Multi-graph Affinity Embeddings for Multilingual Knowledge Graphs

Muhao Chen, Tao Zhou, Pei Zhou, Carlo Zaniolo
Department of Computer Science, UCLA
{muhaochen, taozhou, zpcs03, zaniolo}@cs.ucla.edu

Abstract

Multilingual knowledge graph embeddings provide important latent semantic representations for knowledge-driven cross-lingual NLP tasks. Learning such embeddings is currently based on the training of a weakly supervised alignment model in joint with monolingual knowledge models. However, existing techniques for alignment model suffer significantly from the multilingual inconsistency of knowledge graphs. In this paper, we propose an improved model by learning a generalized affine-map-based alignment model. We find that the proposed approach effectively addresses the limitations of existing approaches, especially in handling the incoherence of embedding spaces on different languages. Experimental results show that the proposed approach offers better performance for both entity and triple-wise knowledge alignment tasks.

1 Introduction

Knowledge bases (KBs) like ConceptNet [19], WordNet [2], and DBpedia [14] constitute essential sources of knowledge. These KBs store knowledge graphs (KGs) that represent two aspects of knowledge: the *monolingual knowledge* which models relation facts of entities as triples, and the *cross-lingual knowledge* that synchronizes monolingual knowledge among various human languages.

Embedding models for monolingual KGs have been extensively studied in the past half decade. These models provide efficient and versatile methods to infuse the symbolic knowledge of KGs into machine learning, by encoding entities in low-dimensional embedding spaces, and supporting relational inferences between entity embeddings via simple vector algebra. For example, the translation-based model TransE [3] represents a relation as the vector translation between two entities, and quantifies the semantic relatedness of entities as vector distance. Models of this kind have been widely applied to NLP-related tasks, such as KG completion [15], relation extraction [23], question answering (QA) [5], and visual semantic labeling [8].

Recently, embeddings have been leveraged to connect the KGs of multiple languages [6]. This advancement is significant as it provides KG embeddings with generic multilinguality despite the fact that cross-lingual knowledge usually covers small parts of the KG. This undoubtedly benefits vast NLP tasks such as knowledge alignment, cross-lingual QA, and machine translation [7]. Successful learning of such embeddings is qualified by two model components. A *knowledge model* (KM) distributes knowledge of each language in a separated embedding space. On top of that, an *alignment model* (AM) learns cross-lingual transitions without losing the encoded monolingual structures, for which existing models employ several techniques, including joint techniques based on axis calibration, translation vectors, and linear transforms [6], as well as off-line maps [20], and parameter sharing [26].

Although effective, the alignment techniques in existing models still suffer significantly from the inconsistency of multilingual KGs. This is because language-specific versions of the KG are usually extended asynchronously [14, 22]. Therefore, the vocabulary of entities and relations, and the relation

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

facts are very inconsistent, which easily lead to the incoherence of embedding spaces for involved language (as shown in Fig. 1). We find off-line techniques and parameter sharing to be ill-suited when facing this issue. Although joint techniques adapt well to the incoherence, we show that current forms of alignment models are still hindered by the conflict between accurate cross-lingual transitions and embedding normalization.

In this paper, we propose MTransE-Af , a simple but effective model, which enhances MTransE [6] with a more generalized affine-map-based AM. We find that the affine map efficaciously addresses the limitations of previous techniques for handling the multilingual inconsistency. We evaluate MTransE-Af on entity and triple-wise cross-lingual knowledge alignment tasks using three benchmark datasets of trilingual KGs. Experimental results confirm the effectiveness of MTransE-Af by offering notably better performance than previous models on the cross-lingual tasks.

2 Related Work

KG embedding models are first explored in the monolingual scenario. Previous works have made significant advances on *translation-based* models. To characterize a triple (h, r, t) where r is a relation between entities h and t, the forerunner TransE [3] of this family follows the objective $\mathbf{h} + \mathbf{r} \approx \mathbf{t}$ by representing r as a translation between entity vectors $\mathbf{h}$ and $\mathbf{t}$. Following TransE, later works such as TransH [24], TransR [15], TransD [11], and TransA [12] differentiate such encoding process to separated embedding spaces using different forms of relation-specific projections, and offer better performance on KG completion and relation extraction tasks. In addition to these, there are non-translation-based models, including successful neural models such as RESCAL [17], HolE [16], and ComplEx [21]. These models perform comparably to translation-based models at the cost of higher model complexity.

To perform multilingual learning on KGs, MTransE [6] connects monolingual models with a jointly trained alignment model, for which three aligment techniques are employed, i.e., axis calibration that adjusts embedding spaces to collocate cross-lingual counterparts (MTransE-AC), cross-lingual vector translation (MTransE-TV), and linear transforms across embedding spaces (MTransE-LT) for different languages. The MTransE-LT thereof achieves the best performance on knowledge alignment tasks. JAPE is introduced in [20] which strengthens the alignment learning of MTransE-AC based on identical entity attributes. This model performs well on KBs that provide entity attributes, though such attributes are not generally available in many KBs such as ConceptNet and WordNet. Another relevant model ITransE [26] enforces parameter sharing on aligned entities. ITransE has been used to align entities for monolingual KGs where vocabularies and relation facts are very coherent. Though it can be used for multilingual learning, we find this strict technique does not adapt well to the substantially inconsistent multilingual scenario. Note that, projection-based models for words, i.e. LM, CCA [9], and joint orthogonal transforms OT [25] can also be extended to KGs, but have been outperformed by MTransE-LT on cross-lingual tasks.

3 Multi-graph Affinity Embeddings

We follow the definition of multilingual KGs in [6]. In a KB, $\mathcal{L}$ denotes the set of languages, and $\mathcal{L}^2$ denotes unordered language pairs. For each $L \in \mathcal{L}$, G_L denotes the language-specific KG of L, and E_L and R_L respectively denote the corresponding vocabularies of entities and relations. $T = (h, r, t)$ denotes a triple in G_L such that $h, t \in E_L$ and $r \in R_L$. Boldfaced $\mathbf{h}$, $\mathbf{r}$, $\mathbf{t}$ respectively represent the embedding vectors of head h, relation r, and tail t. For $(L_1, L_2) \in \mathcal{L}^2$, $\delta(L_1, L_2)$ denotes the alignment set which contains a small portion of already-aligned triple pairs between L_1 and L_2. For each G_L, a k-dimensional embedding space $\mathbb{R}_L^k$ is assigned for vectors of E_L and R_L, for which $\mathbb{R}$ is the field of real numbers.

3.1 Knowledge Model

KM preserves each language version of KG in a separated embedding space. Like previous works [6, 20, 26], we employ TransE as KM:

$$S_K = \sum_{L \in \{L_i, L_j\}} \sum_{(h,r,t) \in G_L} \|\mathbf{h} + \mathbf{r} - \mathbf{t}\|$$

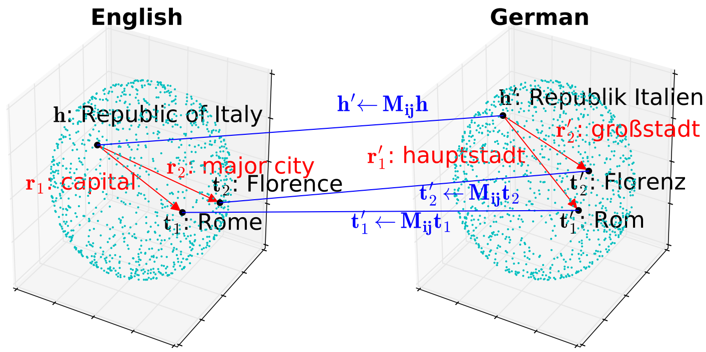

Figure 1: 3D PCA projection of MTransE-LT trained on WK3l-15k, which shows incoherence of the embedding spaces of English and German. In both embedding spaces, entity vectors are normalized to the unit hypersphere, whereas entity counterparts are embedded incoherently, and involved relation counterparts are captured with very unequal l_2 norms ($\|\mathbf{r_1}\| = 0.33$, $\|\mathbf{r'_1}\| = 0.59$, $\|\mathbf{r_2}\| = 0.49$, $\|\mathbf{r'_2}\| = 0.19$).

Unlike other translation-based models, TransE benefits cross-lingual tasks by representing embeddings uniformly in different contexts of relations. Although neural models like RESCAL and HolE may be used as well, in this paper, we choose to keep S_K as a controlled variable and focus on improving S_A (Alignment Model). Like MTransE, S_K does not adopt negative sampling, as we find it does not contribute to cross-lingual tasks.

Note that KM enforces norm contraint such that the l_2-norm of any entity vector is 1. This is an important constraint to prevent the learning process from reaching a trivial solution where all vectors collapse to zero, and is widely enforced by MTransE, other KG embeddings [10, 4, 3, 5, 21, 20], and multilingual word embeddings [25, 1].

3.2 Alignment Model

It is important for AM to be general as needed to capture well the comprehensive transitions between incoherent embeddings of cross-lingual counterparts. We have already discussed that, MTransE-LT learns a linear transform $\mathbf{M}_{ij}$ jointly with KM, such that $\mathbf{M}_{ij}\mathbf{x} \approx \mathbf{x}'$ given (x, x') as a pair of counterpart entities in L_i and L_j. As embedding vectors are normalized, it is easy to show that $\mathbf{M}_{ij}$ is expected to become a linear isometry. This technique outperforms the others due to higher generality. This also explains why, if we constrain $\mathbf{M}_{ij}$ to be orthogonal (OT), the performance of the model becomes worse as it is shown in Section 4.

However, we argue that, the AM adopted by MTransE-LT should be further generalized, as it lacks the ability to model various forms of invertible transforms, such as translation and scaling. The limitation of this technique hinders the precision of cross-lingual transitions, as an example is shown in Fig. 1 with a conflict on the embedded counterpart triples (h, r_1, t_1) and (h', r'_1, t'_1). Because $\mathbf{M}_{ij}\mathbf{h} \approx \mathbf{h}'$ and $\mathbf{M}_{ij}\mathbf{t} \approx \mathbf{t}'$, we have $\mathbf{M}_{ij}(\mathbf{h} - \mathbf{t}) \approx \mathbf{h}' - \mathbf{t}'$, i.e. $\mathbf{M}_{ij}\mathbf{r}_1 \approx \mathbf{r}'_1$ according to S_K. Since $\mathbf{M}_{ij}$ is expected to be isometry, $\|\mathbf{r}'_1\| \approx \|\mathbf{M}_{ij}\mathbf{r}_1\| = \|\mathbf{r}_1\|$ is conflict to the observation where $\|\mathbf{r}_1\|$ and $\|\mathbf{r}'_1\|$ are largely unequal. Such conflicts also occur to r_2 and r'_2, as well as all relation counterparts for which we observe a mean deviance of $l2$-norm that is as large as 0.38.

We hence define AM with a more generalized affine map as below.

$$\sigma_{ij}(\mathbf{x}) = \mathbf{A_{ij}}\mathbf{x} + \mathbf{b_{ij}} \; s.t. \; \mathbf{A_{ij}} \in \mathbb{R}^{k \times k}, \; \mathbf{b_{ij}} \in \mathbb{R}^k$$

The score function of AM is given as,

$$S_A = \sum_{(T,T') \in \delta(L_i, L_j)} S_a(T, T')$$

for which the alignment score S_a is defined as below, where σ^e_{ij} and σ^r_{ij} are the entity and relation-dedicated affine maps respectively.

$$S_a = \left\|\sigma^e_{ij}(\mathbf{h}) - \mathbf{h}'\right\| + \left\|\sigma^r_{ij}(\mathbf{r}) - \mathbf{r}'\right\| + \left\|\sigma^e_{ij}(\mathbf{t}) - \mathbf{t}'\right\|$$

Table 1: Statistics on the datasets. Two versions of WK3l are Wikipedia-based and CN3l is ConceptNet-based. Each dataset has small portions of alignment sets between English-French and English-German, and extra entity inter-lingual links (ILLs) to evaluate entity matching.

Dataset	#En triples	#Fr triples	#De triples	#Align triples	#ILLs
WK3l15k	203,502	170,605	145,616	Fr&En:16,470 De&En:37,170	Fr-En:3,815 De-En:1,610
WK3l120k	1,376,011	767,750	391,108	Fr&En:124,433 De&En:69,413	Fr-En:41,513 De-En:5,921
CN3l	47,696	18,624	25,560	Fr&En:3,668 De&En:8,588	Fr-En:2,146 De-En:3,813

Table 2: Cross-lingual entity matching result.

Dataset	CN3l						WK3l-15k						WK3l-120k			
Language	Fr-En			De-En			Fr-En			De-En			Fr-En		De-En	
Metrics	$H@1$	$H@10$	$Mean$	$H@1$	$H@10$	$Mean$	$H@1$	$H@10$	$Mean$	$H@1$	$H@10$	$Mean$	$H@1$	$H@10$	$H@1$	$H@10$
LM	2.52	20.16	1884.7	1.49	18.04	1487.9	1.56	10.42	3661.0	0.38	15.21	6114.1	1.02	14.26	0.34	13.58
CCA	3.81	26.40	1204.9	3.03	25.30	1740.8	3.78	19.44	3017.9	3.56	22.30	5855.6	1.60	12.85	1.62	20.39
OT	49.77	67.06	42.3	58.23	72.34	33.9	33.27	40.92	461.2	39.07	49.24	145.5	27.15	37.19	28.43	34.21
ITransE	67.56	71.19	186.2	67.80	84.01	107.6	55.33	58.42	690.2	41.39	52.3	127.1	26.79	38.09	49.47	50.57
MTransE-AC	30.04	69.27	55.2	19.37	63.56	33.6	22.21	46.64	436.5	25.53	50.60	167.0	9.54	36.52	23.17	47.79
MTransE-TV	41.05	77.02	29.3	39.21	70.96	14.8	7.57	36.44	464.6	29.86	52.16	151.8	11.49	36.45	34.99	52.24
MTransE-LT	80.57	86.05	**16.6**	77.92	95.67	7.8	58.78	61.52	199.6	63.58	68.53	42.3	38.98	47.43	58.31	67.75
MTransE-Af	**83.17**	**86.94**	16.9	**96.72**	**98.25**	**1.76**	**59.73**	**65.07**	**181.3**	**73.96**	**75.41**	**15.34**	**39.78**	**49.25**	**66.79**	**74.56**

The affine map is strong enough to model almost all regular forms of invertible vector transforms [13]. It is able to tolerate the vector differences of relation counterparts with a translation component, and does not require M_{ij} to be isometry. The affine map effectively enhances cross-lingual transitions for both entities and relations, as shown in the experiments.

3.3 Training

Training MTransE-Af is to minimize $J = S_K + \alpha S_A$ via on-line SGD, for which α is a hyperparameter. We initialize all vectors with uniform distribution on a unit hypersphere, and all matrices with random orthogonal initialization [18].

4 Experiments

In this section, we evaluate the proposed model on two cross-lingual tasks first introduced in [6]: cross-lingual entity matching, and triple alignment verification. Results are reported on three trilingual datasets as shown in Table 1.

4.1 Cross-lingual Entity Matching

The objective of this task is to match the same entities from different languages in KB. The ILLs are used as test cases, on which we aggregate three metrics, i.e. precision $H@1$ (%), the proportion of ranks no larger than 10 $H@10$ (%), and mean rank $Mean$. Higher $H@1$ and $H@10$, and lower $Mean$ indicate better outcomes.

MTransE-Af is compared against LM, CCA, OT, and three MTransE variants using the same settings in [6], as well as ITransE which enforces iterative parameter sharing between two TransE. To evaluate all models under controlled variables, for each dataset, we apply the configuration of MTransE to all models. In detail, we fix $\alpha = 5$, l_2 norm for all settings, while we use learning rate $\lambda = 0.001$ on CN3l, $\lambda = 0.01$ on WN3l datasets. k is set as 50, 75, 100 respectively on CN3l, WK3l15k, and WK3l120k. Norm threshold is set as $\theta = 1.0$ for ITransE to extend alignment seeds. Training is limited to 400 epochs on WK3l datasets and 200 epochs on CN3l.

Results are reported in Table 2. As expected, other baselines are substantially outperformed by the more flexible MTransE-LT. Between English-German graphs, MTransE-Af outperforms the results of MTransE-LT with 18.8%, 10.38%, and 8.48% increment of accuracy, as well as notably higher $H@10$ and lower $Mean$. The English-French settings are more consistent, and we see increments of accuracy ranging from 0.80% to 2.60%. In conclusion, the affine-map-based AM effectively

Table 3: Accuracy of triple alignment verification (%).

Dataset	CN3l		WK3l15k		WK3l120k	
Language	Fr&En	De&En	Fr&En	De&En	Fr&En	De&En
LM	60.53	51.55	52.23	63.61	59.98	59.98
CCA	81.57	79.54	52.28	66.49	65.89	61.01
OT	93.01	87.59	93.20	87.97	88.65	85.24
ITransE	76.40	78.35	89.45	87.87	81.77	82.54
MTransE-AC	93.92	91.89	93.25	91.24	91.27	91.35
MTransE-TV	88.95	84.80	90.38	84.24	87.99	87.04
MTransE-LT	97.46	96.63	94.58	95.03	93.48	93.06
MTransE-Af	**98.19**	**97.13**	**98.75**	**99.34**	**95.67**	**95.31**

Table 4: Results of tail and relation prediction (H@10).

Predict	Tail		Relation	
Language	En	Fr	En	Fr
TransE	**42.19**	25.06	61.79	62.55
MTransE-AC	40.37	23.45	60.18	60.73
MTransE-TV	40.97	22.26	58.32	59.44
MTransE-LT	41.03	25.46	**63.74**	64.77
MTransE-Af	42.07	**29.91**	62.42	**66.87**

improves cross-lingual matching of entity embeddings, and is more capable of handling multilingual inconsistency.

4.2 Triple Alignment Verification

This task is to produce a binary classifier to verify the correctness of triple alignments. We follow the steps in previous works, to create positive cases by isolating 20% of the alignment set, and corrupt the positive cases by (i) randomly replacing one of the six elements, or (ii) randomly substituting one of the two triples with a false alignment. Cases (i) and (ii) contribute negative cases that are as many as 100% and 50% of positive cases respectively. We use 10-fold cross-validation to train and evaluate a simple classifier that finds a threshold τ on the dissimilarity score S_a (Section 3.2) for a given triple alignment, for which $S_a < \tau$ implies positive, otherwise negative. The value of τ is decided by maximizing the accuracy on the training cases. This simple classifier requests the embedding model to precisely represent cross-lingual transitions for the entire triples. Model configurations from the pervious experiment are carried forward to evaluate under controlled variables.

Table 3 reports the results. Although MTransE-LT has already shown satisfying results that are better than the rest models, MTransE-Af further increases the accuracy by 0.50%-0.73% on CN3l, 4.17%-4.31% on WK3l15k, and 2.19%-2.25% on WK3l120k. This indicates that the enhancement by the affine map on learning cross-lingual transitions is effective on both entities and relations.

4.3 Monolingual Tasks

Besides the above two tasks, we also evaluate MTransE-Af on monolingual tasks of tail prediction (i.e. predicting t given h and r) and relation prediction (predicting r given h and t) using the English and French versions of WK3l15k. Like previous works, [3, 24, 12], for each language version, 10% triples are selected as the test set, and the remaining becomes the training set. We compare with TransE and MTransE on these two tasks. For training, MTransE-Af and MTransE variants are trained upon both language versions of the training set for the knowledge model, while the intersection between the alignment set and the training set is used for the alignment models. TransE is trained separately on each training set.

The results for H@10 reported in Table 4 show that MTransE-Af performs at least as well as MTransE variants and its monolingual counterpart TransE, and even better in some cases due to the correlation of the two languages. This indicates that MTransE-Af preserves well the structure of monolingual KGs, and affine-map-based AM does not interfere the learning of KM.

5 Conclusion and Future Work

In this paper, we have proposed a multilingual KG embedding model, which addresses the limitations of previous models in handling the multilingual inconsistency of KGs by introducing a generalized affine-map-based alignment technique. Experimental results show the effectiveness of the model for both entity and triple-wise knowledge alignment, and precise encoding of monolingual knowledge. For future work, we plan to strengthen the weakly supervision of multilingual learning process with more information, such as text and structural entity descriptions.

References

[1] Mikel Artetxe, Gorka Labaka, and Eneko Agirre. Learning principled bilingual mappings of word embeddings while preserving monolingual invariance. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2289–2294, 2016.

[2] Francis Bond. Linking and extending an open multilingual Wordnet. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 1352–1362, 2013.

[3] Antoine Bordes, Nicolas Usunier, et al. Translating embeddings for modeling multi-relational data. In *Advances in Neural Information Processing Systems (NIPS)*, pages 2787–2795, 2013.

[4] Antoine Bordes, Jason Weston, Ronan Collobert, and Yoshua Bengio. Learning structured embeddings of knowledge bases. In *25th Conference on Artificial Intelligence (AAAI)*, pages 301–306, 2011.

[5] Antoine Bordes, Jason Weston, and Nicolas Usunier. Open question answering with weakly supervised embedding models. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases (ECML-PKDD)*, pages 165–180. Springer, 2014.

[6] Muhao Chen, Yingtao Tian, et al. Multilingual knowledge graph embeddings for cross-lingual knowledge alignment. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2017.

[7] Muhao Chen and Carlo Zaniolo. Learning multi-faceted knowledge graph embeddings for natural language processing. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2017.

[8] Yuan Fang, Kingsley Kuan, Jie Lin, Cheston Tan, and Vijay Chandrasekhar. Object detection meets knowledge graphs. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2017.

[9] Manaal Faruqui and Chris Dyer. Improving vector space word representations using multilingual correlation. *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, 2014.

[10] Rodolphe Jenatton, Nicolas L Roux, Antoine Bordes, and Guillaume R Obozinski. A latent factor model for highly multi-relational data. In *Advances in Neural Information Processing Systems (NIPS)*, pages 3167–3175, 2012.

[11] Guoliang Ji, Shizhu He, et al. Knowledge graph embedding via dynamic mapping matrix. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 687–696, 2015.

[12] Yantao Jia, Yuanzhuo Wang, Hailun Lin, et al. Locally adaptive translation for knowledge graph embedding. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, pages 992–998, 2016.

[13] Thomas F Jordan. Affine maps of density matrices. *Physical Review A*, 71(3), 2005.

[14] Jens Lehmann, Robert Isele, Max Jakob, Anja Jentzsch, et al. Dbpedia–a large-scale, multilingual knowledge base extracted from Wikipedia. *Semantic Web*, 6(2):167–195, 2015.

[15] Yankai Lin, Zhiyuan Liu, et al. Learning entity and relation embeddings for knowledge graph completion. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, pages 2181–2187, 2015.

[16] Maximilian Nickel, Lorenzo Rosasco, et al. Holographic embeddings of knowledge graphs. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, pages 1955–1961, 2016.

[17] Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. A three-way model for collective learning on multi-relational data. In *Proceedings of the 28th international conference on machine learning (ICML)*, pages 809–816, 2011.

[18] Andrew M Saxe, James L McClelland, and Surya Ganguli. Exact solutions to the nonlinear dynamics of learning in deep linear neural networks. *International Conference on Learning Representations (ICLR)*, 2014.

[19] Robert Speer, Joshua Chin, and Catherine Havasi. Conceptnet 5.5: An open multilingual graph of general knowledge. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, pages 4444–4451, 2017.

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

[21] Théo Trouillon, Johannes Welbl, et al. Complex embeddings for simple link prediction. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 2071–2080, 2016.

[22] Denny Vrandečić. Wikidata: a free collaborative knowledgebase. *Communications of the ACM*, 57(10):78–85, 2014.

[23] Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. Knowledge graph and text jointly embedding. In *Proceedings of the Conference on Empirical Methods on Natural Language Processing (EMNLP)*, pages 1591–1601, 2014.

[24] Zhen Wang, Jianwen Zhang, Jianlin Feng, et al. Knowledge graph embedding by translating on hyperplanes. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, pages 1112–1119, 2014.

[25] Chao Xing, Dong Wang, Chao Liu, and Yiye Lin. Normalized word embedding and orthogonal transform for bilingual word translation. In *Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL HLT)*, pages 1006–1011, 2015.

[26] Hao Zhu, Ruobing Xie, et al. Iterative entity alignment via knowledge embeddings. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2017.