Multilingual Knowledge Graph Completion via Ensemble Knowledge Transfer

Xuelu Chen[1], Muhao Chen[1,2,3], Changjun Fan[4*],
Ankith Uppunda[1], Yizhou Sun[1] & Carlo Zaniolo[1]

[1]Department of Computer Science, UCLA, USA
[2]Information Sciences Institute, USC, USA
[3]Department of Computer and Information Science, UPenn, USA
[4]College of Systems Engineering, NUDT, China
shirleychen@ucla.edu; muhaoche@usc.edu;
fanchangjun@nudt.edu.cn; {auppunda, yzsun, zaniolo}@cs.ucla.edu

Abstract

Predicting missing facts in a knowledge graph (KG) is a crucial task in knowledge base construction and reasoning, and it has been the subject of much research in recent works using KG embeddings. While existing KG embedding approaches mainly learn and predict facts within a single KG, a more plausible solution would benefit from the knowledge in multiple language-specific KGs, considering that different KGs have their own strengths and limitations on data quality and coverage. This is quite challenging, since the transfer of knowledge among multiple independently maintained KGs is often hindered by the insufficiency of alignment information and the inconsistency of described facts. In this paper, we propose KEnS, a novel framework for embedding learning and ensemble knowledge transfer across a number of language-specific KGs. KEnS embeds all KGs in a shared embedding space, where the association of entities is captured based on self-learning. Then, KEnS performs ensemble inference to combine prediction results from embeddings of multiple language-specific KGs, for which multiple ensemble techniques are investigated. Experiments on five real-world language-specific KGs show that KEnS consistently improves state-of-the-art methods on KG completion, via effectively identifying and leveraging complementary knowledge.

1 Introduction

Knowledge graphs (KGs) store structured representations of real-world entities and relations, constituting actionable knowledge that is crucial to various knowledge-driven applications (Koncel-Kedziorski et al., 2019; Chen et al., 2018a; Bordes et al., 2014). Recently, extensive efforts have been invested in KG embedding models, which encode entities as low-dimensional vectors and capture relations as algebraic operations on entity vectors. These models provide a beneficial tool to complete KGs by discovering previously unknown knowledge from latent representations of observed facts. Representative models including translational models (Bordes et al., 2013; Wang et al., 2014) and bilinear models (Yang et al., 2015; Trouillon et al., 2016) have achieved satisfactory performance in predicting missing facts.

Existing methods mainly investigate KG completion within a single monolingual KG. As different language-specific KGs have their own strengths and limitations on data quality and coverage, we investigate a more natural solution, which seeks to combine embedding models of multiple KGs in an ensemble-like manner. This approach offers several potential benefits. First, embedding models of well-populated KGs (e.g. English KGs) are expected to capture richer knowledge because of better data quality and denser graph structures. Therefore, they would provide ampler signals to facilitate inferring missing facts on sparser KGs. Second, combining the embeddings allows exchanging complementary knowledge across different language-specific KGs. This provides a versatile way of leveraging specific knowledge that is better known in some KGs than the others. For example, consider the facts about the oldest Japanese novel *The Tale of Genji*. English DBpedia (Lehmann et al., 2015) only records its genre as *Monogatari* (story), whereas Japanese DBpedia identifies more genres, including *Love Story*, *Royal Family Related Story*, *Monogatari* and *Literature-Novel*. Similarly, it is reasonable to expect a Japanese KG embedding model to offer significant advantages in inferring knowledge about other Japanese cultural entities such as *Nintendo* and *Mount Fuji*. Moreover, ensemble inference provides a mechanism to assess the credibility of different knowledge sources and thus leads to a
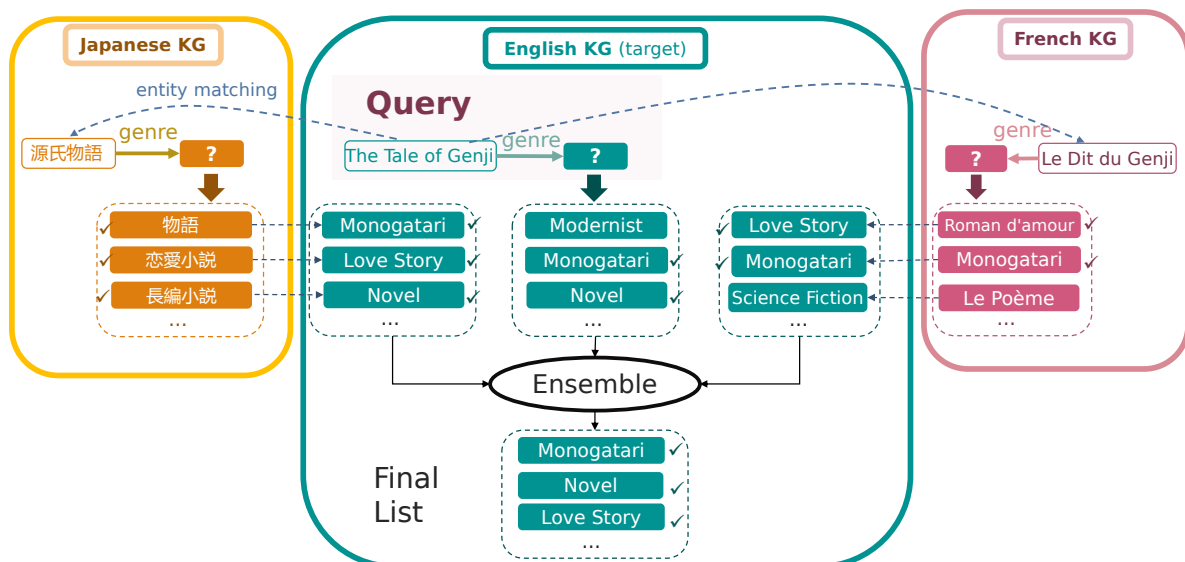

Figure 1: A depiction of the ensemble inference process answering the query *(The Tale of Genji, genre, ?t)* with multiple language-specific KG embeddings. Ground truth answers are marked✓. *Monogatari* is a traditional Japanese literary form.

more accurate final prediction.

Despite the potential benefits, combining predictions from multiple KG embeddings represents a non-trivial technical challenge. On the one hand, knowledge transfer across different embeddings is hindered by the lack of reliable alignment information that bridges different KGs. Recent works on multilingual KG embeddings provide support for automated entity matching (Chen et al., 2017, 2018b; Sun et al., 2018, 2020a). However, the performance of the state-of-the-art (SOTA) entity matching methods is still far from perfect (Sun et al., 2020a), which may cause erroneous knowledge transfer between two KGs. On the other hand, independently extracted and maintained language-specific KGs may inconsistently describe some facts, therefore causing different KG embeddings to give inconsistent predictions and raising a challenge to identifying the trustable sources. For instance, while the English DBpedia strictly distinguishes the network of a TV series (e.g. BBC) from its channel (e.g. BBC One) with two separate relations, i.e., `network` and `channel`, the Greek DBpedia only uses `channel` to represent all of those. Another example of inconsistent information is that Chinese DBpedia labels the birth place of the ancient Chinese poet *Li Bai* as *Sichuan, China*, which is mistakenly recorded as *Chuy, Kyrgyz* in English DBpedia. Due to the rather independent extraction process of each KG, such inconsistencies are inevitable, calling upon a reliable approach to identify credible knowledge among various sources.

In this paper, we propose KEnS (Knowledge Ensemble), which, to the best of our knowledge, is the first ensemble framework of KG embedding models. Fig. 1 gives a depiction showing the ensemble inference process of KEnS. KEnS seeks to improve KG completion in a multilingual setting, by combining predictions from embedding models of multiple language-specific KGs and identifying the most probable answers from those prediction results that are not necessarily consistent. Experiments on five real-world language-specific KGs show that KEnS significantly improves SOTA fact prediction methods that solely rely on a single KG embedding. We also provide detailed case studies to interpret how a sparse, low-resource KG can benefit from embeddings of other KGs, and how exclusive knowledge in one KG can be broadcasted to others.

2 Related Work

We hereby discuss three lines of work that are closely related to this topic.

Monolingual KG Embeddings. Monolingual KG embedding models embed entities and relations in a low-dimensional vector space and measure triple plausibility using these vectors. Translational models assess the plausibility of a triple (h, r, t) by the distance between two entity vectors $\boldsymbol{h}$ and $\boldsymbol{t}$, after applying a relation-specific translation vector $\boldsymbol{r}$. The representative models include TransE (Bordes et al., 2013) and its extensions TransD (Ji

3228

et al., 2015). Despite their simplicity, translational models achieve satisfactory performance on KG completion and are robust against the sparsity of data (Hao et al., 2019). RotatE (Sun et al., 2019b) employs a complex embedding space and models the relation r as the rotation instead of translation of the complex vector h toward t, which leads to the SOTA performance on KG embedding. There are also various methods falling into the groups of Bilinear models such as RESCAL (Nickel et al., 2011) and DistMult (Yang et al., 2015), as well as neural models like HolE (Nickel et al., 2016) and ConvE (Dettmers et al., 2018). Due to the large body of work in this line of research, we only provide a highly selective summary here. Interested readers are referred to recent surveys (Wang et al., 2017; Ji et al., 2020) for more information.

Multilingual KG Embeddings. Recent studies have extended embedding models to bridge multiple KGs, typically for KGs of multiple languages. MTransE (Chen et al., 2017) jointly learns a transformation across two separate translational embedding spaces along with the KG structures. BootEA (Sun et al., 2018) introduces a bootstrapping approach to iteratively propose new alignment labels to enhance the performance. MuGNN (Cao et al., 2019) encodes KGs via multi-channel Graph Neural Network to reconcile the structural differences. Some others also leverage side information to enhance the alignment performance, including entity descriptions (Chen et al., 2018b; Zhang et al., 2019), attributes (Trsedya et al., 2019; Sun et al., 2017; Yang et al., 2019), neighborhood information (Wang et al., 2018; Yang et al., 2015; Li et al., 2019; Sun et al., 2019a, 2020a) and degree centrality measures (Pei et al., 2019). A systematic summary of relevant approaches is given in a recent survey by Sun et al. (2020b). Although these approaches focus on the KG alignment that is different from the problem we tackle here, such techniques can be leveraged to support entity matching between KGs, which is a key component of our framework.

Ensemble methods. Ensemble learning has been widely used to improve machine learning results by combining multiple models on the same task. Representative approaches include voting, bagging (Breiman, 1996), stacking (Wolpert, 1992) and boosting (Freund and Schapire, 1997). Boosting methods seek to combine multiple weak models into a single strong model, particularly by learning model weights from the sample distribution.

Representative methods include AdaBoost (Freund and Schapire, 1997) and RankBoost (Freund et al., 2004), which target at classification and ranking respectively. AdaBoost starts with a pool of weak classifiers and iteratively selects the best one based on the sample weights in that iteration. The final classifier is a linear combination of the selected weak classifiers, where each classifier is weighted by its performance. In each iteration, sample weights are updated according to the selected classifier so that the subsequent classifiers will focus more on the *hard* samples. RankBoost seeks to extend AdaBoost to ranking model combination. The model weights are learned from the ranking performance in a boosting manner. In this paper, we extend RankBoost to combine ranking results from multiple KG embedding models. This technique addresses KG completion by combining knowledge from multiple sources and effectively compensates for the inherent errors in any entity matching processes.

3 Method

In this section, we introduce KEnS, an embedding-based ensemble inference framework for multilingual KG completion.

KEnS conducts two processes, i.e. *embedding learning* and *ensemble inference*. The embedding learning process trains the *knowledge model* that encodes entities and relations of every KG in a shared embedding space, as well as the *alignment model* that seizes the correspondence in different KGs and enables the projection of queries and answers across different KG embeddings. The ensemble inference process combines the predictions from multiple KG embeddings to improve fact prediction. Particularly, to assess the confidence of predictions from each source, we introduce a boosting method to learn entity-specific weights for knowledge models.

3.1 Preliminaries

A KG G consists of a set of (relational) facts $\{(h, r, t)\}$, where h and t are the head and tail entities of the fact (h, r, t), and r is a relation. Specifically, $h, t \in E$ (the set of entities in G), and $r \in R$ (the set of relations). To cope with KG completion, the fact prediction task seeks to fill in the right entity for the missing head or tail of an unseen triple. Without loss of generality, we hereafter discuss the case of predicting missing tails. We refer to a triple

with a missing tail as a *query* $q = (h, r, \underline{?t})$. The answer set $\Omega(q)$ consists of all the right entities that fulfill q. For example, we may have a query *(The Tale of Genji, genre, ?t)*, and its answer set will include *Monogatari*, *Love Story*, etc.

Given KGs in M languages $G_1, G_2, \ldots, G_M$ ($|E_i| \leq |E_j|, i < j$), we seek to perform fact prediction on each of those by transferring knowledge from the others. We consider fact prediction as a ranking task in the KG embedding space, which is to transfer the query to external KGs and to combine predictions from multiple embedding models into a final ranking list. Particularly, given the existing situation of the major KGs, we use the following settings: (i) entity alignment information is available between any two KGs, though limited; and (ii) relations in different language-specific KGs are represented with a unified schema. The reason for the assumption is that unifying relations is usually feasible, since the number of relations is often much smaller compared to the enormous number of entities in KGs. This has been de facto achieved in a number of influential knowledge bases, including DBpedia (Lehmann et al., 2015), Wikidata (Vrandečić and Krötzsch, 2014) and YAGO (Rebele et al., 2016). In contrast, KGs often consist of numerous entities that cannot be easily aligned, and entity alignment is available only in small amounts.

3.2 Embedding Learning

The embedding learning process jointly trains the *knowledge model* and the *alignment model* following Chen et al. (2017), while self-learning is added to improve the alignment learning. The details are described below.

Knowledge model. A knowledge model seeks to encode the facts of a KG in the embedding space. For each language-specific KG, it characterizes the plausibility of its facts. Notation-wise, we use boldfaced $\boldsymbol{h}, \boldsymbol{r}, \boldsymbol{t}$ as embedding vectors for head h, relation r and tail t respectively. The learning objective is to minimize the following margin ranking loss:

$$\mathcal{J}_K^G = \sum_{\substack{(h,r,t) \in G, \\ (h',r,t') \notin G}} [f(\boldsymbol{h'}, \boldsymbol{r}, \boldsymbol{t'}) - f(\boldsymbol{h}, \boldsymbol{r}, \boldsymbol{t}) + \gamma]_+ \quad (1)$$

where $[\cdot]_+ = \max(\cdot, 0)$, and f is a model-specific triple scoring function. The higher score indicates the higher likelihood that the fact is true. γ is a hyperparameter, and (h', r, t') is a negative sampled

triple obtained by randomly corrupting either head or tail of a true triple (h, r, t).

We here consider two representative triple scoring techniques: TransE (Bordes et al., 2013) and RotatE (Sun et al., 2019b). TransE models relations as translations between head entities and tail entities in a Euclidean space, while RotatE models relations as rotations in a complex space. The triple scoring functions are defined as follows.

$$f_{\text{TransE}}(h, r, t) = -\|\boldsymbol{h} + \boldsymbol{r} - \boldsymbol{t}\|_2 \quad (2)$$
$$f_{\text{RotatE}}(h, r, t) = -\|\boldsymbol{h} \circ \boldsymbol{r} - \boldsymbol{t}\|_2 \quad (3)$$

where $\circ : \mathbb{C}^d \times \mathbb{C}^d \to \mathbb{C}^d$ denotes Hadamard product for complex vectors, and $\|\cdot\|_2$ denotes L_2 norm.
Alignment model. An alignment model is trained to match entity counterparts between two KGs on the basis of a small amount of seed entity alignment. We embed all KGs in one vector space and make each pair of aligned entities embedded closely. Given two KGs G_i and G_j with $|E_i| \leq |E_j|$, the alignment model loss is defined as:

$$\mathcal{J}_A^{G_i \leftrightarrow G_j} = \sum_{(e_i, e_j) \in \Gamma_{G_i \leftrightarrow G_j}} \|\boldsymbol{e}_i - \boldsymbol{e}_j\|_2 \quad (4)$$

where $e_i \in E_i, e_j \in E_j$ and $\Gamma_{G_i \leftrightarrow G_j}$ is the set of seed entity alignment between G_j and G_i. Assuming the potential inaccuracy of alignment, we do not directly assign the same vector to aligned entities of different language-specific KGs.

Particularly, as the seed entity alignment is provided in small amounts, the alignment process conducts self-learning, where training iterations incrementally propose more training data on unaligned entities to guide subsequent iterations. At each iteration, if a pair of unaligned entities in two KGs are mutual nearest neighbors according to the CSLS measure (Conneau et al., 2018), KEnS adds this highly confident alignment to the training data.
Learning objective. We conduct joint training of knowledge models for multiple KGs and alignment models between each pair of them via minimizing the following loss function:

$$\mathcal{J} = \sum_{m=1}^{M} \mathcal{J}_K^{G_m} + \lambda \sum_{i=1}^{M} \sum_{j=i+1}^{M} \mathcal{J}_A^{G_i \leftrightarrow G_j} \quad (5)$$

where $\mathcal{J}_K^{G_m}$ is the loss of the knowledge model on G_m as defined in Eq (1), $\mathcal{J}_A^{G_i \leftrightarrow G_j}$ is the alignment loss between G_i and G_j. λ is a positive hyperparameter that weights the two model components.

Following Chen et al. (2017), instead of directly optimizing $\mathcal{J}$ in Eq. (5), our implementation optimizes each $\mathcal{J}_K^G$ and each $\lambda \mathcal{J}_A^{G_i \leftrightarrow G_j}$ alternately in separate batches. In addition, we enforce L_2-regularization to prevent overfitting.

3.3 Ensemble Inference

We hereby introduce how KEnS performs fact prediction on multiple KGs via ensemble inference.

Cross-lingual query and knowledge transfer. To facilitate the process of completing KG G_i with the knowledge from another KG G_j, KEnS first predicts the alignment for entities between G_i and G_j. Then, it uses the alignment to transfer queries from G_i to G_j, and transfer the results back. Specifically, alignment prediction is done by performing an kNN search in the embedding space for each entity in the smaller KG (i.e. the one with fewer entities) and find the closest counterpart from the larger KG. Inevitably, some entities in the larger KG will not be matched with a counterpart due to the 1-to-1 constraint. In this case, we do not transfer queries and answers for that entity.

Weighted ensemble inference. We denote the embedding models of $G_1, \ldots, G_M$ as $f_1, \ldots, f_M$. On the target KG where we seek to make predictions, given each query, the entity candidates are ranked by the weighted voting score of the models:

$$s(e) = \sum_{i=1}^M w_i(e) N_i(e) \qquad (6)$$

where e is an entity on the target KG, and $w_i(e)$ is an entity-specific model weight, $N_i(e)$ is 1 if e is ranked among top K by f_i, otherwise 0.

We propose three variants of KEnS that differ in the computing of $w_i(e)$, namely KEnS$_b$, KEnS$_v$ and KEnS$_m$. Specifically, KEnS$_b$ learns an entity-specific weight $w_i(e)$ for each entity in a _boosting_ manner, KEnS$_v$ fixes $w_i(e) = 1$ for all f_i and e (i.e. _majority voting_), and KEnS$_m$ adopts _mean reciprocal rank_ (MRR) of f_i on the validation set of the target KG as $w_i(e)$. We first present the technical details of the boosting-based KEnS$_b$.

3.3.1 Boosting Based Weight Learning

KEnS$_b$ seeks to learn model weights for ranking combination, which aims at reinforcing correct beliefs and compensating for alignment error. An embedding model that makes more accurate predictions should receive a higher weight. Inspired by RankBoost (Freund et al., 2004), we reduce

the ranking combination problem to a classifier ensemble problem. KEnS$_b$ therefore learns model weights in a similar manner as AdaBoost.

Validation queries and critical entity pairs. To compute entity-specific weights $w_i(e)$, KEnS$_b$ evaluates the performance of f_i on a set of _validation queries_ related to e. These queries are converted from all the triples in the validation set that mention e. An example of validation queries for the entity _The Tale of Genji_ is given as below.

Example 3.1. Examples of triples and validation queries for the entity _The Tale of Genji_.

Triples:
```
{(The Tale of Genji, country, Japan)
(The Tale of Genji, genre, Monogatari)
(The Tale of Genji, genre, Love Story)}
```
Queries:
```
Q = {q₁ = (The Tale of Genji, country, ?t)
     q₂ = (The Tale of Genji, genre, ?t)}
```
$Q = \{q_1 = $ (The Tale of Genji, country, ?t)

$\quad q_2 = $ (The Tale of Genji, genre, ?t)$\}$

Similar to RankBoost (Freund et al., 2004), given a query q, KEnS$_b$ evaluates the ranking performance of a model by checking if each of the _critical entity pairs_ $\{(e, e')\}$ is ranked in correct order, where e is a ground truth tail and e' is an incorrect one. An example of critical entity pairs is given as below:

Example 3.2. Critical entity pairs for the query _(The Tale of Genji, genre, ?t)_. Ground truth tails are boldfaced. Pairs with x-marks indicate wrong prediction orders.

Correct ranking :

Monogatari, Love Story, Modernist, Science Fiction

Predicted ranking:

Modernist, **Monogatari, Love Story**, Science Fiction

Critical pair ranking results:

(**Monogatari**, Modernist) ✗, (**Love Story**, Modernist) ✗

(**Monogatari**, Science Fiction) ✓,

(**Love Story**, Science Fiction) ✓

Uncritical pairs:

(Monogatari, Love Story), (Modernist, Science Fiction)

Ranking loss. The overall objective of KEnS$_b$ is to minimize the sum of ranks of all correct answers in the combined ranking list $\sum_q \sum_{e \in \Omega(q)} r(e)$, where $\Omega(q)$ is the answer set of query q and $r(e)$ is the rank of entity e in the combined ranking list of the ensemble inference. Essentially, the above objective is minimizing the number of mis-ordered critical entity pairs in the combined ranking list. Let the set of all the critical entity pairs from all the validation queries of an entity as P. Freund et al. (2004) have proved that, when using RankBoost,

this ranking loss is bounded as follows:

$$|\{p : p \in P, p \text{ is mis-ordered}\}| \leq |P| \prod_{m=1}^{M} Z^m$$

where M is the number of KGs and therefore the maximum number of rounds in boosting. Z^m is the weighted ranking loss of the m-th round:

$$Z^m = \sum_{p \in P} D^m(p) e^{-w^m [\![p]\!]^m} \qquad (7)$$

where $[\![p]\!]^m = 1$ if the critical entity pair p is ranked in correct order by the selected embedding model in the m-th round, otherwise $[\![p]\!]^m = -1$, $D^m(p)$ is the weight of the critical entity pair p in the m-th round, and w^m is the weight of the chosen model in that round. Now the ranking combination problem is reduced to a common classifier ensemble problem.

Boosting procedure. The boosting process alternately repeats two steps: (i) Evaluate the ranking performance of the embedding models and choose the best one f^m according to the entity pair weight distribution in that round; (ii) Update entity pair weights to put more emphasis on the pairs which f_m ranks incorrectly.

Entity pair weights are initialized uniformly over P as $D^1(p) = \frac{1}{|P|}, p \in P$. In the m-th round ($m = 1, 2, ..., M$), KEnS$_b$ chooses an embedding model f^m and sets its weight w^m, seeking to minimize the weighted ranking loss Z^m defined in Eq.(7). By simple calculus, when choosing the embedding model f_i as the model of the m-th round, w_i^m should be set as follows to minimize Z^m:

$$w_i^m = \frac{1}{2} \ln \left(\frac{\sum_{p \in P, [\![p]\!]=1} D^m(p)}{\sum_{p \in P, [\![p]\!]=-1} D^m(p)} \right) \qquad (8)$$

As we can see from Eq. (8), the higher w_i^m indicates the better performance of f_i under the current entity pair weight distribution D^m. We select the best embedding model in the m-th round f^m based on the maximum weight $w^m = \max\{w_1^m, ..., w_M^m\}$.

After choosing the best model f^m at this iteration, we update the entity pair weight distribution to put more emphasis on what f^m ranked wrong. The new weight distribution D^{m+1} is updated as:

$$D^{m+1}(p) = \frac{1}{Z^m} D^m(p) e^{-w^m [\![p]\!]^m} \qquad (9)$$

where Z^m works as a normalization factor. KEnS$_b$ decreases the weight of $D(p)$ if the selected

Lang.	EN	FR	ES	JA	EL
#Ent.	13,996	13,176	12,382	11,805	5,231
#Rel.	831	178	144	128	111
#Triples	80,167	49,015	54,066	28,774	13,839

Table 1: Statistics of DBP-5Ldataset. *Ent.* and *Rel.* stand for entities and relations respectively.

model ranks the entity pair in correct order and increases the weight otherwise. Thus, $D(p)$ will tend to concentrate on the pairs whose relative ranking is hardest to determine.

For queries related to a specific entity, this process is able to recognize the embedding models that perform well on answering those queries and rectify the mistakes made in the previous iteration.

3.3.2 Other Ensemble Techniques

We also investigate two other model variants with simpler ensemble techniques.

Majority vote (KEnS$_v$): A straightforward ensemble method is to re-rank entities by their nomination counts in the prediction of all knowledge models, which substitutes the voting score (Eq. 6) with $s(e) = \sum_{i=1}^{M} N_i(e)$, where $N_i(e)$ is 1 if e is ranked among the top K by the knowledge model f_i, otherwise 0. When there is a tie, we order by the MRR given by the models on the validation set.
MRR weighting (KEnS$_m$): MRR is a widely-used metric for evaluating the ranking performance of a model (Bordes et al., 2013; Yang et al., 2015; Trouillon et al., 2016), which may also serve as a weight metric for estimating the prediction confidence of each language-specific embedding in ensemble inference (Shen et al., 2017). Let the MRR of f_i be u_i on the validation set, the entities are ranked according to the weighted voting score $s(e) = \sum_{i=1}^{M} u_i N_i(e)$.

4 Experiments

In this section, we conduct the experiment of fact prediction by comparing KEnS variants with various KG embeddings. We also provide a detailed case study to help understand the principle of ensemble knowledge transfer.

4.1 Experiment Settings

To the best of our knowledge, existing datasets for fact prediction contain only one monolingual KG or bilingual KGs. Hence, we prepared a new dataset DBP-5L, which contains five language-specific KGs extracted from English (EN), French (FR), Spanish (ES) and Japanese (JA) and Greek

KG	Greek			Japanese			Spanish			French			English		
Hits@k (%)	1	3	10	1	3	10	1	3	10	1	3	10	1	3	10
TransD	2.8	16.9	29.8	4.2	16.3	28.8	2.12	20.4	11.5	3.3	14.4	25.7	2.9	15.4	27.4
DistMult	8.9	13.0	11.3	9.3	18.4	27.5	7.4	15.0	22.4	6.1	14.3	23.8	8.8	19.4	30.0
HolE	4.2	9.5	18.3	25.5	29.5	32.8	20.1	26.8	29.4	22.4	24.4	28.9	12.3	20.4	25.4
TransE	13.1	23.4	43.7	21.1	34.4	48.5	13.5	29.4	45.0	17.5	33.1	48.8	7.3	16.4	29.3
KEnS_v(TransE)	23.1	36.7	64.7	22.6	35.2	52.5	15.0	28.3	**49.0**	18.7	29.4	52.0	10.8	20.4	**39.4**
KEnS_m(TransE)	26.3	42.1	65.8	26.1	37.7	55.3	16.8	**32.9**	48.6	20.5	35.6	52.8	11.4	21.2	31.3
KEnS_b(TransE)	**26.4**	**42.4**	**66.1**	**26.7**	**39.8**	**56.4**	**17.4**	32.6	48.3	**20.8**	**35.9**	**53.1**	**11.7**	**21.8**	32.0
RotatE	14.5	18.8	36.2	26.4	36.2	60.2	21.2	31.6	53.9	23.2	29.4	55.5	12.3	25.4	30.4
RotatE+PARIS	-	-	-	-	-	-	20.8	39.4	59.1	22.8	32.4	60.8	12.4	22.7	31.5
KEnS_v(RotatE)	20.5	34.3	50.1	31.9	50.0	65.0	20.8	41.0	59.9	23.7	42.7	61.9	13.4	23.6	34.2
KEnS_m(RotatE)	22.0	35.0	51.4	32.0	49.9	65.0	21.2	41.6	60.0	24.5	44.8	62.5	12.1	24.5	34.3
KEnS_b(RotatE)	**27.5**	**40.6**	**56.5**	**32.9**	**49.9**	**64.8**	**22.3**	**42.4**	**60.6**	**25.2**	**44.5**	**62.6**	**14.4**	**27.0**	**39.6**

Table 2: Fact prediction results on DBP-5L. The overall best results are under-scored.

(EL) DBpedia (Lehmann et al., 2015). Table 1 lists the statistics of the contributed dataset DBP-5L. The relations of the five KGs are represented in a unified schema, which is consistent with the problem definition in Section 3.1. The English KG is the most populated one among the five. To produce KGs with a relatively consistent set of entities, we induce the subgraphs by starting from a set of seed entities where we have alignment among all language-specific KGs and then incrementally collecting triples that involve other entities. Eventually between any two KGs, the alignment information covers around 40% of entities. Based on the same set of seed entities, the Greek KG ends up with a notably smaller vocabulary and fewer triples than the other four. We split the facts in each KG into three parts: 60% for training, 30% for validation and weight learning, and 10% for testing.

Experimental setup. We use the Adam (Kingma and Ba, 2014) as the optimizer and fine-tune the hyper-parameters by grid search based on $Hits@1$ on the validation set. We select among the following sets of hyper-parameter values: learning rate $lr \in \{0.01, 0.001, 0.0001\}$, dimension $d \in \{64, 128, 200, 300\}$, batch size $b \in \{256, 512, 1024\}$, and TransE margin $\gamma \in \{0.3, 0.5, 0.8\}$. The best setting is $\{lr = 0.001, d = 300, b = 256\}$ for KEnS(TransE) and $\{lr = 0.01, d = 200, b = 512\}$ for KEnS(RotatE). The margin for TransE is 0.3. The L_2 regularization coefficient is fixed as 0.0001.

Evaluation protocol. For each test case (h, r, t), we consider it as a query $(h, r, ?t)$ and retrieve top K prediction results for $?t$. We compare the propor-

tion of queries with correct answers ranked within top K retrieved entities. We report three metrics with K as $1, 3, 10$. $Hits@1$ is equivalent to accuracy. All three metrics are preferred to be higher. Although another common metric, Mean Reciprocal Rank (MRR), has been used in previous works (Bordes et al., 2013), it is not applicable to the evaluation of our framework because our ensemble framework combines the top entity candidates from multiple knowledge models and yields top K final results without making any claims for entities out of this scope. Following previous works, we use the "filtered" setting with the premise that the candidate space has excluded the triples that have been seen in the training set (Wang et al., 2014).

Competitive methods. We compare six variants of KEnS, which are generated by combining two knowledge models and three ensemble inference techniques introduced in in Section 3. For baseline methods, besides the single-embedding TransE (Bordes et al., 2013) and RotatE (Sun et al., 2019b), we also include DistMult (Yang et al., 2015), TransD (Ji et al., 2015), and HolE (Nickel et al., 2016). After extensive hyperparameter tuning, the baselines are set to their best configurations. We also include a baseline named *RotatE+PARIS*, which trains RotatE on 5 KGs and uses the representative non-embedding symbolic entity alignment tool PARIS (Suchanek et al., 2011) for entity matching. PARIS delivered entity matching predictions for 58%-62% entities in the English, French, and Spanish KG, but almost no matches are delivered for entities in the Greek and Japanese KG, since PARIS mainly relies on entity label similarity.
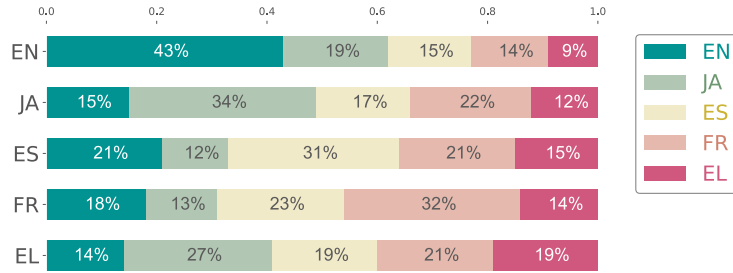
Figure 2: Average model weights learned by KEnS$_b$(TransE).

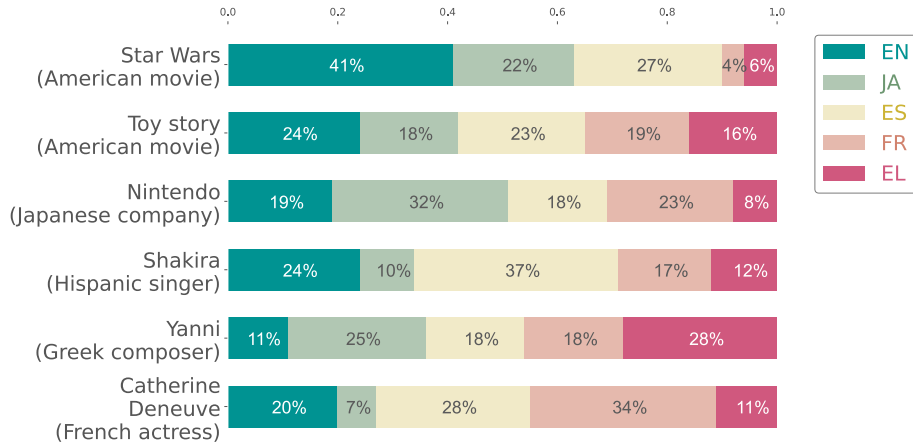

Figure 3: Examples of language-specific model weights learned by KEnS$_b$(TransE). Percentages have been rounded.

The results on the Greek and Japanese KG are thus omitted for RotatE+PARIS.

4.2 Main Results

The results are reported in Table 2. As shown, the ensemble methods by KEnS lead to consistent improvement in fact prediction. Overall, the ensemble inference leads to 1.1%-13.0% of improvement in $Hits@1$ over the best baseline methods. The improved accuracy shows that it is effective to leverage complementary knowledge from external KGs for KG completion. We also observe that KEnS brings larger gains on sparser KGs than on the well-populated ones. Particularly, on the low-resource Greek KG, KEnS$_b$(RotatE) improves $Hits@1$ by as much as 13.0% over its single-KG counterpart. This finding corroborates our intuition that the KG with lower knowledge coverage and sparser graph structure benefits more from complementary knowledge.

Among the variants of ensemble methods, KEnS$_m$ offers better performance than KEnS$_v$, and KEnS$_b$ outperforms the other two in general. For example, on the Japanese KG, KEnS$_v$(TransE) improves $Hits@1$ by 3.5% from the single-KG

TransE, while KEnS$_m$ leads to a 5.0% increase, and KEnS$_b$ further provides a 5.6% of improvement. The results suggest that MRR is an effective measure of the trustworthiness of knowledge models during ensemble inference. Besides, KEnS$_b$ is able to assess trustworthiness at a finer level of granularity by learning entity-specific model weights and can thus further improve the performance.

In summary, the promising results by KEnS variants show the effectiveness of transferring and leveraging cross-lingual knowledge for KG completion. Among the ensemble techniques, the boosting technique represents the most suitable one for combining the prediction results from different models.

4.3 Case Studies

In this section, we provide case studies to show how KEnS is able to transfer cross-lingual knowledge to populate different KGs.

Model weights. The key to the significantly enhanced performance of KEnS$_b$ is the effective combination of multilingual knowledge from multiple sources. Fig 2 shows the average model weight learnt by KEnS$_b$(TransE), which depicts how external knowledge from cross-lingual KGs contributes

Model	Top 3 Predicted Tails
English	Television, Publishing, Information technology
Japanese	**Video game**, Anime, **Consumer electronics**
Spanish	Music, Telecommunication, Retail
French	Retail, Television, **Video game**,
Greek	Nintendo, Music, Wii
KEnS_v	[**Video game**, Television](tie), Music
KEnS_m	Television, **Video game**, Music
KEnS_b	**Video game**, Television, **Consumer electronics**

Table 3: An example of fact prediction on the English KG by the English knowledge model, four supporter knowledge models, and KEnS(TransE) variants. Top 3 predicted tails for the query `(Nintendo, industry, ?t)` are listed. Ground truths are boldfaced.

to target KG completion in general. The model weights imply that sparser KGs benefit more from the knowledge transferred from others. Particularly, when predicting for the Greek KG, the weights of other languages sums up to 81%. This observation indicates that the significant boost received on the Greek KG comes with the fact that it has accepted the most complementary knowledge from others. In contrast, when predicting on the most populated English KG, the other language-specific models give a lesser total weight of 57%.

Among the three KEns variants, the superiority of KEnS_b is attributed to identification of more credible knowledge sources, thus making more accurate predictions. For language-specific KGs, the higher level of credibility often stems from the cultural advantage the KG has over the entity. Fig 3 presents the model weights for 6 culture-related entities learned by KEnS_b(TransE). It shows that KEns can locate the language-specific knowledge model that has a cultural advantage and assign it with a higher weight, which is the basis of an accurate ensemble prediction.

Ensemble inference. To help understand how the combination of multiple KGs improves KG completion and show the effectiveness of leveraging complementary culture-specific knowledge , we present a case study about predicting the fact `(Nintendo, industry, ?t)` for English KG. Table 3 lists the top 3 predicted tails yielded by the KEnS(TransE) variants, along with those by the English knowledge model and supporter knowledge models before ensemble. The predictions made by the Japanese KG are the closest to the ground truths. The reason may be that Japanese KG has documented much richer knowledge about this Japanese video game company, including many of the video games that this com-

pany has released. Among the three KEnS variants, KEnS_b correctly identifies Japanese as the most credible source and yields the best ranking.

5 Conclusion

In this paper, we have proposed a new ensemble prediction framework aiming at collaboratively predicting unseen facts using embeddings of different language-specific KGs. In the embedding space, our approach jointly captures both the structured knowledge of each KG and the entity alignment that bridges the KGs. The significant performance improvements delivered by our model on the task of KG completion were demonstrated by extensive experiments. This work also suggests promising directions of future research. One is to exploit the potential of KEnS on completing low-resource KGs, and the other is to extend the ensemble transfer mechanism to population sparse domain knowledge in biological (Hao et al., 2020) and medical knowledge bases (Zhang et al., 2020). Pariticularly, we also seek to ensure the global logical consistency of predicted facts in the ensemble process by incorporating probabilistic constraints (Chen et al., 2019).

References

Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *Advances in Neural Information Processing Systems (NIPS)*, pages 2787–2795.

Antoine Bordes, Jason Weston, and Nicolas Usunier. 2014. Open question answering with weakly su-

pervised embedding models. In *Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases (ECML-PKDD*, volume 8724, pages 165–180. Springer.

Leo Breiman. 1996. Bagging predictors. *Machine learning*, 24(2):123–140.

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 Annual Meeting of Associations for Computational Linguistics (ACL)*, pages 1452–1461. Association for Computational Linguistics.

Muhao Chen, Changping Meng, Gang Huang, and Carlo Zaniolo. 2018a. Neural article pair modeling for wikipedia sub-article matching. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 3–19. Springer.

Muhao Chen, Yingtao Tian, Kai-Wei Chang, Steven Skiena, and Carlo Zaniolo. 2018b. Co-training embeddings of knowledge graphs and entity descriptions for cross-lingual entity alignment. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, pages 3998–4004. International Joint Conferences on Artificial Intelligence Organization.

Muhao Chen, Yingtao Tian, Mohan Yang, and Carlo Zaniolo. 2017. Multilingual knowledge graph embeddings for cross-lingual knowledge alignment. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1511–1517. International Joint Conferences on Artificial Intelligence.

Xuelu Chen, Muhao Chen, Weijia Shi, Yizhou Sun, and Carlo Zaniolo. 2019. Embedding uncertain knowledge graphs. In *Proceedings of AAAI Conference on Artificial Intelligence (AAAI)*, pages 3363–3370. AAAI Press.

Alexis Conneau, Guillaume Lample, Marc'Aurelio Ranzato, Ludovic Denoyer, and Hervé Jégou. 2018. Word translation without parallel data. In *International Conference on Learning Representations (ICLR)*.

Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. 2018. Convolutional 2d knowledge graph embeddings. In *Proceedings of AAAI Conference on Artificial Intelligence (AAAI)*, pages 1811–1818. AAAI Press.

Yoav Freund, Raj Iyer, Robert E. Schapire, and Yoram Singer. 2004. RankBoost: An efficient boosting algorithm for combining preferences. *Journal of Machine Learning Research (JMLR)*, 4(6):933–969.

Yoav Freund and Robert E Schapire. 1997. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1):119–139.

Junheng Hao, Muhao Chen, Wenchao Yu, Yizhou Sun, and Wei Wang. 2019. Universal representation learning of knowledge bases by jointly embedding instances and ontological concepts. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 1709–1719. ACM.

Junheng Hao, Chelsea Ju, Muhao Chen, Yizhou Sun, Carlo Zaniolo, and Wei Wang. 2020. Bio-joie: Joint representation learning of biological knowledge bases. In *Proceedings of the 11st ACM Conference on Bioinformics, Computational Biology and Biomedicine (BCB)*. ACM.

Guoliang Ji, Shizhu He, Liheng Xu, Kang Liu, and Jun Zhao. 2015. Knowledge graph embedding via dynamic mapping matrix. In *Proceedings of the Annual Meeting of Associations for Computational Linguistics (ACL)*, pages 687–696. The Association for Computer Linguistics.

Shaoxiong Ji, Shirui Pan, Erik Cambria, Pekka Marttinen, and Philip S Yu. 2020. A survey on knowledge graphs: Representation, acquisition and applications. *arXiv preprint arXiv:2002.00388*.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*.

Rik Koncel-Kedziorski, Dhanush Bekal, Yi Luan, Mirella Lapata, and Hannaneh Hajishirzi. 2019. Text generation from knowledge graphs with graph transformers. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers) (NAACL)*, pages 2284–2293. Association for Computational Linguistics.

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

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 *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 2723–2732. Association for Computational Linguistics.

Maximilian Nickel, Lorenzo Rosasco, Tomaso A Poggio, et al. 2016. Holographic embeddings of knowledge graphs. In *Proceedings of AAAI Conference on Artificial Intelligence (AAAI)*, pages 1955–1961. AAAI Press.

Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. 2011. A three-way model for collective learning on multi-relational data. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 809–816. Omnipress.

Shichao Pei, Lu Yu, Robert Hoehndorf, et al. 2019. semi-supervised entity alignment via knowledge graph embedding with awareness of degree difference. In *Proceedings of the Web Confererence (WWW)*, pages 3130–3136. ACM.

Thomas Rebele, Fabian Suchanek, Johannes Hoffart, Joanna Biega, Erdal Kuzey, and Gerhard Weikum. 2016. Yago: A multilingual knowledge base from wikipedia, wordnet, and geonames. In *Proceedings of the International Semantic Web Conference (ISWC)*, volume 9982 of *Lecture Notes in Computer Science*, pages 177–185. Springer.

Jiaming Shen, Zeqiu Wu, Dongming Lei, Jingbo Shang, Xiang Ren, and Jiawei Han. 2017. Setexpan: Corpus-based set expansion via context feature selection and rank ensemble. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, volume 10534 of *Lecture Notes in Computer Science*, pages 288–304. Springer.

Fabian M Suchanek, Serge Abiteboul, et al. 2011. Paris: Probabilistic alignment of relations, instances, and schema. *Proceedings of the VLDB Endowment (PVLDB)*, 5(3).

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

Zequn Sun, Wei Hu, Qingheng Zhang, and Yuzhong Qu. 2018. Bootstrapping entity alignment with knowledge graph embedding. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, pages 4396–4402. International Joint Conferences on Artificial Intelligence Organization.

Zequn Sun, Chengming Wang, Wei Hu, Muhao Chen, Jian Dai, Wei Zhang, and Yuzhong Qu. 2020a. Knowledge graph alignment network with gated multi-hop neighborhood aggregation. In *Proceedings of AAAI Conference on Artificial Intelligence (AAAI)*, pages 222–229. AAAI Press.

Zequn Sun, Jiacheng Huang Wang, Wei Hu, Muhao Chen, and Yuzhong Qu. 2019a. Transedge: Translating relation-contextualized embeddings for knowledge graphs. In *Proceedings of the International Semantic Web Conference (ISWC)*, pages 612–629. Springer.

Zequn Sun, Qingheng Zhang, Wei Hu, Chengming Wang, Muhao Chen, Farahnaz Akrami, and Chengkai Li. 2020b. A benchmarking study of embedding-based entity alignment for knowledge graphs. *Proceedings of the VLDB Endowment*, 13:2326–2340.

Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. 2019b. Rotate: Knowledge graph embedding by relational rotation in complex space. In *International Conference on Learning Representations (ICLR)*.

Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. 2016. Complex embeddings for simple link prediction. In *Proceedings of the International Conference on Machine Learning (ICML)*, volume 48, pages 2071–2080. PMLR.

Bayu Distiawan Trsedya, Jianzhong Qi, and Rui Zhang. 2019. Entity alignment between knowledge graphs using attribute embeddings. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, pages 297–304. AAAI Press.

Denny Vrandečić and Markus Krötzsch. 2014. Wikidata: a free collaborative knowledge base. *Communications of ACM*, 57(10):78–85.

Quan Wang, Zhendong Mao, Bin Wang, and Li Guo. 2017. Knowledge graph embedding: A survey of approaches and applications. *IEEE Transactions on Knowledge and Data Engineering*, 29(12):2724–2743.

Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. 2014. Knowledge graph embedding by translating on hyperplanes. In *Proceedings of AAAI Conference on Artificial Intelligence (AAAI)*, pages 1112–1119. AAAI Press.

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

David H Wolpert. 1992. Stacked generalization. *Neural networks*, 5(2):241–259.

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

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)*, pages 4430–4440. Association for Computational Linguistics.

Qingheng Zhang, Zequn Sun, Wei Hu, Muhao Chen, Lingbing Guo, and Yuzhong Qu. 2019. Multi-view knowledge graph embedding for entity alignment.

In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, pages 5429–5435. International Joint Conferences on Artificial Intelligence Organization.

Tianran Zhang, Muhao Chen, and Alex Bui. 2020. Diagnostic prediction with sequence-of-sets representation learning for clinical event. In *Proceedings of the 18th International Conference on Artificial Intelligence in Medicine (AIME)*. Springer.