

Singapore Management University

Institutional Knowledge at Singapore Management University

Research Collection School Of Computing and Information Systems

School of Computing and Information Systems

10-2022

Interactive contrastive learning for self-supervised entity alignment

Kaisheng ZENG

Zhenhao DONG

Lei HOU

Yixin CAO

Singapore Management University, yxcao@smu.edu.sg

Minghao HU

See next page for additional authors

Follow this and additional works at: https://ink.library.smu.edu.sg/sis_research



Part of the [Databases and Information Systems Commons](#), and the [Graphics and Human Computer Interfaces Commons](#)

Citation

ZENG, Kaisheng; DONG, Zhenhao; HOU, Lei; CAO, Yixin; HU, Minghao; YU, Jifan; LV, Xin; CAO, Lei; WANG, Xin; LIU, Haozhuang; HUANG, Yi; WAN, Jing; and LI, Juanzi. Interactive contrastive learning for self-supervised entity alignment. (2022). *Proceedings of the 31st ACM International Conference on Information and Knowledge Management, Atlanta, Georgia, 2022 October 17-21*. 2465-2475.

Available at: https://ink.library.smu.edu.sg/sis_research/7452

This Conference Proceeding Article is brought to you for free and open access by the School of Computing and Information Systems at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection School Of Computing and Information Systems by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email cherylids@smu.edu.sg.

Author

Kaisheng ZENG, Zhenhao DONG, Lei HOU, Yixin CAO, Minghao HU, Jifan YU, Xin LV, Lei CAO, Xin WANG, Haozhuang LIU, Yi HUANG, Jing WAN, and Juanzi LI

Interactive Contrastive Learning for Self-Supervised Entity Alignment

Kaisheng Zeng
zks19@mails.tsinghua.edu.cn
Tsinghua University
Beijing, China

Zhenhao Dong
dzh0129@outlook.com
Beijing University of Chemical
Technology
Beijing, China

Lei Hou*
houlei@tsinghua.edu.cn
Department of Computer Science and
Technology, BNRist, Tsinghua
Beijing, China

Yixin Cao
Singapore Management University
Singapore

Minghao Hu
Information Research Center of
Military Science
Beijing, China

Jifan Yu
Xin Lv
Tsinghua University
Beijing, China

Lei Cao
Tsinghua University
Beijing, China

Xin Wang
Tsinghua University
Beijing, China

Haozhuang Liu
Tsinghua University
Beijing, China

Yi Huang
Junlan Feng
China Mobile Research Institute
Beijing, China

Jing Wan
Beijing University of Chemical
Technology
Beijing, China

Juanzi Li
Ling Feng
CST, BNRist, Tsinghua University
Beijing, China

ABSTRACT

Self-supervised entity alignment (EA) aims to link equivalent entities across different knowledge graphs (KGs) without the use of pre-aligned entity pairs. The current state-of-the-art (SOTA) self-supervised EA approach draws inspiration from contrastive learning, originally designed in computer vision based on instance discrimination and contrastive loss, and suffers from two shortcomings. Firstly, it puts unidirectional emphasis on pushing sampled negative entities far away rather than pulling positively aligned pairs close, as is done in the well-established supervised EA. Secondly, it advocates the minimum information requirement for self-supervised EA, while we argue that self-described KG's side information (e.g., entity name, relation name, entity description) shall preferably be explored to the maximum extent for the self-supervised EA task. In this work, we propose an interactive contrastive learning model for self-supervised EA. It conducts bidirectional contrastive learning via building pseudo-aligned entity pairs as pivots to achieve direct cross-KG information interaction. It further exploits the integration of entity textual and structural information and elaborately designs encoders for better utilization in the self-supervised setting. Experimental results show that our approach outperforms the previous best self-supervised method by a large margin (over 9% Hits@1

absolute improvement on average) and performs on par with previous SOTA supervised counterparts, demonstrating the effectiveness of the interactive contrastive learning for self-supervised EA. The code and data are available at <https://github.com/THU-KEG/ICLEA>.

CCS CONCEPTS

• **Computing methodologies** → **Neural networks**; • **Information systems** → **Information integration**.

KEYWORDS

Knowledge Graph, Entity Alignment, Self-Supervised Learning, Contrastive Learning

ACM Reference Format:

Kaisheng Zeng, Zhenhao Dong, Lei Hou, Yixin Cao, Minghao Hu, Jifan Yu, Xin Lv, Lei Cao, Xin Wang, Haozhuang Liu, Yi Huang, Junlan Feng, Jing Wan, Juanzi Li, and Ling Feng. 2022. Interactive Contrastive Learning for Self-Supervised Entity Alignment. In *Proceedings of the 31st ACM Int'l Conference on Information and Knowledge Management (CIKM '22)*, Oct. 17–21, 2022, Atlanta, GA, USA. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3511808.3557364>

1 INTRODUCTION

Knowledge Graphs (KGs) (e.g., DBpedia [19], YAGO [28], and Wikidata [43]) provide structural knowledge about the entities and relations in the real world. These separately constructed KGs contain heterogeneous but complementary knowledge [36]. Entity Alignment (EA) integrates the complementary knowledge in these KGs via identifying equivalent entities [39] and thus benefits various knowledge-driven applications such as information extraction [14], question answering [7], and recommendation system [3, 4, 11].

*Corresponding author.



This work is licensed under a Creative Commons Attribution International 4.0 License.

The mainstream solutions in the literature are based on deep representation learning, which embeds entities into a latent space, and then calculates the distance between embeddings as the evidence for EA. These approaches [5, 27, 41, 46] utilize seed alignments to guide the representation learning of entity and relation and propagate those seed alignments to the entire KG. Despite the excellent performance, they rely on some pre-aligned entity pairs provided by humans during the training process. As the acquisition of seed alignments is usually time-consuming, labor-intensive, and error-prone, it is hard to apply them to real EA-related scenarios [21].

Hence, self-supervised EA under the *unsupervised* setting (i.e., matching entities without any pre-aligned label) starts to attract research attention. One pioneer work is EVA [24], which used visual semantic representations of entities to align entities in heterogeneous KGs in a fully unsupervised setting. Very recently, SelfKG [25] drew inspiration from contrastive learning approaches, which originally targeted at computer vision tasks by leveraging instance discrimination and contrastive loss (e.g., MoCo [16], SimCLR [6]), and provided a contrastive learning framework for self-supervised entity alignment. Although SelfKG achieved a promising performance, it suffers from two shortcomings.

First, it struggles with the dilemma of precise entity alignment situation, because it puts unidirectional emphasis on pushing sampled negative entities far away for the uniformity of the whole representation space rather than pulling positively aligned pairs close to enhance the alignment. Its self-negative sampling strategy can only sample entities from the same source KG to avoid conflict, which blocks the direct cross-KG information interaction. EA is essentially building inter-KG links, and thus the cross-KG information interaction is critical. Existing methods achieve such interaction via explicit supervision (seed alignments [41]) or implicit signals (images as visual pivots [24]). We argue that pushing sampled negative entities away and pulling possibly aligned entities close are equally important. The latter is actually the core in well-established supervised EA approaches.

Second, it follows the minimum information requirement for self-supervised EA and only focuses on entity names, which ignores rich entity semantics in KGs. As shown in Figure 1, the entity relationship and description provide rich information, while sometimes relying only on entity names can introduce ambiguity problems. Entity name ambiguity problem [15] brings severe challenges to self-supervised EA; as shown in Figure 1, entity names can be deceptive to EA while KG’s side information containing extensive entity semantics may bring great benefit. So we argue that self-described KG’s side information (e.g., entity name, relation name, entity description) shall preferably be explored to the maximum extent for the self-supervised EA task.

To validate the above two hypotheses, we propose a model named Interactive Contrastive Learning for self-supervised Entity Alignment (ICLEA). We design a novel interactive contrastive learning mechanism in a self-supervised EA framework by constructing pseudo-aligned entity pairs as virtual pivots to establish a direct information interaction channel for the two KGs. Inspired by previous supervised approaches, we introduce more important side information into self-supervised EA. Specifically, we separately encode entity name labels and descriptions by different pre-trained models, then organically combine them to provide powerful initial

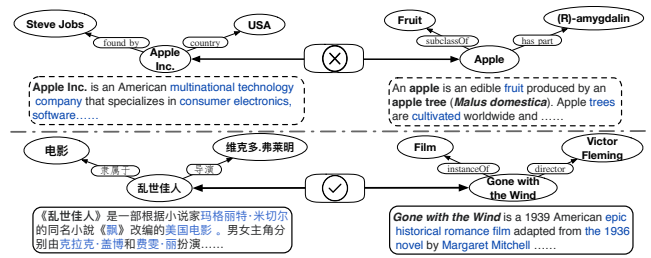


Figure 1: Examples of KG’s side information (e.g., entity description) enhancing EA when entities’ names are deceptive.

embeddings for entities. We also propose a relation-aware neighborhood aggregator to better leverage the structural and semantic information brought by the KGs’ relations.

The main contributions of our work are threefold:

- We propose an interactive contrastive learning to achieve the direct cross-KG information interaction via building pseudo-aligned entity pairs in a self-supervised EA.
- We propose to better utilize and integrate the structural and semantic entity information for a self-supervised EA framework with the help of pre-trained language models.
- Experimental results show that ICLEA outperforms the best self-supervised baseline by a large margin (over 9% Hits@1 absolute improvement on average), and performs on par with previous SOTA supervised methods while maintaining more stable model training. Our work significantly narrows the gap between supervised and self-supervised EA approaches.

2 RELATED WORK

We discuss two lines of work that are relevant to ours according to their usage of pre-aligned entity pairs during the model’s training process: (semi-)supervised and self/un-supervised approaches.

2.1 (Semi-)Supervised EA Approaches

In (semi-)supervised approaches, pre-aligned entity pairs are provided during the training process. Early EA methods focused on Bayesian-based probability estimation [40] and hand-crafted relevant similarity factors [21]. Before the wave of representation learning arrived, most EA approaches either relied on extra resources [45] or human efforts [28] to discover identical entity pairs.

The deep embedding-based EA approaches seek to encode a KG into a low-dimensional latent space. Earlier approaches focus on how to handle this task using only the structure of KG [39, 54]. MTransE [5] uses TransE to learn the vector space of a single KG and then learns a linear transformation to align two vector spaces. IPTransE [35] models relational paths by inferring the equivalence between direct relationships and multi-hop paths. DAT [58] offers a framework for EA with emphasis on long-tail entities. GNNs [1, 8, 42] are well suited for modeling graph structure, and have been widely used for modeling the multifaceted information in KGs to assist embedding-based EA recently [2, 20, 27, 30, 38, 47, 53, 55, 61, 63]. The emergence of pre-trained language models (PLMs) [18] has recently brought natural language processing to a new era. Some PLMs can bring high-quality multilingual initial embeddings,

which is crucial for the EA task. HMAN [50] uses multilingual BERT to initialize the embedding of entities by its side information which bridges multilingual gaps. BERT-INT [41] proposes an EA strategy based on directly fine-tuned multilingual BERT. Besides the high performance achieved by these (semi-) supervised models, researchers begin to consider the efficiency and practicality with new methods such as adversarial learning, active learning, etc [23, 29, 33, 56]. Despite the success of these approaches, the imperative need for labeled data still bridges the gap between these models and real-world applications. The acquisition of pre-aligned seeds is naturally time-consuming and labor-intensive [26]. To address this problem, self-supervised or unsupervised EA has gained increasing attention from the academic community.

2.2 Self-Supervised or Unsupervised EA Approaches

In self-supervised or unsupervised¹ [17, 26] approaches, none of the pre-aligned entity pairs is provided during the training process. MultiKE [59] designs some cross-KG inference methods to enhance the alignment between two KGs. EVA [24] uses images as pivots for aligning entities in different KGs in an unsupervised setting.

Self-supervised learning methods are proposed to learn data co-occurrence relationships from unlabeled data without using any human-annotated labels [17]. Self-supervised image feature learning in the computer vision (CV) community has obtained great success. The margin between self-supervised methods' performance and supervised methods' performance on some downstream tasks becomes very small [6, 16]. Strikingly, the success of self-supervised training in computer vision community provides new directions for unsupervised EA models. As there are few attempts that launch self-supervised training in this task, such as SelfKG [25], we recognize that it is urgent and purposeful to set up a paradigm so that the experience of existing EA methods can be adapted to this new direction. Specifically, SelfKG designs a unidirectional contrastive learning strategy between two KGs for self-supervised EA and uses self-negative sampling to sample entities from the same source KG to avoid conflict, which blocks the direct cross-KG information interaction. Therefore, we propose ICLEA, which jointly considers the structural and semantic information in the self-training process and constructs pseudo-aligned entity pairs as virtual pivots to guide the direct bidirectional cross-KG information interaction, and calls for more efforts in self-supervised EA.

3 PROBLEM DEFINITION

DEFINITION 1 (KNOWLEDGE GRAPH). A knowledge graph is represented as $\mathcal{G}=(\mathcal{E}, \mathcal{R}, \mathcal{T}, \mathcal{S})$, where each $e \in \mathcal{E}$, $r \in \mathcal{R}$, $t = (e_i, r, e_j) \in \mathcal{T}$ ($e_i, e_j \in \mathcal{E}$) represent an entity, a relation and a fact triple respectively, and $s(e) = \{n_e, d_e, a_e\} \in \mathcal{S}$ denotes the side information of entity e , i.e., entity name label, textual description and attribute value. We denote the set of one-hop neighbors of entity e as \mathcal{N}_e of size $|\mathcal{N}_e|$, namely, the entities that are directly connected to e in knowledge graph \mathcal{G} via fact triples, where $\mathcal{N}_e = \{e' \mid (e, r, e') \in \mathcal{T} \wedge (e, e' \in \mathcal{E})\} \cup \{e' \mid (e', r, e) \in \mathcal{T} \wedge (e, e' \in \mathcal{E})\}$.

¹Self-supervised learning is a form of unsupervised learning. Their distinction is informal in the existing literature. In this work, we do not distinguish between them in a strict sense.

Knowledge graphs are often separately constructed for various goals, and thus contain heterogeneous but complementary knowledge. Entity alignment is to identify entities from different knowledge graphs (different languages or sources) that describe the same real-world object, and can be formally defined as follows.

DEFINITION 2 (ENTITY ALIGNMENT). Given two different KGs \mathcal{G}_1 and \mathcal{G}_2 , EA is to learn a ranking function $f : \mathcal{E}_1 \times \mathcal{E}_2 \rightarrow \mathbb{R}$ to calculate the similarity score between two entities, based on which we rank the correctly aligned entity e_2 as high as possible among all entities of \mathcal{E}_2 with a queried entity $e_1 \in \mathcal{E}_1$. Pre-aligned entity pairs or named seed alignments $\mathcal{I} = \{(e_1, e_2) \mid e_1 \in \mathcal{E}_1, e_2 \in \mathcal{E}_2, e_1 \leftrightarrow e_2\}$ are provided, where \leftrightarrow means that the entity e_1 from \mathcal{G}_1 and the entity e_2 from \mathcal{G}_2 are equivalent.

According to the use of seed alignments \mathcal{I} , the EA task is classified into (semi-) supervised and self-supervised or unsupervised settings. The (semi-) supervised setting leverages part of \mathcal{I} as supervision signals for learning, while the self-supervised or unsupervised setting does not require any supervision.

4 THE PROPOSED APPROACH

Although supervised EA methods achieve SOTA performance, their dependence on supervision signals limits real-world EA applications. In contrast, self-supervised methods show the ability to obtain competitive performance without any supervision, and have much better scalability. To fully exploit the potential of self-supervised EA approaches, three crucial issues remain to be addressed.

- How to obtain the entity initial representation embedding efficiently and effectively by KG's side information?
- How to fully utilize and integrate the structural and semantic KG information in a self-supervised EA framework?
- How to establish a direct cross-KG information interaction channel for a self-supervised EA framework instead of only performing unidirectional intra-KG contrastive learning?

To address the above issues, we introduce the interactive contrastive learning for self-supervised EA framework, which contains three parts: **1. A KG semantic encoder** is utilized to encode the names, descriptions of entities and relations for providing efficient initial entity embeddings. **2. A relation-aware neighborhood aggregator** is introduced to fully exploit relations' structural and semantic information so as to update entity representation embeddings. **3. An interactive contrastive learning mechanism** is proposed to construct pseudo-aligned entity pairs as virtual pivots for guiding contrastive learning between two KGs, thus promoting the direct learning of cross-KG interactions. Figure 2 shows the overall framework of our proposed approach ICLEA.

4.1 KG Semantic Encoder

KG semantic encoder aims to fully capture the semantics of entities and relations by leveraging their names and descriptions. Since pre-trained language models (PLMs) have achieved remarkable progress in Natural Language Processing (NLP), we decide to leverage LaBSE [9], a language-agnostic pre-trained sentence model, for encoding the name. Besides, we choose SentenceTransformers², a

²<https://github.com/UKPLab/sentence-transformers>

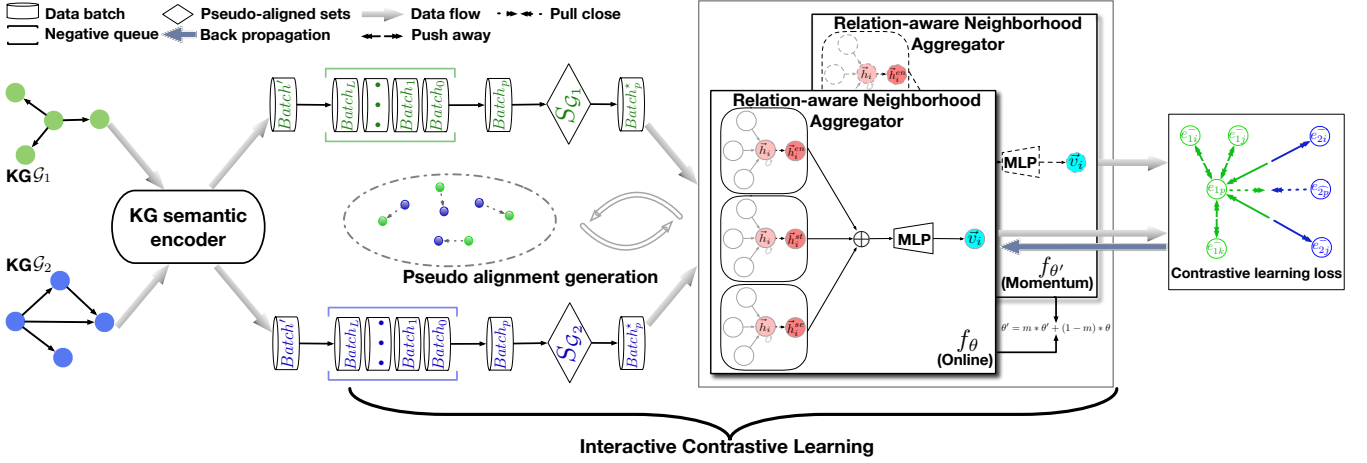


Figure 2: The overall framework of ICLEA. It consists of three main parts: (1) KG semantic encoder; (2) The relation-aware neighborhood aggregator; (3) Interactive contrastive learning mechanism.

transformer-based sentence embedding framework, to obtain entity description embeddings. LaBSE is a SOTA language-agnostic sentence embedding PLM which is trained on 109 different languages. SentenceTransformers is a Python framework for SOTA sentence, text and image embeddings. We choose them due to their outstanding empirical performance in capturing precise semantics for phrases and long sentences while helping us to cross the multilingual barrier.

KG Name Encoder. The names of each entity e and relation r are usually composed of phrases. We utilize the tokenizer of LaBSE to obtain the tokens of the entity name n_e and relation name n_r , and get corresponding embeddings through LaBSE model $f_{\theta_{LaBSE}}$. Next, a mean pooling operation is applied to the embedding of each token, followed by an L_2 normalization.

$$\vec{h}_{n_x} = \|\text{Mean}(f_{\theta_{LaBSE}}(n_x))\|_{L_2}, n_x \in \{n_e, n_r\}. \quad (1)$$

Entity Description Encoder. Moreover, entity description contains rich semantics that allows PLMs to encode with more context information. The entity description is usually composed of one or more sentences and contains multifaceted features related to the entity. For each description, we select the first L_{max} characters to feed into the SentenceTransformers model $f_{\theta_{ST}}$ and get the representation of entity description d_e ,

$$\vec{h}_{d_e} = \|\text{Mean}(f_{\theta_{ST}}(d_e))\|_{L_2}. \quad (2)$$

Finally, we get the entity e 's representation embedding \vec{h}_e by concatenating entity name embedding \vec{h}_{n_e} and entity description embedding \vec{h}_{d_e} , and more processing details are available in Section 5.1,

$$\vec{h}_e = \text{Concat}(\vec{h}_{n_e}, \vec{h}_{d_e}). \quad (3)$$

4.2 Relation-Aware Neighborhood Aggregator

Relation-aware neighborhood aggregator aims to update entity embeddings by performing message passing with the help of KG's

relation information which can aggregate neighborhood entities' helpful information for EA to the center node entity. Refer to Section 4.3, our self-supervised EA framework has the online and momentum parts as shown in Figure 2. These two parts have the same model structure, and they have different parameter updating strategies. We take the online part as an example to introduce this module. The KG's relations can bring us two aspects of crucial information: **structural** – neighbor entities, which provide valuable context information for understanding the center node entity, and **semantic** – neighbor relations, which capture rich semantic information of edges adjacent to the given center node entity. In this work, we apply GAT [42] and its variants as the backbone networks due to its effectiveness.

Structural Aggregator. The aggregation of neighbor entities is performed by considering both entities' importance and their relations to the center node. One vanilla GAT aggregates all neighbor entity embeddings for the center node to model the importance of different entities, which treats all relations equally, formally,

$$\vec{h}_i^{en} = \|\sum_{k=1}^K \sigma \left(\sum_{j \in \mathcal{N}_i \cup \{e_i\}} \alpha_{ij}^k \mathbf{W}^k \vec{h}_j \right)\|, \quad (4)$$

$$\alpha_{ij} = \frac{\exp \left(\sigma \left(\vec{\mathbf{q}}^T \left[\mathbf{W} \vec{h}_i \| \mathbf{W} \vec{h}_j \right] \right) \right)}{\sum_{k \in \mathcal{N}_i \cup \{e_i\}} \exp \left(\sigma \left(\vec{\mathbf{q}}^T \left[\mathbf{W} \vec{h}_i \| \mathbf{W} \vec{h}_k \right] \right) \right)},$$

where \mathbf{W} is the learnable linear transformation's weight matrix. The attention mechanism is a single-layer feedforward neural network parametrized by a weight vector $\vec{\mathbf{q}}$. K is the multi-head attention number. σ is a nonlinear activation function such as LeakyReLU. α_{ij} is the normalized attention coefficient. $(\cdot)^T$ represents matrix transposition and $\|$ is vector concatenation operation.

Another extended GAT is proposed to model the importance of different adjacent relations. It uses relation specific attention heads as relation-wise gates to control information flow from neighbor

entities and obtains updated entity embedding \vec{h}_i^{st} ,

$$\begin{aligned} \vec{h}_i^{st} &= \|\|_{k=1}^K \sigma \left(\sum_{j \in \mathcal{N}_i} \beta_{ij}^k \mathbf{W}^k \vec{h}_j \right), \\ \beta_{ij} &= \frac{\exp(Y_{ij})}{\sum_{k \in \mathcal{N}_i} \exp(Y_{ik})}, \\ Y_{ij} &= \sigma \left(\vec{r}_{ij} \mathbf{W}_{r_{ij}^1} + \mathbf{B}_{r_{ij}^1} \right) \mathbf{W}_{r_{ij}^2} + \mathbf{B}_{r_{ij}^2}, \end{aligned} \quad (5)$$

where \vec{r}_{ij} is trainable relation specific embedding between entity e_i and entity e_j . $\mathbf{W}_{r_{ij}}$ and $\mathbf{B}_{r_{ij}}$ are relation's transformation and bias weight matrix.

Semantic Aggregator. Relation's name label in KGs usually contains certain semantic information. For example, the "sisters"³ relation in DBpedia describes a kinship relation between the head and tail entities, we argue that information is helpful for obtaining better entity representation embeddings for self-supervised EA. The semantic aggregator aims to aggregate and fuse neighbor relations' semantic information into the center node entity representation embedding. In this work, we apply a one-hop GAT that is the same as Equation 4 to fuse neighbor relation name label embeddings into corresponding center entity embeddings. Specifically, for each \vec{h}_j in Equation 4, we replace it with a average neighbor relation embedding $\hat{h}_j = \frac{1}{K_{ij}} \sum_{x=0}^{K_{ij}} \vec{h}_{n_{r_x}}$, where n_{r_x} is the x -th relation's name label between entity e_i and entity e_j , K_{ij} is the total number of relations between entity e_i and entity e_j , and $\vec{h}_{n_{r_x}}$ is the corresponding relation name label embedding, which can be calculated from Equation 1. For \vec{h}_k in Equation 4, we replace it with \hat{h}_k using a similar method. This way, we can obtain \vec{h}_i^{se} of entity e_i based on its neighbor relation name labels.

Finally, we use a fully-connected layer to fuse three aspects of embeddings \vec{h}_i^{en} , \vec{h}_i^{st} , and \vec{h}_i^{se} to obtain the final entity representation embedding \vec{v}_i ,

$$\vec{v}_i = \text{MLP}(\text{Concat}(\vec{h}_i^{en}, \vec{h}_i^{st}, \vec{h}_i^{se})). \quad (6)$$

4.3 Interactive Contrastive Learning

Interactive contrastive learning mechanism is designed to learn direct cross-KG interactions for self-supervised EA. It mainly consists of three parts: (1) Momentum contrastive learning mechanism performs intra-KG contrastive learning for self-supervised EA, which samples negative entities and pushes them far away from the positive one in the same source KG, thus the aligned entity pairs in different KGs are relatively drawn close. (2) Negative sample queues store previously processed batches as negative samples for the positive batch, which can provide a large number of negative samples for the training process. (3) Interactive contrastive learning mechanism constructs pseudo-aligned pairs as virtual pivots during training and establishes cross-KG direct information interaction for self-supervised EA. It pulls the positive samples and pseudo-aligned entities closer, while pushing them far away from the negative samples in different KGs.

Momentum Contrastive Learning Mechanism. Given a training KG \mathcal{G}_1 's initial entity embeddings $\mathbf{H}_1 = \{\vec{h}_1, \vec{h}_2, \dots, \vec{h}_{|\mathcal{E}_1|}\}$

which can be calculated from Equation 3, self-supervised EA representation learning aims to learn an online embedding transformation function f_θ that maps \mathbf{H}_1 to $\mathbf{V}_1 = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_{|\mathcal{E}_1|}\}$ with $\vec{v}_x = f_\theta(\vec{h}_x)$, such that \vec{v}_x best describes \vec{h}_x . Instance-wise contrastive learning achieves this objective by optimizing a contrastive loss. Following SelfKG [25], we use the Noise Contrastive Estimation (NCE) loss [13]. In practice, we jointly optimize the NCE loss on both source KG \mathcal{G}_1 and target KG \mathcal{G}_2 , taking \mathcal{G}_1 as example, the NCE loss can be defined as,

$$\begin{aligned} \mathcal{L}_{\text{ncc}_{\mathcal{G}_1}} &= \sum_{x=1}^{|\mathcal{E}_1|} -\log \frac{\exp(\vec{v}_x \cdot \vec{v}'_x / \tau)}{\exp(\vec{v}_x \cdot \vec{v}'_x / \tau) + \sum_{k=1}^r \exp(\vec{v}_x \cdot \vec{v}'_k / \tau)} \\ &= \underbrace{\sum_{x=1}^{|\mathcal{E}_1|} -\frac{1}{\tau} \vec{v}_x \cdot \vec{v}'_x}_{\text{Alignment}} + \underbrace{\log(\exp(\vec{v}_x \cdot \vec{v}'_x / \tau) + \sum_{k=1}^r \exp(\vec{v}_x \cdot \vec{v}'_k / \tau))}_{\text{Uniformity}}, \end{aligned} \quad (7)$$

where, the "Alignment" part pulls the positive pair closer while the "Uniformity" part pushes the negative pairs away, \vec{v}_x and \vec{v}'_x are positive embeddings for entity e_x which can be obtained from the online encoder f_θ and momentum encoder $f_{\theta'}$, and \vec{v}'_k includes r negative samples' embeddings, and τ is a temperature hyperparameter. The online encoder's parameter θ is instantly updated with the backpropagation. The negative sample's embedding is obtained by feeding corresponding \vec{h}_x to the momentum encoder parameterized by θ' , $\vec{v}'_x = f_{\theta'}(\vec{h}_x)$, where θ' is a moving average of θ ,

$$\theta' \leftarrow m \times \theta' + (1 - m) \times \theta, m \in [0, 1]. \quad (8)$$

Negative Sample Queues. While performing contrastive learning for self-supervised EA, we need to sample a large number of negative entities from the same source KG to avoid the conflict by simply excluding the positive one. We maintain two negative queues for both KGs that store previously processed batches as negative samples. As shown in the middle part of Figure 2, when a new data batch $Batch'$ arrives, we add it to the corresponding negative queue tail, and the head $Batch_0$ in the same negative queue is dequeued as a positive sample batch $Batch_p$. In the early stage of training, we do not perform any parameter update until one of the negative queues reaches the predefined length $L + 1$, where L for the number of previous processed batches used as negative samples and "1" for the dequeued positive batch. Let the numbers of the entities in KGs $|\mathcal{E}_1|$, $|\mathcal{E}_2|$, the batch size B and L is constraint by,

$$(L + 1) \times B \leq \min(|\mathcal{E}_1|, |\mathcal{E}_2|). \quad (9)$$

Finally, the real number of negative samples used for each positive batch's entity is $(L + 1) \times B - 1$.

Interactive Contrastive Learning Mechanism. The interactive contrastive learning mechanism aims to build a direct information interaction channel for two KGs during training. We construct pseudo-aligned entity sets $S_{\mathcal{G}_1}$ and $S_{\mathcal{G}_2}$ for each source KG to the corresponding target KG at the beginning of each training epoch. Given initial embeddings of source and target KGs \mathcal{G}_1 and \mathcal{G}_2 , \mathbf{H}_1 and \mathbf{H}_2 , we feed them into the online encoder f_θ to obtain the corresponding entity embeddings \mathbf{V}_1 and \mathbf{V}_2 . For each entity e_{1p} in \mathcal{G}_1 and e_{2q} in \mathcal{G}_2 , we match the most similar entities $e_{2\hat{p}}$ and $e_{1\hat{q}}$ from

³<https://dbpedia.org/property/sisters>

Algorithm 1: Interactive Contrastive Learning Process.

Input: Online encoder f_θ , momentum encoder $f_{\theta'}$, training datasets X_1, X_2 of KGs $\mathcal{G}_1, \mathcal{G}_2$, distance threshold λ , momentum m .
 $\theta' = \theta$. // initialize momentum encoder as the online encoder
 $H_1, H_2 = \text{KG_Semantic_Encoder}(X_1, X_2)$ // get the initial entity embeddings for source KG \mathcal{G}_1 and target KG \mathcal{G}_2
 $NQ_1, NQ_2 = \emptyset$ // initialize negative sample queues for \mathcal{G}_1 and \mathcal{G}_2
while not MaxEpoch **do**
 // Get pseudo-aligned entities */
 $V_1 = f_\theta(H_1), V_2 = f_\theta(H_2)$ // get features for all data
 $Pair_{1 \rightarrow 2} = \text{Faiss.search}(V_1, V_2, 1, \lambda)$
 $Pair_{2 \rightarrow 1} = \text{Faiss.search}(V_2, V_1, 1, \lambda)$
 // Find the top 1 nearest neighbor with L2 distance less than λ for each entity
 $Pair_{all} = \text{Merge}(Pair_{1 \rightarrow 2}, Pair_{2 \rightarrow 1})$
 // Training */
for $Batch'$ in Dataloader($X_1 + X_2$) **do** // load a minibatch $Batch'$
if $Batch'$ belongs to X_1 **then**
 | $NQ_1.add_tail(Batch')$, $NQ = NQ_1$, $\mathcal{G} = \mathcal{G}_1$
else
 | $NQ_2.add_tail(Batch')$, $NQ = NQ_2$, $\mathcal{G} = \mathcal{G}_2$
if $NQ.Length == L + 1$ **then**
 | $Batch_p = NQ.dequeue_head()$ // the head batch of
 | corresponding negative sample queue dequeues as the positive batch
 | $\vec{v}_x = f_\theta(Batch_p), \vec{v}'_x = f_{\theta'}(Batch_p)$
 | $\vec{v}''_x = f_{\theta'}(Pair_{all}[Batch_p])$ // forward pass through the
 | online encoder and momentum encoder, $Batch_p$ as positive samples
 | $NV_{\mathcal{G}_1} = f_{\theta'}(NQ_1), NV_{\mathcal{G}_2} = f_{\theta'}(NQ_2)$
 | $Loss =$
 | $\mathcal{L}_{nce_{\mathcal{G}}}(\vec{v}_x, \vec{v}'_x, NV_{\mathcal{G}}) + \mathcal{L}_{icl_{\mathcal{G}}}(\vec{v}_x, \vec{v}''_x, NV_{\mathcal{G}_1}, NV_{\mathcal{G}_2})$
 | // calculate loss with Equations 7 and 10
 | $\theta = \text{Adam}(Loss, \theta)$ // update online encoder's parameters
 | $\theta' = m \times \theta' + (1 - m) \times \theta$ // update momentum encoder's parameters

the corresponding \mathcal{G}_2 and \mathcal{G}_1 . We predefine an L_2 distance threshold λ , if $\text{Dis}(e_{1p}, e_{2p}^-)$ or $\text{Dis}(e_{2q}, e_{1q}^-)$ is less than λ , we add the pair of entities $\langle e_{1p}, e_{2p}^- \rangle$ or $\langle e_{2q}, e_{1q}^- \rangle$ to the corresponding pseudo-aligned set $S_{\mathcal{G}_1}$ or $S_{\mathcal{G}_2}$. We use these pseudo-aligned entity pairs as virtual pivots to guide the entire contrastive learning process, thus establishing the cross-KG direct information interaction for self-supervised EA. We apply Faiss⁴ for obtaining pseudo-aligned sets efficiently. It is worth noting that distinct from the previous iterative bootstrapping strategy [37] that maintains high-confidence aligned entity pairs iteratively, our approach focuses on generating larger pseudo-aligned sets automatically, which liberates us from the caution of introducing additional noises. Although the pseudo-aligned sets we construct also contain some noises, our experimental results in the next section demonstrate that our model is less sensitive to those noises, and the advantages of the pseudo-aligned sets on self-supervised EA far outweigh the disadvantages of noises. To incorporate pseudo-aligned entity pairs into the contrastive learning process we introduce an additional NCE loss \mathcal{L}_{icl} into the training process, and in Equation 10 we take the example of the KG \mathcal{G}_1 , where $\vec{v}_x = f_\theta(\vec{h}_{1p}), \vec{v}'_x = f_{\theta'}(\vec{h}_{2p}^-), \vec{h}_{1p}$ and \vec{h}_{2p}^- are the embeddings of the corresponding positive entity and its pseudo-aligned entity. The hyperparameter β balances the sensitiveness of the model to

⁴<https://github.com/facebookresearch/faiss>

the negative sample \vec{v}'_{1_k} and \vec{v}'_{2_k} from different KGs,

$$\mathcal{L}_{icl_{\mathcal{G}_1}} = \sum_{x=1}^{|\mathcal{E}_1|} \beta \cdot \left[-\log \frac{\exp(\vec{v}_x \cdot \vec{v}'_x / \tau)}{\exp(\vec{v}_x \cdot \vec{v}'_x / \tau) + \sum_{k=1}^r \exp(\vec{v}_x \cdot \vec{v}'_{1_k} / \tau)} \right] + (1 - \beta) \cdot \left[-\log \frac{\exp(\vec{v}_x \cdot \vec{v}'_x / \tau)}{\exp(\vec{v}_x \cdot \vec{v}'_x / \tau) + \sum_{k=1}^r \exp(\vec{v}_x \cdot \vec{v}'_{2_k} / \tau)} \right]. \quad (10)$$

Based on the above discussion, we can get the overall optimization goal \mathcal{L} of ICLEA in Equation 11. A pseudo-code of our proposed algorithm is given in Algorithm 1,

$$\mathcal{L} = \sum_{i \in \{1, 2\}} \mathcal{L}_{nce_{\mathcal{G}_i}} + \mathcal{L}_{icl_{\mathcal{G}_i}}. \quad (11)$$

5 EXPERIMENTS

In this section, we evaluate our proposed approach on DBP15K, a widely used benchmark for EA. We first introduce the experimental settings, then report the overall results, and finally conduct ablation studies as well as parameter sensitivity analyses.

5.1 Experimental Settings

Dataset. The DBP15K dataset is originally built by [36], which includes three cross-lingual datasets extracted from DBpedia⁵. Each contains 15,000 reference alignments between English (EN) and one of the other languages, i.e., Chinese (ZH), Japanese (JA) and French (FR). Table 1 presents the detailed statistics about DBP15K. To make wide comparisons, we report the results on both original [36] and translated version [51], corresponding to two different experimental settings: multilingual setting and monolingual setting. In the multilingual setting, all the used data are extracted from original multilingual DBpedia without using any translation tools. Moreover, to make the comparisons fair, following previous works [48, 51], we use Google Translate to translate all non-English entity descriptions into English when using translated DBP15K in the monolingual setting.

Table 1: Statistics of the simplified DBP15K datasets.

Datasets		Ent.	Rel.	R-Tri.	Ent Alignments.
ZH_EN	ZH	19,388	1,700	70,414	15,000
	EN	19,572	1,322	95,142	15,000
JA_EN	JA	19,814	1,298	77,214	15,000
	EN	19,780	1,152	93,484	15,000
FR_EN	FR	19,661	902	105,998	15,000
	EN	19,993	1,207	115,722	15,000

Evaluation Metrics. We use Hits@N as the evaluation metric. Hits@N means the proportion of correct entities that rank no larger than N (N is 1 and 10), and higher Hits@N indicates better performance. We further calculate the average Hits@1 of three subtasks to measure the overall performance.

Baselines. We compare our ICLEA against two groups of baselines, i.e., semi-supervised and un/self-supervised EA approaches. The latter include MultiKE [59], EVA [24] and SelfKG [26], and the former are further divided into three branches: (1) KG structure

⁵<http://downloads.dbpedia.org/2016-04/>

Table 2: Overall results on DBP15K. Methods marked with “*” use a translated version of DBP15K in the monolingual setting. The best results in supervised/unsupervised or self-supervised settings are marked in underline/bold.

Model	DBP15K _{ZH_EN}		DBP15K _{JA_EN}		DBP15K _{FR_EN}		AVG
	Hits@1	Hits@10	Hits@1	Hits@10	Hits@1	Hits@10	Hits@1
Supervised							
MTransE	30.8	61.4	27.9	57.5	24.4	55.6	27.7
JAPE	41.2	74.5	36.3	68.5	32.4	66.7	36.6
IPTransE	40.6	73.5	36.7	69.3	33.3	68.5	36.9
SEA	42.4	79.6	38.5	78.3	40.0	79.7	40.3
KECG	47.8	83.5	49.0	84.4	48.6	85.1	48.5
MuGNN	49.4	84.4	50.1	85.7	49.5	87.0	49.7
RSNs	50.8	74.5	50.7	73.7	51.6	76.8	51.0
AliNet	53.9	82.6	54.9	83.1	55.2	85.2	54.7
BootEA	62.9	84.8	62.2	85.4	65.3	87.4	63.5
NAEA	65.0	86.7	64.1	87.3	67.3	89.4	65.5
MRPEA	68.1	86.7	65.5	85.9	67.7	89.0	67.1
PREA	71.5	92.9	71.3	93.3	73.9	94.6	72.2
GM-Align*	67.9	78.5	74.0	87.2	89.4	95.2	77.1
RDGCN*	70.8	84.6	76.7	89.5	88.6	95.7	78.7
HGCN*	72.0	85.7	76.6	89.7	89.2	96.1	79.3
AttrGNN*	79.6	92.9	78.3	92.1	91.9	97.8	83.3
DGMC*	80.1	87.5	84.8	89.7	93.3	96.0	86.1
RNM*	84.0	91.9	87.2	94.4	93.8	95.4	88.3
EPEA*	88.5	95.3	92.4	96.9	95.5	98.6	92.1
CEAFF	79.5	-	86.0	-	96.4	-	87.3
HMAN	87.1	98.7	93.5	99.4	97.3	<u>99.8</u>	92.6
BERT-INT	<u>96.8</u>	<u>99.0</u>	<u>96.4</u>	<u>99.1</u>	<u>99.2</u>	<u>99.8</u>	97.5
Unsupervised or Self-supervised							
MultiKE	50.9	57.6	39.3	48.9	63.9	71.2	51.4
EVA	75.2	89.5	73.7	89.0	73.1	90.9	74.0
SelfKG	74.5	86.6	81.6	91.3	95.7	99.2	84.0
SelfKG*	82.9	91.9	89.0	95.3	95.9	99.2	89.3
ICLEA	88.4	97.2	91.9	97.5	98.6	99.9	93.0
ICLEA*	92.1	98.1	95.5	98.8	99.2	99.9	95.6

embedding-based methods that only use the KGs’ structure information: MTransE [5], JAPE [36], IPTransE [60], SEA [32], KECG [20], MuGNN [2], RSNs [12], AliNet [38], BootEA [37], NAEA [62], MRPEA [34] and PREA [31]. (2) Graph-based methods that leverage GNNs to utilize various KG information: GM-Align[51], RDGCN [49], HGCN[50], AttrGNN [27], DGMC [10], RNM [63], EPEA [48] and CEAFF [57]. (3) BERT-based methods that directly use multilingual BERT to deal with EA: HMAN [52] and BERT-INT [41].

Data Preprocessing. The name labels of entity and relation in DBP15K usually have a string with meaningless prefixes, such as “http://dbpedia.org/resource/Jay_Chou”, we remove the meaningless prefixes and replace the underscores with empty spaces, producing the final entity and relation name labels. [41] provides the original multilingual description data of entities in DBP15K, we directly use these entity description data in the multilingual setting. In the monolingual setting, we leverage an open source Google Translate tool ⁶ to translate the non-English entity descriptions into the corresponding English versions. To obtain one-hop neighbors for each entity, we treat KGs as undirected graphs, that

means we use the relation triples in the datasets to find all entities connected to a given entity, regardless of the connection’s direction. **Model and Training Details.** Our model is implemented with Pytorch 1.7.0. We employ Adam as our optimizer with a learning rate $1e-6$ and gradually reduce it to maintain the stability of training. The number of training epochs is 300, the batch size is 64, momentum m is 0.9999, similarity threshold λ is 1.0, temperature τ is 0.08 and the size of negative queue is 32, the entity description length is 512. We use one-hop GAT in our model for its efficiency, during our experiments we find that using multi-hop GAT brings some performance degradation, which may be caused by the KGs’ heterogeneity and the amplified neighbor noises in the self-supervised EA setting, that is consistent with the findings of previous work [26]. We utilize the data about a center entity and its up to 15 neighboring entities as the input of GAT. In the relation-aware neighborhood aggregator layer, the dimension of both input and output embeddings is equal to $LaBSE_DIM + DESC_DIM$, and in the fully-connected layer, the input dimension is $LaBSE_DIM \times 7$ and the output dimension is $LaBSE_DIM \times 5$, where $LaBSE_DIM$ and $DESC_DIM$ are equal to 768. We choose L_2 distance as entity embedding distance metric for pseudo alignment generation and set β to 0.9 in interactive contrastive learning. We randomly select 5% from the original training set as the validation set in order to early stopping. All experiments are conducted on a Ubuntu server with eight GPUs (GeForceRTX 3090), each evaluation is repeated five times with the same random seed (set to 37) and the averaged results are reported.

5.2 Main Results

Table 2 lists the overall performance. For all baselines, we take the reported results from the original papers and [25]. From the results, we have the following key observations:

Comparisons with semi-supervised methods. ICLEA outperforms almost all previous supervised models on both original and translated DBP15K, and achieves results comparable to BERT-INT (the strongest baseline). They use multilingual BERT directly to help with entity alignment. ICLEA achieves the same excellent results as BERT-INT on the FR_EN subtask, where the BERT-INT’s Hits@1 is 0.6% higher than that of ICLEA, while the ICLEA’s Hits@10 is 0.1% higher than that of BERT-INT. The gap between ICLEA and BERT-INT is only 4.5% in average Hits@1, and it is further narrowed to 1.9% on the translated dataset (though BERT-INT does not translate dataset directly, it uses multilingual BERT). It is worth mentioning that BERT-INT actually fine-tunes the whole multilingual BERT, whose parameter scale is ten times larger than ICLEA (110M vs. 11M). In a nutshell, ICLEA significantly narrows the gap between supervised and self-supervised EA approaches.

Comparisons with un/self-supervised methods. In the unsupervised or self-supervised setting, ICLEA notably outperforms all those baselines with a large margin. Even compared with the strongest baseline SelfKG, ICLEA improves the average Hits@1 score by 9.0% on the original datasets in the multilingual setting and 6.3% on the translated datasets in the monolingual setting. Note that the improvement is more significant in original datasets in the multilingual setting, especially in ZH_EN subtask, indicating that ICLEA is less language dependent and suitable for dealing with the cross-lingual entity alignment tasks.

⁶<https://github.com/ssut/py-googletrans>

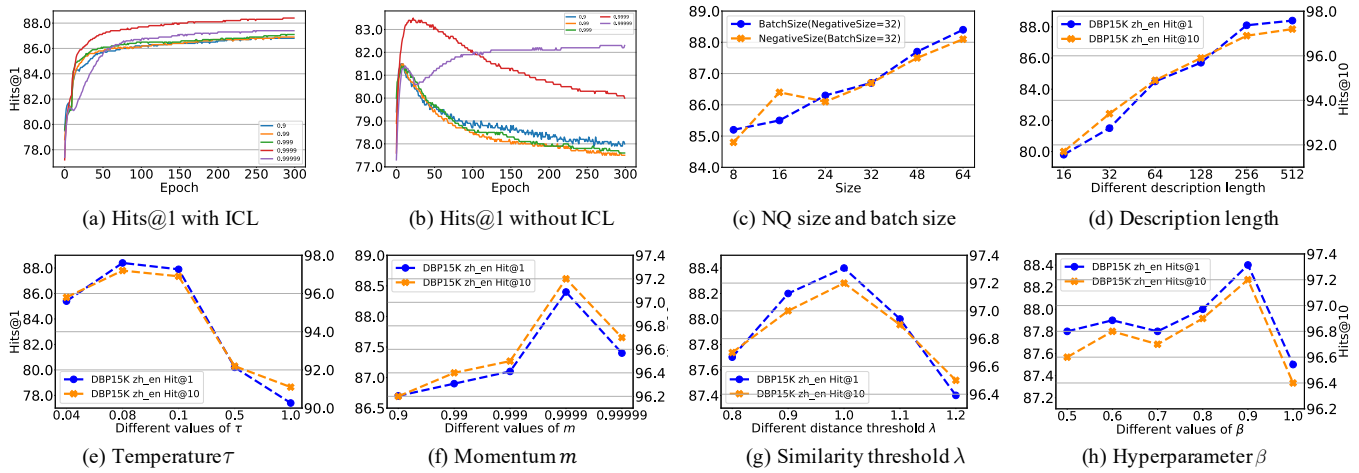


Figure 3: Study on negative queue (NQ) size, batch size, description length, temperature τ , momentum m , L_2 distance threshold λ and hyperparameter β on DBP15K_{ZH_EN}. (a), (b) present the test Hits@1 curve throughout the training with and without ICL.

Table 3: Ablation study of ICLEA on DBP15K. “” means using translated version of DBP15K in the monolingual setting.**

Model	DBP15K _{ZH_EN}		DBP15K _{JA_EN}		DBP15K _{FR_EN}		DBP15K _{ZH_EN} *		DBP15K _{JA_EN} *		DBP15K _{FR_EN} *	
	Hits@1	Hits@10	Hits@1	Hits@10	Hits@1	Hits@10	Hits@1	Hits@10	Hits@1	Hits@10	Hits@1	Hits@10
ICLEA	88.4	97.2	91.9	97.5	98.6	99.9	92.1	98.1	95.5	98.8	99.2	99.9
w/o Nam.	72.0	92.7	74.1	94.1	89.5	99.3	75.8	94.0	82.0	97.0	89.2	99.1
w/o Des.	80.4	91.4	87.3	93.1	97.3	99.5	86.5	93.4	92.4	96.6	97.1	99.4
w/o ICL	84.3	94.1	88.8	96.2	97.1	99.8	89.3	96.3	93.3	98.1	97.3	99.8
w/o MCL	86.7	96.5	90.2	97.1	97.6	99.7	91.4	97.9	94.2	98.3	97.9	99.8
w/o Rel.	87.0	96.7	90.6	97.2	97.8	99.9	90.5	97.3	94.9	98.5	98.1	99.8

Multilingual bias phenomenon. ICLEA exhibits multilingual bias consistent with most previous methods, i.e., performing best on the FR_EN subtask and worst on the ZH_EN subtask. Early models that rely on only KG’s structure information to handle EA e.g. MTransE, JAPE, IPTransE, SEA, KECG, etc., do not exhibit such multilingual bias phenomenon. This may be due to the fact that as EA evolves, more and more methods incorporate entity-related semantic information in the model to assist EA, while leading to the multilingual bias phenomenon described earlier. Note that EVA has a relatively balanced performance on three datasets due to the introduction of language-independent entity images. This inspires us that multi-modal entity information utilization in self-supervised EA would be a promising future direction.

5.3 Ablation Study

We perform an ablation study to evaluate the effectiveness of all model components. Accordingly, we implement five variants of ICLEA by removing entity name information (*w/o Nam.*), entity description information (*w/o Des.*), relation-aware neighborhood aggregator (*w/o Rel.*), momentum contrastive learning mechanism (*w/o MCL*) and interactive contrastive learning mechanism (*w/o ICL*). Table 3 presents the results, which show that the removal of

each component has a negative impact on the performance. From the results, we have the following key observations:

The entity name information has the greatest impact on the model’s accuracy. Specifically, *w/o Nam.* has the largest negative impact on Hits@1 value, demonstrating the entity name information plays a fundamental role in the model’s accuracy. Intuitively, the entity description information contains richer semantic information than entity name labels and should play a more critical role in the EA model’s accuracy. However, the experimental results are inconsistent with the intuition, and we believe this may be due to the heterogeneity of entity descriptions in different KGs. There may be differences in the focus, order, etc. of different KGs for the same entity’s description, so that noise is introduced along with entity semantics, which may be the reason for this phenomenon. We believe how to solve the entity description’s heterogeneity problem and better incorporate description information in self-supervised EA would be a valuable open future problem to explore.

The entity description information has the greatest impact on the model’s Hits@10. As shown in Table 3, in most cases, *w/o Des.* has the largest negative impact on the Hits@10 value, demonstrating the effectiveness of entity description. The Hits@10 value responds to a certain extent to the recall ability of the model,

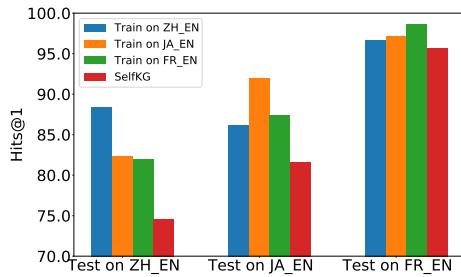


Figure 4: Inductive learning of ICLEA in multilingual setting.

which indicates that entity description information plays a crucial role in improving the model’s recall.

The relation information can also improve the model’s accuracy. *w/o Rel.* causes relatively more significant decreases in Hits@1 than Hits@10, indicating the importance of relation and structure in distinguishing more ambiguous cases.

In addition to the high-performance improvement, the interactive contrastive learning mechanism also brings more stable model training process. *w/o ICL* brings a more notable performance degradation, illustrating the importance of the interactive contrastive learning mechanism for self-supervised EA. Besides, we further investigate how the performance changes during the model training epochs with and without ICL. As shown in Figure 3a and 3b, without ICL, the test Hits@1 score usually reaches the peak after tens of epochs and then shows a sharp drop, while using ICL brings a more stable model training process.

5.4 Hyper-parameters Analyses

Negative queue size and batch size. For these two parameters, we perform a grid search from 8 to 64. Figure 3c shows that the performance exhibits a fluctuating upward trend when fixing batch size to 64 and increasing the size of the negative queue. Meanwhile, Figure 3c also shows the performance increases steadily with larger batch size when setting queue size to 64. While both expanding negative queue and batch size can improve the performance, a larger batch size usually brings more computational cost.

Entity description length. As shown in Figure 3d, a longer description brings better performance because it can provide more semantic information. According to our statistical results, the description length of most entities in original DBP15K is less than 512. Considering the computation cost and the description length distribution in the datasets, we set the maximal length to 512.

Temperature τ and momentum coefficient m . The temperature hyperparameter τ regulates the degree of attention to difficult samples [44] and the momentum coefficient m prevents model’s sensitive update [16]. We empirically choose both parameters from finite sets and present the results in Figure 3e and Figure 3f, which show that a relatively larger m (e.g., 0.9999) leads to better performance and $\tau = 0.08$ is a good cut-off point.

Distance threshold λ and hyperparameter β . Distance threshold λ in Equation 10 controls the selection of pseudo alignments. A larger λ means more pseudo alignments for cross-KG interaction, while possibly introducing more noises. We find in our experiments

that our model is less sensitive to these noises, and a proper proportion of noise can even bring some help to the model’s performance. As shown in Figure 3g, it is a good balance point when $\lambda = 1.0$. It will be an interesting research problem to design a dynamic adjustment mechanism for λ to further balance the effect of noise on the model. The hyperparameter β is the trade-off factor between the positive and negative samples from different KGs. As shown in Figure 3h, we find that the sensitivity of the model is more preferred to the source KG’s negative sample queue, and thus set β to 0.9.

5.5 Inductive Learning

We train ICLEA on one dataset of original DBP15K, then directly transfer and evaluate them on the other two datasets which can evaluate our model’s cross-lingual transference ability, and show the results in Figure 4. When training ICLEA on FR_EN and testing on ZH_EN/JA_EN, the Hits@1 values show a significant decrease, with relative reductions of 7.4% and 4.9%, indicating that the cross-lingual transference from French to Chinese and Japanese is extremely challenging for the model. When training ICLEA on ZH_EN/JA_EN and testing on JA_EN/ZH_EN, the Hits@1 values decrease by 6.3% and 6.9%, respectively, which reveals Chinese and Japanese are almost the same difficult to be transferred between each other. No matter training on any dataset, the test performance on FR_EN is comparable to the normal one (i.e., train and test on FR_EN), this is very likely due to the fact that French and English are more similar to each other, making the alignment avoids the affect of different languages. We also find that no matter training on ZH_EN, JA_EN, or FR_EN, ICLEA’s test performance on the other two datasets outperforms SelfKG by a large margin (Hits@1 values increased by at least 9.9%, 5.5%, and 3.1%, respectively), which demonstrates ICLEA is better at overcoming the multilingual challenge in self-supervised entity alignment.

6 CONCLUSION AND FUTURE WORK

In this work, we propose a model named ICLEA for self-supervised EA. To better utilize the entity’s structural and semantic information, we separately encode entity names and descriptions with the help of PLMs and propose a relation-aware neighborhood aggregator to better leverage the structural and semantic information brought by KGs’ relations. We design an innovative interactive contrastive learning mechanism by constructing pseudo-aligned entity pairs to establish a direct information interaction channel for the two KGs. Experimental results show that ICLEA performs on par with previous SOTA supervised counterparts and outperforms previous best self-supervised results by a large margin. We also present two promising directions for self-supervised EA. (1) Aligned entities usually have similar attribute-value pairs in different KGs. How to combine entity attributes into a self-supervised EA framework will be a focus of our future work. (2) PCL [22] introduces prototypes as latent variables to help find the maximum-likelihood estimation of model’s parameters in an Expectation-Maximization framework. KG entities also have potential prototype information, such as people, buildings, etc. How to let the model learn this information and assist in self-supervised EA is also an open challenge.

7 ACKNOWLEDGEMENT

This work is supported by the National Key Research and Development Program of China (2020AAA0106501), the NSFC Youth Project (61825602), grants from Joint Institute of Tsinghua University–China Mobile Communications Group Co.,Ltd., BAAI and the Institute for Guo Qiang, Tsinghua University (2019GQB0003).

REFERENCES

- [1] Joan Bruna, Wojciech Zaremba, Arthur Szlam, and Yann LeCun. 2014. Spectral networks and deep locally connected networks on graphs. In *2nd International Conference on Learning Representations, ICLR 2014*.
- [2] Yixin Cao, Zhiyuan Liu, Chengjiang Li, Juanzi Li, and Tat-Seng Chua. 2019. Multi-Channel Graph Neural Network for Entity Alignment. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 1452–1461.
- [3] Yixin Cao, Xiang Wang, Xiangnan He, Zikun Hu, and Tat-Seng Chua. 2019. Unifying knowledge graph learning and recommendation: Towards a better understanding of user preferences. In *The world wide web conference*. 151–161.
- [4] Chong Chen, Min Zhang, Weizhi Ma, Yiqun Liu, and Shaoping Ma. 2020. Jointly non-sampling learning for knowledge graph enhanced recommendation. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 189–198.
- [5] Muhao Chen, Yingtao Tian, Mohan Yang, and Carlo Zaniolo. 2017. Multilingual Knowledge Graph Embeddings for Cross-lingual Knowledge Alignment. <https://doi.org/10.24963/ijcai.2017/209>
- [6] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. 2020. A simple framework for contrastive learning of visual representations. In *ICML*. PMLR, 1597–1607.
- [7] Wanyun Cui, Yanghua Xiao, Haixun Wang, Yangqiu Song, Seung-won Hwang, and Wei Wang. 2017. KBQA: Learning Question Answering over QA Corpora and Knowledge Bases. *Proceedings of the VLDB Endowment* 10, 5 (2017).
- [8] Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. 2016. Convolutional neural networks on graphs with fast localized spectral filtering. *Advances in neural information processing systems* 29 (2016), 3844–3852.
- [9] Fangxiaoyu Feng, Yinfei Yang, Daniel Cer, Naveen Arivazhagan, and Wei Wang. 2022. Language-agnostic BERT Sentence Embedding. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 878–891.
- [10] Matthias Fey, Jan E Lenssen, Christopher Morris, Jonathan Masci, and Nils M Kriege. 2020. Deep graph matching consensus. *arXiv preprint arXiv:2001.09621* (2020).
- [11] Zuohui Fu, Yikun Xian, Ruoyuan Gao, Jieyu Zhao, Qiaoying Huang, Yingqiang Ge, Shuyuan Xu, Shijie Geng, Chirag Shah, Yongfeng Zhang, et al. 2020. Fairness-aware explainable recommendation over knowledge graphs. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 69–78.
- [12] Lingbing Guo, Zequn Sun, and Wei Hu. 2019. Learning to exploit long-term relational dependencies in knowledge graphs. In *ICML*. PMLR, 2505–2514.
- [13] Michael Gutmann and Aapo Hyvärinen. 2010. Noise-contrastive estimation: A new estimation principle for unnormalized statistical models. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*. JMLR Workshop and Conference Proceedings, 297–304.
- [14] Xu Han, Zhiyuan Liu, and Maosong Sun. 2018. Neural knowledge acquisition via mutual attention between knowledge graph and text. In *AAAI*.
- [15] Xianpei Han and Jun Zhao. 2009. Named entity disambiguation by leveraging wikipedia semantic knowledge. In *Proceedings of the 18th ACM conference on Information and knowledge management*. 215–224.
- [16] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. 2020. Momentum contrast for unsupervised visual representation learning. In *CVPR*. 9729–9738.
- [17] Longlong Jing and Yingli Tian. 2020. Self-supervised visual feature learning with deep neural networks: A survey. *IEEE transactions on pattern analysis and machine intelligence* (2020).
- [18] Jacob Devlin Ming-Wei Chang Kenton and Lee Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT*. 4171–4186.
- [19] Jens Lehmann, Robert Isele, and Jakob et al. 2015. DBpedia—a large-scale, multilingual knowledge base extracted from Wikipedia. *Semantic web* 6, 2 (2015), 167–195.
- [20] 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*. 2723–2732.
- [21] Juanzi Li, Jie Tang, Yi Li, and Qiong Luo. 2008. Rimom: A dynamic multistrategy ontology alignment framework. *TKDE* 21, 8 (2008), 1218–1232.
- [22] Junnan Li, Pan Zhou, Caiming Xiong, and Steven Hoi. 2020. Prototypical Contrastive Learning of Unsupervised Representations. In *International Conference on Learning Representations*.
- [23] Bing Liu, Harrison Scells, Guido Zuccon, Wen Hua, and Genghong Zhao. 2021. ActiveEA: Active Learning for Neural Entity Alignment. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. 3364–3374.
- [24] Fangyu Liu, Muhao Chen, Dan Roth, and Nigel Collier. 2021. Visual Pivoting for (Unsupervised) Entity Alignment. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*.
- [25] Xiao Liu, Haoyun Hong, Xinghao Wang, Zeyi Chen, Evgeny Kharlamov, Yuxiao Dong, and Jie Tang. 2022. SelfKG: Self-Supervised Entity Alignment in Knowledge Graphs. In *Proceedings of the ACM Web Conference 2022 (Virtual Event, Lyon, France) (WWW '22)*. Association for Computing Machinery, New York, NY, USA, 860–870. <https://doi.org/10.1145/3485447.3511945>
- [26] Xiao Liu, Fanjin Zhang, Zhenyu Hou, Li Mian, Zhaoyu Wang, Jing Zhang, and Jie Tang. 2021. Self-supervised learning: Generative or contrastive. *IEEE Transactions on Knowledge and Data Engineering* (2021).
- [27] Zhiyuan Liu, Yixin Cao, Liangming Pan, Juanzi Li, and Tat-Seng Chua. 2020. Exploring and Evaluating Attributes, Values, and Structure for Entity Alignment. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 6355–6364.
- [28] Farzaneh Mahdisoltani, Joanna Biega, and Fabian Suchanek. 2014. Yago3: A knowledge base from multilingual wikipedias. In *7th biennial conference on innovative data systems research*. CIDR Conference.
- [29] Xin Mao, Wenting Wang, Yuanbin Wu, and Man Lan. 2021. Are Negative Samples Necessary in Entity Alignment? An Approach with High Performance, Scalability and Robustness. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. 1263–1273.
- [30] 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 *Proceedings of the 13th International Conference on Web Search and Data Mining*. 420–428.
- [31] Xin Mao, Wenting Wang, Huimin Xu, Yuanbin Wu, and Man Lan. 2020. Relational reflection entity alignment. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. 1095–1104.
- [32] 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*. 3130–3136.
- [33] Shichao Pei, Lu Yu, Guoxian Yu, and Xiangliang Zhang. 2020. Rea: Robust cross-lingual entity alignment between knowledge graphs. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2175–2184.
- [34] Xiaofei Shi and Yanghua Xiao. 2019. Modeling multi-mapping relations for precise cross-lingual entity alignment. In *EMNLP*. 813–822.
- [35] M Sun, H Zhu, R Xie, and Z Liu. 2017. Iterative Entity Alignment Via Joint Knowledge Embeddings [C]. In *International Joint Conference on Artificial Intelligence*. AAAI Press.
- [36] Zequn Sun, Wei Hu, and Chengkai Li. 2017. Cross-Lingual Entity Alignment via Joint Attribute-Preserving Embedding. 628–644. https://doi.org/10.1007/978-3-319-68288-4_37
- [37] Zequn Sun, Wei Hu, Qingheng Zhang, and Yuzhong Qu. 2018. Bootstrapping Entity Alignment with Knowledge Graph Embedding. In *IJCAI*, vol.18. 4396–4402.
- [38] 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 *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 222–229.
- [39] Zequn Sun, Qingheng Zhang, Wei Hu, Chengming Wang, Muhao Chen, Farahnaz Akrami, and Chengkai Li. 2020. A Benchmarking Study of Embedding-based Entity Alignment for Knowledge Graphs. *Proceedings of the VLDB Endowment* 13, 11 (2020).
- [40] Jie Tang, Juanzi Li, Bangyong Liang, Xiaotong Huang, Yi Li, and Kehong Wang. 2006. Using Bayesian decision for ontology mapping. *Journal of web semantics* 4, 4 (2006), 243–262.
- [41] Xiaobin Tang, Jing Zhang, Bo Chen, Yang Yang, Hong Chen, and Cuiping Li. 2020. BERT-INT: A BERT-based Interaction Model For Knowledge Graph Alignment. In *IJCAI*. 3174–3180.
- [42] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2018. Graph Attention Networks. In *International Conference on Learning Representations*.
- [43] Vrandečić. 2014. Wikidata: a free collaborative knowledgebase. *Commun. ACM* 57, 10 (2014), 78–85.
- [44] Feng Wang and Huaping Liu. 2021. Understanding the behaviour of contrastive loss. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2495–2504.
- [45] Zhichun Wang, Juanzi Li, and Jie Tang. 2013. Boosting cross-lingual knowledge linking via concept annotation. In *Twenty-Third International Joint Conference on Artificial Intelligence*.

- [46] Zhichun Wang, Qingsong Lv, Xiaohan Lan, and Yu Zhang. 2018. Cross-lingual Knowledge Graph Alignment via Graph Convolutional Networks. 349–357. <https://doi.org/10.18653/v1/D18-1032>
- [47] Zhichun Wang, Qingsong Lv, Xiaohan Lan, and Yu Zhang. 2018. Cross-lingual knowledge graph alignment via graph convolutional networks. In *EMNLP*. 349–357.
- [48] Zhichun Wang, Jinjian Yang, and Xiaoju Ye. 2020. Knowledge graph alignment with entity-pair embedding. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 1672–1680.
- [49] Yuting Wu, Xiao Liu, Yansong Feng, Zheng Wang, Rui Yan, and Dongyan Zhao. 2019. Relation-Aware Entity Alignment for Heterogeneous Knowledge Graphs. *IJCAI*.
- [50] Yuting Wu, Xiao Liu, Yansong Feng, Zheng Wang, and Dongyan Zhao. 2019. Jointly Learning Entity and Relation Representations for Entity Alignment. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. 240–249.
- [51] 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 *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 3156–3161.
- [52] Hsiu-Wei Yang, Yanyan Zou, Peng Shi, Wei Lu, Jimmy Lin, and Xu Sun. 2019. Aligning Cross-Lingual Entities with Multi-Aspect Information. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. 4431–4441.
- [53] Donghan Yu, Yiming Yang, Ruohong Zhang, and Yuexin Wu. 2021. Generalized multi-relational graph convolution network. In *The World Wide Web Conference*.
- [54] Kaisheng Zeng, Chengjiang Li, Lei Hou, Juanzi Li, and Ling Feng. 2021. A comprehensive survey of entity alignment for knowledge graphs. *AI Open* 2 (2021), 1–13.
- [55] Kaisheng Zeng, Chengjiang Li, Yan Qi, Xin Lv, Lei Hou, Guozheng Peng, Juanzi Li, and Ling Feng. 2021. Encoding the Meaning Triangle (Object, Entity, and Concept) as the Semantic Foundation for Entity Alignment. In *International Conference on Web Information Systems Engineering*. Springer, 227–241.
- [56] Weixin Zeng, Xiang Zhao, Jiuyang Tang, and Changjun Fan. 2021. Reinforced Active Entity Alignment. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. 2477–2486.
- [57] Weixin Zeng, Xiang Zhao, Jiuyang Tang, and Xuemin Lin. 2019. Collective Embedding-based Entity Alignment via Adaptive Features.
- [58] Weixin Zeng, Xiang Zhao, Wei Wang, Jiuyang Tang, and Zhen Tan. 2020. Degree-aware alignment for entities in tail. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 811–820.
- [59] Qingheng Zhang, Zequn Sun, Wei Hu, Muhao Chen, Lingbing Guo, and Yuzhong Qu. 2019. Multi-view Knowledge Graph Embedding for Entity Alignment. In *IJCAI*.
- [60] Hao Zhu, Ruobing Xie, Zhiyuan Liu, and Maosong Sun. 2017. Iterative Entity Alignment via Joint Knowledge Embeddings. In *IJCAI*, Vol. 17. 4258–4264.
- [61] Qi Zhu, Hao Wei, Bunyamin Sisman, Da Zheng, Christos Faloutsos, Xin Luna Dong, and Jiawei Han. 2020. Collective multi-type entity alignment between knowledge graphs. In *Proceedings of The Web Conference 2020*. 2241–2252.
- [62] Qiannan Zhu, Xiaofei Zhou, Jia Wu, Jianlong Tan, and Li Guo. 2019. Neighborhood-Aware Attentional Representation for Multilingual Knowledge Graphs. In *IJCAI*. 1943–1949.
- [63] Yao Zhu, Hongzhi Liu, Zhonghai Wu, and Yingpeng Du. 2021. Relation-Aware Neighborhood Matching Model for Entity Alignment. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 4749–4756.