. DGMC [8] employs synchronous message passing networks to iteratively re-rank the soft correspondences to reach a matching consensus in local neighborhoods between graphs. RREA [22] uses relational reflection transformation to obtain relation-specific embeddings for each entity in a more efficient way. DINGAL [45] expands the coupling distance between the parameter matrix in GCN and the topology of the underlying graph. Dual-AMN [19] not only models both intra-graph and cross-graph information, but also greatly reduces computational complexity. Although good results have been achieved, high-quality entity alignment seed are required. If the proportion of seeds is reduced, its performance will be greatly reduced.

2.3 EA with auxiliary information

In addition to graph structure information, some studies also add entity features based on auxiliary information into the process of entity encoding, including semantic embedding [39], attributes [18, 35, 38, 47], images [17] and description [46]. GM-Align [43], RDGCN [39] and HGCN [40] are models that combine semantic embedding of entity names. They distinguish different neighbors by initializing the entity representation. Therefore, their performance is better than the original GCN-align model [37]. AttrGNN [18], CEAFF [47] and EPEA [38] are models that combine entity attribute information. In the encoding process of the above methods, entity embedding is optimized by adding entity attribute information. However, additional information is not always available; thus, our proposed model solves entity alignment from a different perspective.

2.4 Model Learning from Pseudo-labelling

Pseudo-labelling has been mainly adopted by methods [14, 26, 27] based on deep learning. For unlabelled samples, the probability distribution of the model prediction is used as an indicator for pseudo-label the data [14]. By using the simple and efficient method, the system can easily add more data to help re-train the model. He and Sun. [11] proved that using a batch of samples with the highest prediction probability of the model can help enhance the performance of the model. For pseudo-label based methods, Lee et al. [14] and Xie et al. [42] generated pseudo labels for unlabeled data and then added them to train with labeled samples for enlarging data quantity jointly. In this regard, the quality of the pseudo label is vital. The difference among these methods is how to promote the accuracy of generated pseudo labels: Lee et al. [14] only utilized the high confidence pseudo labels by setting a threshold and avoided using the low-confidence label that may introduce the label noise; Xie et al. [42] refined the generated pseudo labels round by round with the introduced noisy student model, which is based on the assumption that the model would learn the essential and distinct features under the noisy input of each round and gradually

generate more precise pseudo labels. Our proposed model employs pseudo-labeling but inherits the spirit of label propagation. To avoid exhaustive labeling of unobserved samples, the graph structure information for selecting candidates that can be reliably labelled is exploited. Unlike previous studies, we propose a new loss function that can be used to efficiently train neural network models from pseudo-labeled and unlabeled data.

3 UPLR

Our proposed model builds a gate graph attention network based on the graph attention network and further develops a non-sampling calibration strategy combined with a gradual enhancement strategy to optimize the final knowledge graph embedding. Figure 2 shows the framework of the model, and our proposed model will be described in more detail below.

3.1 Problem Formulation

Formally, a KG is represented as $G = (E, R, T)$ and E, R, T are the sets of entities, relations and triples respectively. Let $G_s = (E_s, R_s, T_s)$ and $G_t = (E_t, R_t, T_t)$ be two cross-lingual knowledge graph to be aligned. That is, an entity in E_s may have its counterpart in E_t in different language or in different surface names. As a starting point, an equivalent cross-lingual knowledge graph entity pairs between E_s and E_t based on the semantic embedding as a pseudo-labels set $L = \{(e_s, e_t) | e_s \in E_s, e_t \in E_t\}$ is established, and an entity alignment task aims to automatically find more correct equivalent entities pair using the entity alignment seed (pseudo-labels) is defined.

3.2 Gate Graph Attention Networks

In this section, we adapt the non-local model of [41] to introduce gate to the Graph Attention Networks (GAT) [36] framework, enabling the entity alignment model to efficiently model relationships between widely separated spatial regions. We call the proposed model Gate Graph Attention Networks (GateGAT) because of its gate module [33] (see Figure 2).

GAT can be interpreted as performing the propagation of the attention message and updating on nodes. At each layer, the node features $e_i \in G$ ($G = (e_1, e_2, \dots, e_i, \dots, e_n)$) will be updated by its attention neighbor messages to $e'_i \in G$ as follows:

$$e'_i = \sigma \left(\sum_{j \in \{i\} \cup \mathcal{N}_i} \alpha_{ij} W e_j \right) \quad (1)$$

where σ is a non-linear activation, \mathcal{N}_i represents the one-hop neighbor set of node s , and W is a trainable parameter. The attention coefficient $\alpha_{i,j}$ is computed as follows, with parameters of a weight vector W . Moreover, to further explore the translation information for relations based on triples, and inspired by TransE [2], our designs a relational translation matrix $M'_{i,j}$ as follows:

$$\alpha_{i,j} = \text{softmax}_j (e_{i,j}) = \frac{\exp(e_{i,j})}{\sum_{k \in \mathcal{N}_s} \exp(e_{ik})} \quad (2)$$

where

$$e_{i,j} = \text{LeakyRelu} \left(M'_{i,j} W^T [W e_i \| W e_j] \right), \quad (3)$$

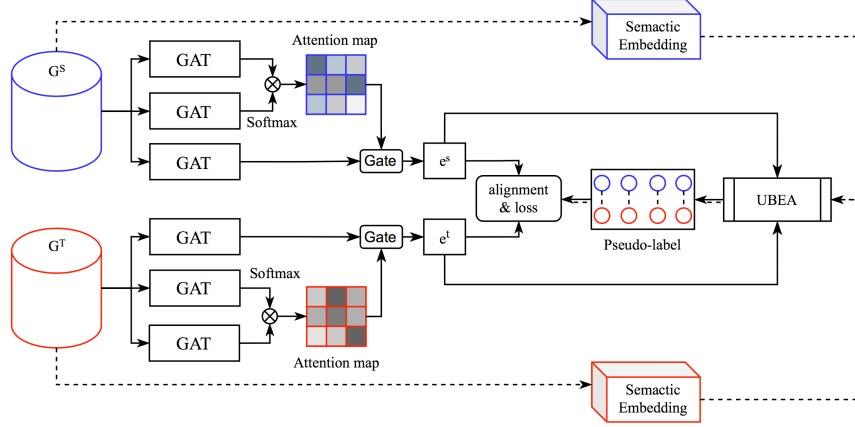


Figure 2: The overall framework of the UPLR. The dotted line represents the process of initializing the pseudo label.

$$M_{i,j}^r = r_i r_j^T \quad (r_i, r_j \in R), \quad (4)$$

where R is the set of r relations in G .

The entity features from the previous hidden layer $e'_i \in \mathbb{R}^{N \times N}$ are first transformed into two feature spaces f, g to calculate the attention, where

$$\beta_{i,j} = \frac{\exp\left(f\left(W_f e'_i\right)^T g\left(W_g e'_j\right)\right)}{\sum_{i=1}^N \exp\left(f\left(W_f e'_i\right)^T g\left(W_g e'_j\right)\right)}, \quad (5)$$

and $\beta_{i,j}$ indicates the extent to which the model attends to the i th location when synthesizing the j th region. $W_f \in \mathbb{R}^{N \times N}$, $W_g \in \mathbb{R}^{C \times N}$. Here, C is the number of channels, N is the number of feature locations of features from the previous hidden layer. The output of the attention layer is $o = (o_1, o_2, \dots, o_i, \dots, o_n) \in \mathbb{R}^{N \times N}$, where

$$o_i = \sum_{j=1}^N \beta_{i,j} \left(f\left(e'_i\right) - g\left(e'_j\right) \right), \quad (6)$$

$$\theta_i = \text{sigmoid}(M o_i + b), \quad (7)$$

where θ as the gate to control the combination of both source entity and target entity. M and b are the weight matrix and bias vector, respectively. Finally, the new vector of the node is obtained by inputting the node e_i vector and the new information vector of node o_i into Equation (8).

$$e_i^{(l+1)} = (1 - \theta_i) \odot o_i^{(l)} + \theta_i \odot e_i^{(l)} \quad (8)$$

where $(1 - \theta_i) \odot$ and $\theta_i \odot$ are working as selectors to choose information to be forgotten and remembered, respectively. $e_i^{(l+1)}$ is the embedding of node i , which is the output feature vector of the $(l+1)$ -th GateGAT layer with respect to node i .

3.3 Graph Interactive Divergence

In order to mine parallel data without supervision signals, we use graph interaction divergence to measure the pair of source entities and target entities. We aim to construct a new metric that can measure the difference score between a pair of entities (e_s, e_t) . If two entities are similar, we expect their differences to be insignificant. We construct the divergence by calculating the distance between the source entity and the target entity, which are defined as follows:

$$D(e_s, e_t) = \cos(e_s, e_t) \quad (9)$$

where $D(\cdot, \cdot)$ is the divergence between two entities, $\cos(\cdot, \cdot)$ is the cosine distance, e_s is the source entity, and e_t is the target entity. To calculate the divergence of the cross-lingual knowledge graph neighbor entity nodes of the source entity and the target entity, respectively, and construct the graph interaction divergence $GID(e_s, e_t)$ between the entities, which is defined as follows:

$$GID(e_s, e_t) = D(e_s, e_t) - \alpha \text{div}(D)(e_s, e_t) \quad (10)$$

where α represents a hyper-parameters, and

$$\text{div}(D)(e_s, e_t) = \overline{\sum_{p \in N(e_t)} D(e_s, p)} + \overline{\sum_{q \in N(e_s)} D(e_t, q)} \quad (11)$$

where $N(\cdot)$ is the set of neighbor nodes of the entity, p and q represent the entity neighbor nodes.

3.4 Building Entity Alignment Seeds

In the Unsupervised algorithm for Building Entity Alignment seeds (UBEAS) algorithm (Algorithm 1), the key idea is to calculate the graph interaction divergence of cross-lingual knowledge graph entity pairs. An unsupervised algorithm is created to build a pseudo-labels set based on the graph interaction divergence. Entity sets e_s and e_t are constructed in the source knowledge graph E_s and the target knowledge graph E_t , respectively. The initial entity embedding is obtained from the initial semantic embedding, and then the similarity of cross-lingual knowledge graph entities through the graph interaction divergence is measured. For each pair of the entities (e_s, e_t) , (e_t, e_s) , our constructs a rule (6 and 7 of Algorithm

Algorithm 1 Unsupervised algorithm for Building Entity Alignment seeds (UBEA)

Input: $E_s, E_t, SE(\text{Semantic Embedding})$
//SE: GloVe embedding;
Output: PLS //Pseudo-Labels Set
//Construct pseudo-labels set through graph interaction divergence.

```

1:  $PLS = []$ 
2: for  $e_s$  in  $E_t$  and  $e_t$  in  $E_s$  do
3:    $e_s, e_t = SE[e_s], SE[e_t]$ 
4:    $A = \text{sorted}(\text{GID}(e_s, e_t))$ 
5:    $B = \text{sorted}(\text{GID}(e_s, e_t))$ 
6:   for  $e_s, e_t$  in  $A$  do
7:     if  $B[e_s][1] = e_s$  : then
8:        $PLS.append[e_s, e_t]$ 
9:     end if
10:  end for
11:  if  $(e_s, e_t)$  in  $PLS$  then
12:     $E_s.remove(e_s)$ 
13:     $E_t.remove(e_t)$ 
14:  end if
15: end for
16: return  $PLS$ 

```

1) to enhance the entity label pair $(e_s, e_t^1), (e_s^1, e_t)$ (1 represents the 1-ranked entity), and aggregate all the cross-lingual knowledge graph entity pairs (e_s^1, e_t^1) mined into a pseudo-labels set. An outline of the algorithm is shown in Algorithm 1.

3.5 Non-sampling calibration strategy

As KGs provide semantic relations between entities, it is natural to incorporate the semantics of the relational facts into entity modeling. As discussed in Section 2, GCN needs the structures of two KGs to be highly similar or relation alignment for entity alignment. Here, we borrow the translational assumption from TransE [2]. The detailed steps are as follows:

$$\mathbf{r} = \frac{1}{|\mathcal{T}_r|} \sum_{(s,t) \in \mathcal{T}_r} (\mathbf{e}_s - \mathbf{e}_t) \quad (12)$$

where \mathcal{T}_r is the subject-target entity pairs of relation r . Then the following relation loss for refinement is minimized:

$$L_r = \sum_{r \in \mathcal{R}} \frac{1}{|\mathcal{T}_r|} \sum_{(s,t) \in \mathcal{T}_r} \|\mathbf{e}_s - \mathbf{e}_t - \mathbf{r}\| \quad (13)$$

where \mathcal{R} is the set of the total relations in the KGs.

An efficient optimization method without sampling is introduced in this section, which is the basis for learning our proposed model from the whole heterogeneous feedback data. The observed interactions are somewhat limited for pseudo-labeled data, while the number of unobserved examples is high. To learn model parameters, He and Sun [12] introduce a weighted regression loss that associates confidence to each prediction in the implicit data matrix. Following this idea, for a batch of pseudo labels $PLS(e_s, e_t)$ and the whole entity set E , their loss of a matrix M_E is:

Algorithm 2 Iterative Learning (IL)

Input:
(1) E_s and E_t
(2) PLS //come from UBEA(Algorithm 1)
(3) total number of iterations T

```

1: while  $t < T$  do
2:    $E = \text{GateGAT}(E_s, E_t, PLS)$ 
3:    $PLS = \text{UBEA}(E_s, E_t, E)$ 
   // Until the iteration reaches  $T$ .
4: end while
5: return  $PLS$ 

```

$$L_s = \sum_{e^s \in P^s} \sum_{e \in E} W_{e^s v} (M_{(e^s)} - M_{(e)})^2 \quad (14)$$

$$= \sum_{e^s \in P^s} \sum_{e \in E} W_{e^s v} \left(M_{(e^s)}^2 - 2M_{(e^s)}M_{(e)} + M_{(e)}^2 \right)$$

$$L_t = \sum_{e^t \in P^t} \sum_{e \in E} W_{e^t v} (M_{(e^t)} - M_{(e)})^2 \quad (15)$$

$$= \sum_{e^t \in P^t} \sum_{e \in E} W_{e^t v} \left(M_{(e^t)}^2 - 2M_{(e^t)}M_{(e)} + M_{(e)}^2 \right)$$

where W is a weight matrix to be learned. The time complexity of computing this loss is $O(|P||E|d)$. The final objective of UPLR is the combination of the relation loss and without negative sampling loss, aiming at injecting relation semantics to the preserved graph structures:

$$L_N = \lambda(2L_r - L_s - L_t) + b \quad (16)$$

where λ and b are hyper-parameters.

3.6 Iterative Learning

Based on the construction algorithm of the unsupervised entity alignment seed (Algorithm 1), a gradual enhancement strategy is proposed to iteratively build the entity pseudo-label set and learn the knowledge graph representation (Algorithm 2). The goal is to optimize the pseudo-labels set and improve the mapping between the two spaces. In Algorithm 2, the pseudo-labels set and knowledge representation of entity are optimized by iterating T times. In the iteration process, updating the pseudo-labels set from the new training model every iteration will be updated (Algorithm 2, line 7).

4 EXPERIMENTS

DBP15K [30] contains three datasets built from DBpedia, namely DBP15K_{ZH-EN} (Chinese-English), DBP15K_{JA-EN} (Japanese-English) and DBP15K_{FR-EN} (French-English). Each dataset has 15 thousand reference entity alignment, and the relation alignment is partially known.

4.1 Evaluation Metrics

Like previous works [3, 5, 15], we use $Hits@k$ (Hk) and Mean Reciprocal Rank (MRR) to evaluate the effectiveness of our proposed model, where a $Hits@k$ (Hk) score (higher is better) is computed by measuring the proportion of correctly ranked cross-lingual knowledge graph entity pairs ranked in the top k .

Table 1: Statistics of the DBP15K

| Datasets | | Entities | Relations | Triples |
|-------------------------|----|----------|-----------|---------|
| DBP15K _{ZH-EN} | ZH | 66,469 | 2,830 | 153,929 |
| | EN | 98,125 | 2,317 | 237,674 |
| DBP15K _{JA-EN} | JA | 65,733 | 2,043 | 164,373 |
| | EN | 95,680 | 2,096 | 233,319 |
| DBP15K _{FR-EN} | FR | 66,858 | 1,379 | 192,191 |
| | EN | 105,889 | 2,209 | 278,590 |

4.2 Implementation Details

The configuration is: *Attention-heads* = 2, $\alpha = 0.5$, $\gamma = 3$, $lr = 0.005$, *Depth* = 3, *Node-hidden* = 125, *Rel-hidden* = 125, *epoch* = 6, $T = 5$. In order to utilize entity names in different KGs for better initialization, they [39] uses Google Translate to translate Chinese, Japanese, and French entity names into English, and then use pre-trained English semantic embedding GloVe.840B.300d to construct the input entity representations for the primal graph. Note: Google Translate can not guarantee accurate translations for named entities without any context. We use Tensorflow along with Python 3.6 to perform our experiments. All experiments are executed on a Linux machine with processor Intel(R) Xeon(R) Gold 5218 CPU @ 2.30GHz, 256GB RAM, and GeForce RTX 1080Ti GPU.

4.3 Settings of Experiments

In order to analyze the performance of the model, we selected some classic and state-of-the-art entity alignment methods as baselines in our experiments (the parameters of the baseline models are all default parameters set in public resources). Our proposed model evaluates the performance of the model from three aspects: (1) Main results; (2) Ablation study; (3) Auxiliary Experiments.

4.4 Main Results

Our proposed model has omitted some methods that require auxiliary entity information that is not used by others (see Section 2). Although the semantic information in the process of unsupervised entity alignment is added, we do not add semantic information to knowledge representation. Therefore, there are three types of comparative experiments designed: Graph structure, Graph structure + Iterative learning, and Unsupervised.

Only use graph structures.

Our proposed model compares the differences in methods based on graph structural features, which includes nine approaches. Table 2 shows the performance comparison of the basic graph structure methods. Obviously, on all evaluation indicators, the model’s performance is consistently rated as the best among all elementary graph structure methods. In particular, compared with the state-of-the-art method RREA, the method on the DBP15K_{ZH-EN} dataset improves by 12.6% on $H1$, and the method on the DBP15K_{JA-EN} dataset increases by 13.8% on $H1$, and the method on the DBP15K_{FR-EN} dataset improves by 18.0% on $H1$. The main reason is that the unsupervised entity alignment algorithm establishes more seeds for entity alignment, and the information between knowledge graphs across languages can be better optimized through unsampled loss functions.

In addition, our proposed model unsupervised entity alignment algorithm can significantly improve the performance of all methods on all datasets. We observe that the performance gap of textual methods in different datasets is much more important than that of structural methods. The performance of the model has increased most in the French dataset. This is because French words are easily mapped into English when constructing entity alignment seeds based on semantic embeddings. All in all, UPLR breaks the upper limit of the performance of the purely structure-based entity alignment method, which proves that the design is efficient.

Use graph structure and iterative learning.

Our proposed model outperforms the eight models using only graph structures and iterative learning models and achieves the best results for all datasets. Specifically, BootEA is a transe-based model with better performance than some models that only use the graph structure because it uses a bootstrap strategy to expand entity alignment seeds iteratively. This demonstrates that the iterative strategy can significantly improve the performance of entity alignment.

Our proposed model adds an iterative strategy to improve $H1$ by at least 6%. Our proposed unsupervised method enables better knowledge representation for KGs, thereby improving the entity alignment performance of the model.

Unsupervised.

Compared with other unsupervised entity alignment methods, Our proposed model has certain advantages. This method is improved on the DBP15K data set, which confirms that the seed based on semantic embedding can improve the performance of unsupervised entity alignment more effectively than the seed based on image information. In reality, image information is often not directly obtained, or it is not easy to get high-quality image-entity translation pairs. In addition, although SEU has achieved good results, it solves the task of entity alignment through static embedding (semantic embedding). Due to the difference between static embedding and real data (We discuss the performance of the static embedding based model in the SE in Table 3), the model performance is not as good as our proposed model. Therefore, this method is considered one of the most effective unsupervised entity alignment methods.

4.5 Ablation Study

In the above experiments, the overall effectiveness of the method is proved. In this section, we demonstrate the validity of each component of UPLR (GateGAT+ L_N + P + IL). Ablation studies were conducted from multiple aspects, and the results are shown in Table 3. L_s represents a margin-based scoring function (this is a commonly [39] used method for calculating loss functions in supervised learning), GAT represents graph attention network, $GateGAT$ represents gate graph attention network, L_S represents supervised loss function [39] requiring seeds, IL represents iterative learning, L_N represents our non-sampling calibration strategy, S represents entity-aligned seeds (seed accounts for 30% of the total dataset) and P represents pseudo-label dataset. SE stands for semantic embedding (it comes from GloVe.840B.300d, and is detailed in (4.2)). In experiments requiring S , the $H1$ value of GateGAT was higher than that of $GAT H1$. However, the $H10$ value of GateGAT is slightly lower than $GAT H10$, which may be because Gate reduces the

Table 2: Performance on DBP15K

| Models | ZH-EN | | | JA-EN | | | FR-EN | | |
|--------------------------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | $H1$ | $H10$ | MRR | $H1$ | $H10$ | MRR | $H1$ | $H10$ | MRR |
| Graph structure | | | | | | | | | |
| MTransE [5] | 0.209 | 0.512 | 0.310 | 0.25 | 0.572 | 0.360 | 0.247 | 0.577 | 0.360 |
| KECG [15] | 0.478 | 0.835 | 0.598 | 0.49 | 0.844 | 0.610 | 0.486 | 0.851 | 0.610 |
| MuGNN [3] | 0.494 | 0.844 | 0.611 | 0.501 | 0.857 | 0.621 | 0.495 | 0.870 | 0.621 |
| RSN [10] | 0.508 | 0.745 | 0.591 | 0.507 | 0.737 | 0.590 | 0.516 | 0.768 | 0.605 |
| GMNN [44] | 0.539 | 0.826 | - | 0.433 | 0.681 | - | 0.465 | 0.728 | - |
| AliNet [33] | 0.539 | 0.826 | 0.628 | 0.549 | 0.831 | 0.552 | 0.645 | 0.852 | 0.657 |
| HyperKA [28] | 0.572 | 0.865 | 0.678 | 0.564 | 0.865 | 0.673 | 0.597 | 0.891 | 0.704 |
| MRAEA(Basic) [21] | 0.638 | 0.886 | 0.736 | 0.646 | 0.891 | 0.735 | 0.666 | 0.912 | 0.765 |
| RREA(Basic) [22] | 0.715 | 0.929 | 0.794 | 0.713 | 0.933 | 0.793 | 0.739 | 0.946 | 0.816 |
| Dual-AMN(Basic) [19] | 0.731 | 0.923 | 0.799 | 0.726 | 0.927 | 0.799 | 0.756 | 0.948 | 0.827 |
| UPLR(Basic) | 0.841 | 0.965 | 0.887 | 0.851 | 0.974 | 0.898 | 0.919 | 0.991 | 0.948 |
| Graph structure + Iterative Learning | | | | | | | | | |
| BootEA [31] | 0.629 | 0.847 | 0.703 | 0.622 | 0.854 | 0.701 | 0.653 | 0.874 | 0.731 |
| NAEA [48] | 0.650 | 0.867 | 0.720 | 0.641 | 0.873 | 0.718 | 0.673 | 0.894 | 0.752 |
| TransEdge [32] | 0.735 | 0.919 | 0.801 | 0.719 | 0.932 | 0.795 | 0.710 | 0.941 | 0.796 |
| MRAEA(Iter) [21] | 0.757 | 0.930 | 0.827 | 0.758 | 0.934 | 0.826 | 0.781 | 0.948 | 0.849 |
| DGMC [8] | 0.801 | 0.874 | - | 0.848 | 0.897 | - | 0.933 | 0.960 | - |
| RREA(Semi) [22] | 0.801 | 0.948 | 0.857 | 0.802 | 0.952 | 0.858 | 0.827 | 0.966 | 0.881 |
| Dual-AMN(Semi) [19] | 0.808 | 0.940 | 0.857 | 0.801 | 0.949 | 0.855 | 0.840 | 0.965 | 0.888 |
| RNM [49] | 0.840 | 0.919 | 0.870 | 0.872 | 0.944 | 0.899 | 0.938 | 0.981 | 0.954 |
| Unsupervised | | | | | | | | | |
| PRASE [25] | 0.651 | - | - | 0.726 | - | - | 0.757 | - | - |
| EVA [17] | 0.761 | 0.907 | 0.814 | 0.762 | 0.913 | 0.817 | 0.793 | 0.942 | 0.847 |
| SEU(w) [20] | 0.816 | 0.923 | 0.854 | 0.865 | 0.952 | 0.896 | 0.953 | 0.989 | 0.967 |
| UPLR | 0.902 | 0.970 | 0.927 | 0.912 | 0.978 | 0.937 | 0.967 | 0.994 | 0.974 |

Table 3: The ablation experiment of each component of the model.

| | ZH-EN | | JA-EN | | FR-EN | |
|-------------------------|-------|-------|-------|-------|-------|-------|
| | $H1$ | $H10$ | $H1$ | $H10$ | $H1$ | $H10$ |
| SE | 0.586 | 0.710 | 0.664 | 0.784 | 0.825 | 0.897 |
| GAT+ L_S + S | 0.707 | 0.931 | 0.705 | 0.935 | 0.729 | 0.952 |
| GAT+ L_S + S + IL | 0.798 | 0.946 | 0.793 | 0.948 | 0.819 | 0.962 |
| GateGAT+ L_S + S + IL | 0.807 | 0.940 | 0.799 | 0.946 | 0.840 | 0.966 |
| UPLR(w/o L_N) | 0.846 | 0.947 | 0.883 | 0.961 | 0.951 | 0.990 |
| UPLR(w/o Gate) | 0.893 | 0.971 | 0.906 | 0.979 | 0.959 | 0.994 |
| UPLR(w/o IL) | 0.841 | 0.965 | 0.851 | 0.974 | 0.919 | 0.991 |
| UPLR | 0.902 | 0.970 | 0.912 | 0.978 | 0.967 | 0.994 |

weight of the noise entity, thereby increasing the critical metric $H1$, while $H10$ is slightly lower. As discussed in Section 3, UPLR has two novel designs : (1) *Gate* and (2) L_N . From Table 3, the performance of both designs is greatly improved. These ablation experiments show that the design is meaningful and has improved significantly.

4.6 Auxiliary Experiments

4.6.1 The impact of pseudo-labeling. Our proposed method constructs a new pseudo-labels set (The results are shown in Figure 3(a)) for DBP15K by random selection.

The randomly selected parameters are based on SE in Table 3. Our proposed method constructs a pseudo-labels set of cross-lingual knowledge graph entity pairs with DBP15K_{ZH-EN}: 58.6%, DBP15K_{JA-EN}: 66.4%, and DBP15K_{FR-EN}: 82.5% accuracy, respectively. The results are presented in Figure 3. Although the randomly selected pseudo-labels set and the pseudo-labels set calculated by semantic embedding have the same number of correct label entity pairs and the same number of incorrect label entity pairs, their final results are different. The performance of the randomly selected pseudo-labels set deteriorates in the three languages pairs. This experiment found that with the same correct labels, entities with different types of mislabeled have different contributions to the model, and mislabeled entity pairs obtained through semantic embedding can play a positive role in the entity alignment process. Therefore, the model’s good performance depends not only on the algorithm but also on the important cross-lingual knowledge graph entity pairs discovered through semantic embedding.

4.6.2 The impact of different ratios of semantic embedding. In practical applications, constructing entity pseudo-labels by semantic embedding consumes many computing resources, especially for large-scale KGs. We hope that this model can maintain a good performance under the condition of limited semantic embedding. The results are presented in Figure 3(b). In order to examine the impact of semantic embedding with different ratios on the performance of the entity alignment model, Our proposed model sets up multiple sets of experiments. To investigate the performance with different

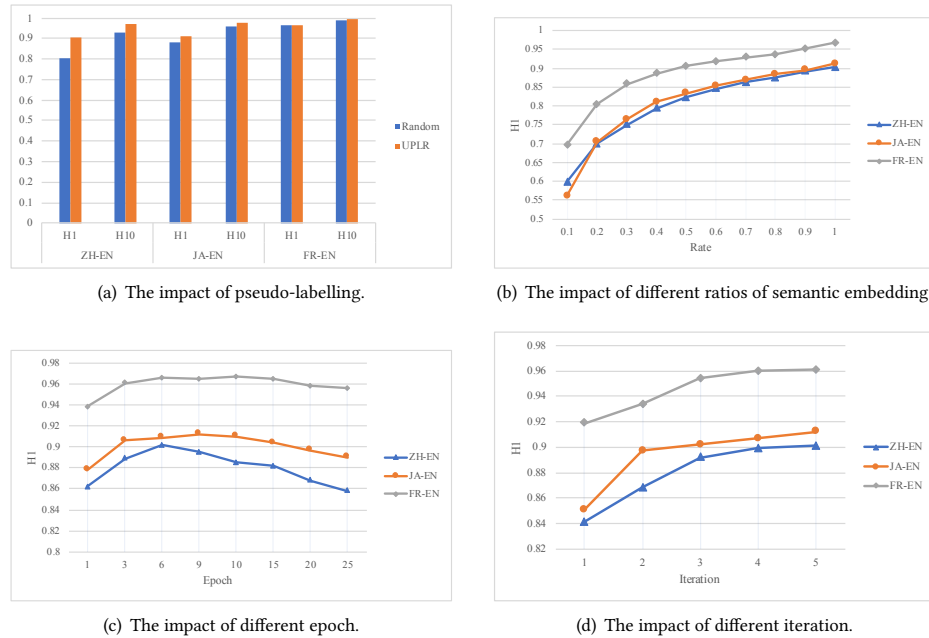


Figure 3: Auxiliary Experiments

pre-aligned ratios, we set the ratios from 0.1 to 1. It is noted that if we only use a 0.1 ratio of semantic embedding information is used, our proposed model can surpass AliNet [33] (AliNet uses 30% seed). The growth rate of the model performance is the highest when semantic embedding with 40% coverage is added to the model. With more semantic embedding added to the model, the model performance increases slowly, although the model can be enhanced. These results show that the unsupervised entity alignment model is effective. A small amount of semantic embedding can achieve good results for entity alignment tasks without entity alignment seeds.

4.6.3 The impact of different epoch. In addition to our proposed architecture, the non-sampling calibration loss is one of our main contributions, and to verify its effectiveness, we compare DBP15K datasets of different epochs in different languages. The result is shown in Figure 3(c).

It is observed that our proposed non-sampling calibration strategy can make the model converge faster and achieve the best performance. In the experiment, the supervised loss function (RREA [22] uses 3000 epochs) usually requires thousands of epochs to converge. In the DBP15K_{ZH-EN} data set, our proposed model with a non-sampling calibration strategy is 5.6% higher than the model with a supervised loss function (Table 3, UPLR (w/o L_N)). Experimental results show that our proposed non-sampling calibration loss function significantly improves the convergence speed without affecting the accuracy.

4.6.4 The impact of different iteration. In order to verify the effectiveness of this method, the impact of different iteration times on performance is further analyzed. The result is shown in Figure 3(d). It is observed that the $H1$ score of our model increases

with the number of iterations. This observation is consistent with our expectations because pseudo-labels are continuously refined through iterations. The UPLR model will quickly converge to the optimal solution if an appropriate number of iterations is selected. However, if the number of iterations is too large, the algorithm cannot converge. However, UPLR still shows satisfactory performance without iteration compared with other methods. It is noted that our proposed model has exceeded the performance of SEU [20] (SEU uses 10 iterations) without iteration.

5 CONCLUSIONS

We break the dependence of entity alignment models on seeds, and we propose an uncertainty-aware pseudo label refinery (UPLR) for entity alignment. We find that the challenge of solving the entity alignment task using pseudo-labels is to reduce the impact of noise on the model. We propose a non-sampling calibration strategy and a progressive enhancement strategy to reduce the influence of noise on the model. The non-sampling calibration strategy can reduce the impact of noise on the model without artificially designing a threshold, and can effectively reduce the impact of noise on the model in model optimization. The gradual enhancement strategy can enhance the similarity of the two domains, so that high-quality pseudo-labels can be generated. In future work, we will apply our proposed model to more datasets.

ACKNOWLEDGMENTS

This work was supported by National Key Research and Development Program of China (Grant No. 2020YFC0833402), National Natural Science Foundation of China (Grant Nos. 61976021, U1811262), and Beijing Academy of Artificial Intelligence (BAAI).

REFERENCES

- [1] Philip Bachman, Ouais Alsharif, and Doina Precup. 2014. Learning with pseudo-ensembles. *Advances in neural information processing systems* 27 (2014), 3365–3373.
- [2] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *NIPS*, 2013.
- [3] Yixin Cao, Zhiyuan Liu, Chengjiang Li, Juanzi Li, and Tat-Seng Chua. 2019. Multi-channel graph neural network for entity alignment. In *ACL*, 2019.
- [4] Mathilde Caron, Piotr Bojanowski, Armand Joulin, and Matthijs Douze. 2018. Deep clustering for unsupervised learning of visual features. In *ECCV*, 2018.
- [5] Muhao Chen, Yingtao Tian, Mohan Yang, and Carlo Zaniolo. 2017. Multilingual knowledge graph embeddings for cross-lingual knowledge alignment. In *IJCAI*, 2017.
- [6] Soham Datta, Prabir Mallick, Sangameshwar Patil, Indrajit Bhattacharya, and Girish Palshikar. 2021. Generating An Optimal Interview Question Plan Using A Knowledge Graph And Integer Linear Programming. In *NAACL*, 2021.
- [7] Hao Fei, Yafeng Ren, Yue Zhang, Donghong Ji, and Xiaohui Liang. 2021. Enriching contextualized language model from knowledge graph for biomedical information extraction. *Briefings in Bioinformatics* 22, 3 (2021).
- [8] Matthias Fey, Jan E Lenssen, Christopher Morris, Jonathan Masci, and Nils M Kriege. 2020. Deep graph matching consensus. In *ECCV*, 2020.
- [9] Fan Gong, Meng Wang, Haofen Wang, Sen Wang, and Mengyue Liu. 2021. SMR: Medical knowledge graph embedding for safe medicine recommendation. *Big Data Research* 23 (2021), 100174.
- [10] Lingbing Guo, Zequn Sun, and Wei Hu. 2019. Learning to exploit long-term relational dependencies in knowledge graphs. In *ICML*, 2019.
- [11] Hangfeng He and Xu Sun. 2017. A unified model for cross-domain and semi-supervised named entity recognition in chinese social media. In *AAAI*, 2017.
- [12] Yifan Hu, Yehuda Koren, and Chris Volinsky. 2008. Collaborative filtering for implicit feedback datasets. In *ICDM*, 2008.
- [13] Samuli Laine and Timo Aila. 2016. Temporal ensembling for semi-supervised learning. *arXiv preprint arXiv:1610.02242* (2016).
- [14] Dong-Hyun Lee et al. 2013. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *ICML*, 2013.
- [15] Chengjiang Li, Yixin Cao, Lei Hou, Jiaxin Shi, Juanzi Li, and Tat-Seng Chua. 2019. Semi-supervised entity alignment via joint knowledge embedding model and cross-graph model. In *EMNLP*, 2019.
- [16] Xixun Lin, Hong Yang, Jia Wu, Chuan Zhou, and Bin Wang. 2019. Guiding cross-lingual entity alignment via adversarial knowledge embedding. In *ICDM*, 2019.
- [17] Fanguy Liu, Muhao Chen, Dan Roth, and Nigel Collier. 2021. Visual Pivoting for (Unsupervised) Entity Alignment. In *AAAI*, 2021.
- [18] Zhiyuan Liu, Yixin Cao, Liangming Pan, Juanzi Li, and Tat-Seng Chua. 2020. Exploring and Evaluating Attributes, Values, and Structures for Entity Alignment. In *EMNLP*, 2020.
- [19] Xin Mao, Wenting Wang, Yuanbin Wu, and Man Lan. 2021. Boosting the Speed of Entity Alignment 10³: Dual Attention Matching Network with Normalized Hard Sample Mining. In *WWW*, 2021.
- [20] Xin Mao, Wenting Wang, Yuanbin Wu, and Man Lan. 2021. From Alignment to Assignment: Frustratingly Simple Unsupervised Entity Alignment. In *EMNLP*, 2021.
- [21] Xin Mao, Wenting Wang, Huimin Xu, Man Lan, and Yuanbin Wu. 2020. MRAEA: an efficient and robust entity alignment approach for cross-lingual knowledge graph. In *WSDM*, 2020.
- [22] Xin Mao, Wenting Wang, Huimin Xu, Yuanbin Wu, and Man Lan. 2020. Relational Reflection Entity Alignment. In *CIKM*, 2020.
- [23] Shichao Pei, Lu Yu, Robert Hoehndorf, and Xiangliang Zhang. 2019. Semi-supervised entity alignment via knowledge graph embedding with awareness of degree difference. In *WWW*, 2019.
- [24] Shichao Pei, Lu Yu, Guoxian Yu, and Xiangliang Zhang. 2020. Rea: Robust cross-lingual entity alignment between knowledge graphs. In *SIGKDD*, 2020.
- [25] Zhiyuan Qi, Ziheng Zhang, Jiaoyan Chen, Xi Chen, Yuejia Xiang, Ningyu Zhang, and Yefeng Zheng. 2021. Unsupervised Knowledge Graph Alignment by Probabilistic Reasoning and Semantic Embedding. In *IJCAI*, 2021.
- [26] Weiwei Shi, Yihong Gong, Chris Ding, Zhiheng MaXiaoYu Tao, and Nanning Zheng. 2018. Transductive semi-supervised deep learning using min-max features. In *ECCV*, 2018.
- [27] Ke Sun, Zhouchen Lin, and Zhanxing Zhu. 2020. Multi-stage self-supervised learning for graph convolutional networks on graphs with few labeled nodes. In *AAAI*, 2020.
- [28] Zequn Sun, Muhao Chen, Wei Hu, Chengming Wang, Jian Dai, and Wei Zhang. 2020. Knowledge Association with Hyperbolic Knowledge Graph Embeddings. In *EMNLP*, 2020.
- [29] Z. Sun, W. Hu, and C. Li. 2017. Cross-lingual Entity Alignment via Joint Attribute-Preserving Embedding. *Springer, Cham* (2017).
- [30] Zequn Sun, Wei Hu, and Chengkai Li. 2017. Cross-lingual entity alignment via joint attribute-preserving embedding. In *ISWC*, 2017.
- [31] Zequn Sun, Wei Hu, Qingheng Zhang, and Yuzhong Qu. 2018. Bootstrapping Entity Alignment with Knowledge Graph Embedding. In *IJCAI*, 2018.
- [32] Zequn Sun, Jiacheng Huang, Wei Hu, Muhao Chen, Lingbing Guo, and Yuzhong Qu. 2019. Transedge: Translating relation-contextualized embeddings for knowledge graphs. In *ISWC*, 2019.
- [33] Zequn Sun, Chengming Wang, Wei Hu, Muhao Chen, Jian Dai, Wei Zhang, and Yuzhong Qu. 2020. Knowledge graph alignment network with gated multi-hop neighborhood aggregation. In *AAAI*, 2020.
- [34] Antti Tarvainen and Harri Valpola. 2017. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *arXiv preprint arXiv:1703.01780* (2017).
- [35] B. D. Trsedya, J. Qi, and Z. Rui. 2019. Entity Alignment between Knowledge Graphs Using Attribute Embeddings. In *AAAI*, 2019.
- [36] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2017. Graph attention networks. *arXiv preprint arXiv:1710.10903* (2017).
- [37] Zhichun Wang, Qingsong Lv, Xiaohan Lan, and Yu Zhang. 2018. Cross-lingual knowledge graph alignment via graph convolutional networks. In *EMNLP*, 2018.
- [38] Zhichun Wang, Jinjian Yang, and Xiaoju Ye. 2020. Knowledge Graph Alignment with Entity-Pair Embedding. In *EMNLP*, 2020.
- [39] Yuting Wu, Xiao Liu, Yansong Feng, Zheng Wang, Rui Yan, and Dongyan Zhao. 2019. Relation-aware entity alignment for heterogeneous knowledge graphs. In *IJCAI*, 2019.
- [40] Yuting Wu, Xiao Liu, Yansong Feng, Zheng Wang, and Dongyan Zhao. 2019. Jointly learning entity and relation representations for entity alignment. In *EMNLP*, 2019.
- [41] Abhinav Gupta Xiaolong Wang, Ross Girshick and Kaiming He. 2018. Non-local Neural Networks. In *CVPR*, 2018.
- [42] Qizhe Xie, Minh-Thang Luong, Eduard Hovy, and Quoc V Le. 2020. Self-training with noisy student improves imagenet classification. In *CVPR*, 2020.
- [43] Kun Xu, Liwei Wang, Mo Yu, Yansong Feng, Yan Song, Zhiguo Wang, and Dong Yu. 2019. Cross-lingual knowledge graph alignment via graph matching neural network. In *ACL*, 2019.
- [44] Kun Xu, Liwei Wang, Mo Yu, Yansong Feng, Yan Song, Zhiguo Wang, and Dong Yu. 2019. Cross-lingual knowledge graph alignment via graph matching neural network. In *ACL*, 2019.
- [45] Yuchen Yan, Lihui Liu, Yikun Ban, Baoyu Jing, and Hanghang Tong. 2021. Dynamic Knowledge Graph Alignment. In *AAAI*, 2021.
- [46] Hsiu-Wei Yang, Yanyan Zou, Peng Shi, Wei Lu, Jimmy Lin, and Xu Sun. 2020. Aligning cross-lingual entities with multi-aspect information. In *EMNLP*, 2020.
- [47] Wexin Zeng, Xiang Zhao, Jiuyang Tang, and Xuemin Lin. 2020. Collective Embedding-based Entity Alignment via Adaptive Features. In *ICDE*, 2020.
- [48] Qiannan Zhu, Xiaofei Zhou, Jia Wu, Jianlong Tan, and Li Guo. 2019. Neighborhood-Aware Attentional Representation for Multilingual Knowledge Graphs. In *IJCAI*, 2019.
- [49] Yao Zhu, Hongzhi Liu, Zhonghai Wu, and Yingpeng Du. 2021. Relation-Aware Neighborhood Matching Model for Entity Alignment. In *AAAI*, 2021.