A Novel Embedding Model for Knowledge Base Completion Based on Convolutional Neural Network

Dai Quoc Nguyen[1], Tu Dinh Nguyen[1], Dat Quoc Nguyen[2], Dinh Phung[1]

[1] PRaDA Centre, Deakin University, Australia

{dai.nguyen,tu.nguyen,dinh.phung}@deakin.edu.au

[2] The University of Melbourne, Australia

dqnguyen@unimelb.edu.au

Abstract

In this paper, we propose a novel embedding model, named ConvKB, for knowledge base completion. Our model ConvKB advances state-of-the-art models by employing a convolutional neural network, so that it can capture global relationships and transitional characteristics between entities and relations in knowledge bases. In ConvKB, each triple *(head entity, relation, tail entity)* is represented as a 3-column matrix where each column vector represents a triple element. This 3-column matrix is then fed to a convolution layer where multiple filters are operated on the matrix to generate different feature maps. These feature maps are then concatenated into a single feature vector representing the input triple. The feature vector is multiplied with a weight vector via a dot product to return a score. This score is then used to predict whether the triple is valid or not. Experiments show that ConvKB achieves better link prediction performance than previous state-of-the-art embedding models on two benchmark datasets WN18RR and FB15k-237.

1 Introduction

Large-scale knowledge bases (KBs), such as YAGO (Suchanek et al., 2007), Freebase (Bollacker et al., 2008) and DBpedia (Lehmann et al., 2015), are usually databases of triples representing the relationships between entities in the form of fact *(head entity, relation, tail entity)* denoted as *(h, r, t)*, e.g., *(Melbourne, cityOf, Australia)*. These KBs are useful resources in many applications such as semantic searching and ranking (Kasneci et al., 2008; Schuhmacher and Ponzetto, 2014; Xiong et al., 2017), question answering (Zhang et al., 2016; Hao et al., 2017) and machine reading (Yang and Mitchell, 2017). However, the KBs are still incomplete, i.e., missing a lot of valid triples (Socher et al., 2013; West et al., 2014). Therefore, much research work has been devoted towards *knowledge base completion* or *link prediction* to predict whether a triple *(h, r, t)* is valid or not (Bordes et al., 2011).

Many embedding models have proposed to learn vector or matrix representations for entities and relations, obtaining state-of-the-art (SOTA) link prediction results (Nickel et al., 2016a). In these embedding models, valid triples obtain lower implausibility scores than invalid triples. Let us take the well-known embedding model TransE (Bordes et al., 2013) as an example. In TransE, entities and relations are represented by k-dimensional vector embeddings. TransE employs a transitional characteristic to model relationships between entities, in which it assumes that if *(h, r, t)* is a valid fact, the embedding of head entity h plus the embedding of relation r should be close to the embedding of tail entity t, i.e. $\boldsymbol{v}_h + \boldsymbol{v}_r \approx \boldsymbol{v}_t$ (here, $\boldsymbol{v}_h$, $\boldsymbol{v}_r$ and $\boldsymbol{v}_t$ are embeddings of h, r and t respectively). That is, a TransE score $\|\boldsymbol{v}_h + \boldsymbol{v}_r - \boldsymbol{v}_t\|_p^p$ of the valid triple *(h, r, t)* should be close to 0 and smaller than a score $\|\boldsymbol{v}_{h'} + \boldsymbol{v}_{r'} - \boldsymbol{v}_{t'}\|_p^p$ of an invalid triple *(h', r', t')*. The transitional characteristic in TransE also implies the global relationships among same dimensional entries of $\boldsymbol{v}_h$, $\boldsymbol{v}_r$ and $\boldsymbol{v}_t$.

Other transition-based models extend TransE to additionally use projection vectors or matrices to translate head and tail embeddings into the relation vector space, such as: TransH (Wang et al., 2014), TransR (Lin et al., 2015b), TransD (Ji et al., 2015), STransE (Nguyen et al., 2016b) and TranSparse (Ji et al., 2016). Furthermore, DISTMULT (Yang et al., 2015) and ComplEx (Trouillon et al., 2016) use a tri-linear dot product to compute the score for each triple. Recent research

has shown that using relation paths between entities in the KBs could help to get contextual information for improving KB completion performance (Lin et al., 2015a; Luo et al., 2015; Guu et al., 2015; Toutanova et al., 2016; Nguyen et al., 2016a). See other embedding models for KB completion in Nguyen (2017).

Recently, convolutional neural networks (CNNs), originally designed for computer vision (LeCun et al., 1998), have significantly received research attention in natural language processing (Collobert et al., 2011; Kim, 2014). CNN learns non-linear features to capture complex relationships with a remarkably less number of parameters compared to fully connected neural networks. Inspired from the success in computer vision, Dettmers et al. (2018) proposed ConvE—the first model applying CNN for the KB completion task. In ConvE, only v_h and v_r are reshaped and then concatenated into an input matrix which is fed to the convolution layer. Different filters of the same 3×3 shape are operated over the input matrix to output feature map tensors. These feature map tensors are then vectorized and mapped into a vector via a linear transformation. Then this vector is computed with v_t via a dot product to return a score for *(h, r, t)*. See a formal definition of the ConvE score function in Table 1. It is worth noting that ConvE focuses on the local relationships among different dimensional entries in each of v_h or v_r, i.e., *ConvE does not observe the global relationships among same dimensional entries of an embedding triple (v_h, v_r, v_t)*, so that ConvE ignores the transitional characteristic in transition-based models, which is one of the most useful intuitions for the task.

In this paper, we present ConvKB—an embedding model which proposes a novel use of CNN for the KB completion task. In ConvKB, each entity or relation is associated with an unique k-dimensional embedding. Let v_h, v_r and v_t denote k-dimensional embeddings of h, r and t, respectively. For each triple *(h, r, t)*, the corresponding triple of k-dimensional embeddings (v_h, v_r, v_t) is represented as a $k \times 3$ input matrix. This input matrix is fed to the convolution layer where different filters of the same 1×3 shape are used to extract the global relationships among same dimensional entries of the embedding triple. That is, these filters are repeatedly operated over every row of the input matrix to produce different

Model	The score function $f(h, r, t)$
TransE	$\|v_h + v_r - v_t\|_p^p$
DISTMULT	$\langle v_h, v_r, v_t \rangle$
ComplEx	$\mathsf{Re}\left(\langle v_h, v_r, \overline{v}_t \rangle\right)$
ConvE	$g\left(\mathsf{vec}\left(g\left(\mathsf{concat}\left(\widehat{v}_h, \widehat{v}_r\right) * \boldsymbol{\Omega}\right)\right) \boldsymbol{W}\right) \cdot v_t$
ConvKB	$\mathsf{concat}\left(g\left([v_h, v_r, v_t] * \boldsymbol{\Omega}\right)\right) \cdot \mathbf{w}$

Table 1: The score functions in previous SOTA models and in our ConvKB model. $\|v\|_p$ denotes the p-norm of v. $\langle v_h, v_r, v_t \rangle = \sum_i v_{h_i} v_{r_i} v_{t_i}$ denotes a tri-linear dot product. g denotes a non-linear function. $*$ denotes a convolution operator. $\cdot$ denotes a dot product. concat denotes a concatenation operator. $\widehat{v}$ denotes a 2D reshaping of v. $\boldsymbol{\Omega}$ denotes a set of filters.

feature maps. The feature maps are concatenated into a single feature vector which is then computed with a weight vector via a dot product to produce a score for the triple *(h, r, t)*. This score is used to infer whether the triple *(h, r, t)* is valid or not.

Our contributions in this paper are as follows:

- We introduce ConvKB—a novel embedding model of entities and relationships for knowledge base completion. ConvKB models the relationships among same dimensional entries of the embeddings. This implies that ConvKB generalizes transitional characteristics in transition-based embedding models.

- We evaluate ConvKB on two benchmark datasets: WN18RR (Dettmers et al., 2018) and FB15k-237 (Toutanova and Chen, 2015). Experimental results show that ConvKB obtains better link prediction performance than previous SOTA embedding models. In particular, ConvKB obtains the best mean rank and the highest Hits@10 on WN18RR, and produces the highest mean reciprocal rank and highest Hits@10 on FB15k-237.

2   Proposed ConvKB model

A knowledge base $\mathcal{G}$ is a collection of valid factual triples in the form of *(head entity, relation, tail entity)* denoted as (h, r, t) such that $h, t \in \mathcal{E}$ and $r \in \mathcal{R}$ where $\mathcal{E}$ is a set of entities and $\mathcal{R}$ is a set of relations. Embedding models aim to define a *score function f* giving an implausibility score for each triple (h, r, t) such that valid triples receive lower scores than invalid triples. Table 1 presents score functions in previous SOTA models.

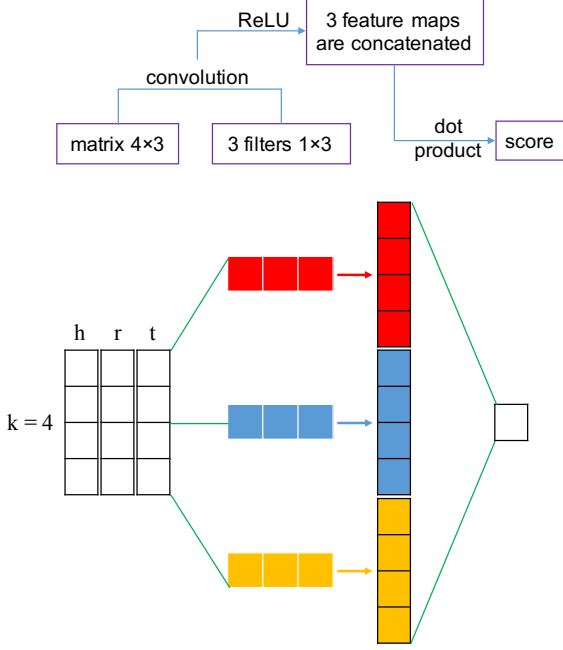We denote the dimensionality of embeddings by k such that each embedding triple (v_h, v_r, v_t) are

Figure 1: Process involved in ConvKB (with the embedding size $k = 4$, the number of filters $\tau = 3$ and the activation function $g =$ ReLU for illustration purpose).

viewed as a matrix $\boldsymbol{A} = [\boldsymbol{v}_h, \boldsymbol{v}_r, \boldsymbol{v}_t] \in \mathbb{R}^{k \times 3}$. And $\boldsymbol{A}_{i,:} \in \mathbb{R}^{1 \times 3}$ denotes the i-th row of $\boldsymbol{A}$. Suppose that we use a filter $\boldsymbol{\omega} \in \mathbb{R}^{1 \times 3}$ operated on the convolution layer. $\boldsymbol{\omega}$ is not only aimed to examine the global relationships between same dimensional entries of the embedding triple $(\boldsymbol{v}_h, \boldsymbol{v}_r, \boldsymbol{v}_t)$, but also to generalize the transitional characteristics in the transition-based models. $\boldsymbol{\omega}$ is repeatedly operated over every row of $\boldsymbol{A}$ to finally generate a feature map $\boldsymbol{v} = [v_1, v_2, ..., v_k] \in \mathbb{R}^k$ as:

$$v_i = g\left(\boldsymbol{\omega} \cdot \boldsymbol{A}_{i,:} + b\right)$$

where $b \in \mathbb{R}$ is a bias term and g is some activation function such as ReLU.

Our ConvKB uses different filters $\in \mathbb{R}^{1 \times 3}$ to generate different feature maps. Let $\boldsymbol{\Omega}$ and τ denote the set of filters and the number of filters, respectively, i.e. $\tau = |\boldsymbol{\Omega}|$, resulting in τ feature maps. These τ feature maps are concatenated into a single vector $\in \mathbb{R}^{\tau k \times 1}$ which is then computed with a weight vector $\mathbf{w} \in \mathbb{R}^{\tau k \times 1}$ via a dot product to give a score for the triple (h, r, t). Figure 1 illustrates the computation process in ConvKB.

Formally, we define the ConvKB score function f as follows:

$$f(h, r, t) = \mathsf{concat}\left(g\left([\boldsymbol{v}_h, \boldsymbol{v}_r, \boldsymbol{v}_t] * \boldsymbol{\Omega}\right)\right) \cdot \mathbf{w}$$

where $\boldsymbol{\Omega}$ and $\mathbf{w}$ are shared parameters, indepen-

Dataset	$|\mathcal{E}|$	$|\mathcal{R}|$	#Triples in train/valid/test		
WN18RR	40,943	11	86,835	3,034	3,134
FB15k-237	14,541	237	272,115	17,535	20,466

Table 2: Statistics of the experimental datasets.

dent of h, r and t; $*$ denotes a convolution operator; and concat denotes a concatenation operator.

If we only use one filter $\boldsymbol{\omega}$ (i.e. using $\tau = 1$) with a fixed bias term $b = 0$ and the activation function $g(x) = |x|$ or $g(x) = x^2$, and fix $\boldsymbol{\omega} = [1, 1, -1]$ and $\mathbf{w} = \mathbf{1}$ during training, ConvKB reduces to the plain TransE model (Bordes et al., 2013). So our ConvKB model can be viewed as an extension of TransE to further model global relationships.

We use the Adam optimizer (Kingma and Ba, 2014) to train ConvKB by minimizing the loss function $\mathcal{L}$ (Trouillon et al., 2016) with L_2 regularization on the weight vector $\mathbf{w}$ of the model:

$$\mathcal{L} = \sum_{(h,r,t) \in \{\mathcal{G} \cup \mathcal{G}'\}} \log\left(1 + \exp\left(l_{(h,r,t)} \cdot f(h, r, t)\right)\right)$$
$$+ \frac{\lambda}{2}\|\mathbf{w}\|_2^2$$

$$\text{in which, } l_{(h,r,t)} = \begin{cases} 1 & \text{for } (h,r,t) \in \mathcal{G} \\ -1 & \text{for } (h,r,t) \in \mathcal{G}' \end{cases}$$

here $\mathcal{G}'$ is a collection of invalid triples generated by corrupting valid triples in $\mathcal{G}$.

3 Experiments

3.1 Datasets

We evaluate ConvKB on two benchmark datasets: WN18RR (Dettmers et al., 2018) and FB15k-237 (Toutanova and Chen, 2015). WN18RR and FB15k-237 are correspondingly subsets of two common datasets WN18 and FB15k (Bordes et al., 2013). As noted by Toutanova and Chen (2015), WN18 and FB15k are easy because they contain many reversible relations. So knowing relations are reversible allows us to easily predict the majority of test triples, e.g. state-of-the-art results on both WN18 and FB15k are obtained by using a simple reversal rule as shown in Dettmers et al. (2018). Therefore, WN18RR and FB15k-237 are created to not suffer from this reversible relation problem in WN18 and FB15k, for which the knowledge base completion task is more realistic. Table 2 presents the statistics of WN18RR and FB15k-237.

Method	WN18RR			FB15k-237		
	MR	MRR	H@10	MR	MRR	H@10
IRN (Shen et al., 2017)	–	–	–	<u>211</u>	–	46.4
KBGAN (Cai and Wang, 2018)	–	0.213	48.1	–	0.278	45.8
DISTMULT (Yang et al., 2015) [★]	5110	0.43	49	254	0.241	41.9
ComplEx (Trouillon et al., 2016) [★]	5261	<u>0.44</u>	<u>51</u>	339	0.247	42.8
ConvE (Dettmers et al., 2018)	5277	**0.46**	48	246	<u>0.316</u>	49.1
TransE (Bordes et al., 2013) (our results)	<u>3384</u>	0.226	50.1	347	0.294	46.5
Our ConvKB model	**2554**	0.248	**52.5**	257	**0.396**	**51.7**
KB$_{LRN}$ (García-Durán and Niepert, 2017)	–	–	–	**209**	0.309	<u>49.3</u>
R-GCN+ (Schlichtkrull et al., 2017)	–	–	–	–	0.249	41.7
Neural LP (Yang et al., 2017)	–	–	–	–	0.240	36.2
Node+LinkFeat (Toutanova and Chen, 2015)	–	–	–	–	0.293	46.2

Table 3: Experimental results on WN18RR and FB15k-237 test sets. MRR and H@10 denote the mean reciprocal rank and Hits@10 (in %), respectively. [★]: Results are taken from Dettmers et al. (2018) where Hits@10 and MRR are rounded to 2 decimal places on WN18RR. The last 4 rows report results of models that exploit information about relation paths (KB$_{LRN}$, R-GCN+ and Neural LP) or textual mentions derived from a large external corpus (Node+LinkFeat). The best score is in **bold**, while the second best score is in <u>underline</u>.

3.2 Evaluation protocol

In the KB completion or link prediction task (Bordes et al., 2013), the purpose is to predict a missing entity given a relation and another entity, i.e, inferring h given (r, t) or inferring t given (h, r). The results are calculated based on ranking the scores produced by the score function f on test triples.

Following Bordes et al. (2013), for each valid test triple (h, r, t), we replace either h or t by each of other entities in $\mathcal{E}$ to create a set of corrupted triples. We use the "**Filtered**" setting protocol (Bordes et al., 2013), i.e., not taking any corrupted triples that appear in the KB into accounts. We rank the valid test triple and corrupted triples in ascending order of their scores. We employ three common evaluation metrics: mean rank (MR), mean reciprocal rank (MRR), and Hits@10 (i.e., the proportion of the valid test triples ranking in top 10 predictions). Lower MR, higher MRR or higher Hits@10 indicate better performance.

3.3 Training protocol

We use the common Bernoulli trick (Wang et al., 2014; Lin et al., 2015b) to generate the head or tail entities when sampling invalid triples. We also use entity and relation embeddings produced by TransE to *initialize* entity and relation embeddings in ConvKB. We employ a TransE implementation available at: https://github.com/datquocnguyen/STransE. We train TransE for 3,000 epochs, using a grid search of hyper-parameters: the dimensionality of embeddings $k \in \{50, 100\}$, SGD learning rate $\in \{1e^{-4}, 5e^{-4}, 1e^{-3}, 5e^{-3}\}$, l_1-norm or l_2-norm, and margin $\gamma \in \{1, 3, 5, 7\}$. The highest Hits@10 scores on the validation set are when using l_1-norm, learning rate at $5e^{-4}$, $\gamma = 5$ and $k = 50$ for WN18RR, and using l_1-norm, learning rate at $5e^{-4}$, $\gamma = 1$ and k = 100 for FB15k-237.

To learn our model parameters including entity and relation embeddings, filters $\boldsymbol{\omega}$ and the weight vector $\mathbf{w}$, we use Adam (Kingma and Ba, 2014) and select its initial learning rate $\in \{5e^{-6}, 1e^{-5}, 5e^{-5}, 1e^{-4}, 5e^{-4}\}$. We use ReLU as the activation function g. We fix the batch size at 256 and set the L_2-regularizer λ at 0.001 in our objective function. The filters $\boldsymbol{\omega}$ are initialized by a truncated normal distribution or by $[0.1, 0.1, -0.1]$. We select the number of filters $\tau \in \{50, 100, 200, 400, 500\}$. We run ConvKB up to 200 epochs and use outputs from the last epoch for evaluation. The highest Hits@10 scores on the validation set are obtained when using $k = 50$, $\tau = 500$, the truncated normal distribution for filter initialization, and the initial learning rate at $1e^{-4}$ on WN18RR; and k = 100, $\tau = 50$, $[0.1, 0.1, -0.1]$ for filter initialization, and the initial learning rate at $5e^{-6}$ on FB15k-237.

3.4 Main experimental results

Table 3 compares the experimental results of our ConvKB model with previous published results, using the same experimental setup. Table 3 shows that ConvKB obtains the best MR and highest Hits@10 scores on WN18RR and also the highest

MRR and Hits@10 scores on FB15k-237.

ConvKB does better than the closely related model TransE on both experimental datasets, especially on FB15k-237 where ConvKB gains significant improvements of $347 - 257 = 90$ in MR (which is about 26% relative improvement) and $0.396 - 0.294 = 0.102$ in MRR (which is 34+% relative improvement), and also obtains $51.7 - 46.5 = 5.2\%$ absolute improvement in Hits@10. Previous work shows that TransE obtains very competitive results (Lin et al., 2015a; Nickel et al., 2016b; Trouillon et al., 2016; Nguyen et al., 2016a). However, when comparing the CNN-based embedding model ConvE with other models, Dettmers et al. (2018) did not experiment with TransE. We reconfirm previous findings that TransE in fact is a strong baseline model, e.g., TransE obtains better MR and Hits@10 than ConvE on WN18RR.

ConvKB obtains better scores than ConvE on both datasets (except MRR on WN18RR and MR on FB15k-237), thus showing the usefulness of taking transitional characteristics into accounts. In particular, on FB15k-237, ConvKB achieves improvements of $0.394 - 0.316 = 0.078$ in MRR (which is about 25% relative improvement) and $51.7 - 49.1 = 2.6\%$ in Hits@10, while both ConvKB and ConvE produce similar MR scores. ConvKB also obtains 25% relatively higher MRR score than the relation path-based model KB_{LRN} on FB15k-237. In addition, ConvKB gives better Hits@10 than KB_{LRN}, however, KB_{LRN} gives better MR than ConvKB. We plan to extend ConvKB with relation path information to obtain better link prediction performance in future work.

4 Conclusion

In this paper, we propose a novel embedding model ConvKB for the knowledge base completion task. ConvKB applies the convolutional neural network to explore the global relationships among same dimensional entries of the entity and relation embeddings, so that ConvKB generalizes the transitional characteristics in the transition-based embedding models. Experimental results show that our model ConvKB outperforms other state-of-the-art models on two benchmark datasets WN18RR and FB15k-237. Our code is available at: https://github.com/daiquocnguyen/ConvKB.

We also plan to extend ConvKB for a new application where we could formulate data in the form of triples. For example, inspired from the work by Vu et al. (2017) for search personalization, we can also apply ConvKB to model *user*-oriented relationships between submitted *queries* and *documents* returned by search engines, i.e. modeling triple representations (query, user, document).

References

Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. 2008. Freebase: A collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*, pages 1247–1250.

Antoine Bordes, Nicolas Usunier, Alberto García-Durán, Jason Weston, and Oksana Yakhnenko. 2013. Translating Embeddings for Modeling Multi-relational Data. In *Advances in Neural Information Processing Systems 26*, pages 2787–2795.

Antoine Bordes, Jason Weston, Ronan Collobert, and Yoshua Bengio. 2011. Learning Structured Embeddings of Knowledge Bases. In *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence*, pages 301–306.

Liwei Cai and William Yang Wang. 2018. KBGAN: Adversarial Learning for Knowledge Graph Embeddings. In *Proceedings of The 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, page to appear.

Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. 2011. Natural language processing (almost) from scratch. *Journal of Machine Learning Research*, 12:2493–2537.

Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. 2018. Convolutional 2D Knowledge Graph Embeddings. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*, page to appear.

Alberto García-Durán and Mathias Niepert. 2017. KBLRN: End-to-end learning of knowledge base representations with latent, relational, and numerical features. *arXiv preprint abs/1709.04676*.

Kelvin Guu, John Miller, and Percy Liang. 2015. Traversing Knowledge Graphs in Vector Space. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 318–327.

Yanchao Hao, Yuanzhe Zhang, Kang Liu, Shizhu He, Zhanyi Liu, Hua Wu, and Jun Zhao. 2017. An end-to-end model for question answering over knowledge base with cross-attention combining global knowledge. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 221–231.

Guoliang Ji, Shizhu He, Liheng Xu, Kang Liu, and Jun Zhao. 2015. Knowledge Graph Embedding via Dynamic Mapping Matrix. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 687–696.

Guoliang Ji, Kang Liu, Shizhu He, and Jun Zhao. 2016. Knowledge Graph Completion with Adaptive Sparse Transfer Matrix. In *Proceedings of the Thirtieth Conference on Artificial Intelligence*, pages 985–991.

Gjergji Kasneci, Fabian M Suchanek, Georgiana Ifrim, Maya Ramanath, and Gerhard Weikum. 2008. Naga: Searching and ranking knowledge. In *Proceedings of the 24th IEEE International Conference on Data Engineering*, pages 953–962.

Yoon Kim. 2014. Convolutional Neural Networks for Sentence Classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pages 1746–1751.

Diederik Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86:2278–2324.

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*, 6:167–195.

Yankai Lin, Zhiyuan Liu, Huanbo Luan, Maosong Sun, Siwei Rao, and Song Liu. 2015a. Modeling Relation Paths for Representation Learning of Knowledge Bases. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 705–714.

Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015b. Learning Entity and Relation Embeddings for Knowledge Graph Completion. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence Learning*, pages 2181–2187.

Yuanfei Luo, Quan Wang, Bin Wang, and Li Guo. 2015. Context-Dependent Knowledge Graph Embedding. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1656–1661.

Dat Quoc Nguyen. 2017. An overview of embedding models of entities and relationships for knowledge base completion. *arXiv preprint*, arXiv:1703.08098.

Dat Quoc Nguyen, Kairit Sirts, Lizhen Qu, and Mark Johnson. 2016a. Neighborhood Mixture Model for Knowledge Base Completion. In *Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning*, pages 40–50.

Dat Quoc Nguyen, Kairit Sirts, Lizhen Qu, and Mark Johnson. 2016b. STransE: a novel embedding model of entities and relationships in knowledge bases. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 460–466.

Maximilian Nickel, Kevin Murphy, Volker Tresp, and Evgeniy Gabrilovich. 2016a. A Review of Relational Machine Learning for Knowledge Graphs. *Proceedings of the IEEE*, 104(1):11–33.

Maximilian Nickel, Lorenzo Rosasco, and Tomaso Poggio. 2016b. Holographic Embeddings of Knowledge Graphs. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, pages 1955–1961.

Michael Schlichtkrull, Thomas Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. 2017. Modeling relational data with graph convolutional networks. *arXiv preprint arXiv:1703.06103*.

Michael Schuhmacher and Simone Paolo Ponzetto. 2014. Knowledge-based graph document modeling. In *Proceedings of the 7th ACM International Conference on Web Search and Data Mining*, pages 543–552.

Yelong Shen, Po-Sen Huang, Ming-Wei Chang, and Jianfeng Gao. 2017. Traversing knowledge graph in vector space without symbolic space guidance. *arXiv preprint arXiv:1611.04642v4*.

Richard Socher, Danqi Chen, Christopher D Manning, and Andrew Ng. 2013. Reasoning With Neural Tensor Networks for Knowledge Base Completion. In *Advances in Neural Information Processing Systems 26*, pages 926–934.

Fabian M. Suchanek, Gjergji Kasneci, and Gerhard Weikum. 2007. Yago: A core of semantic knowledge. In *Proceedings of the 16th International Conference on World Wide Web*, pages 697–706.

Kristina Toutanova and Danqi Chen. 2015. Observed Versus Latent Features for Knowledge Base and Text Inference. In *Proceedings of the 3rd Workshop on Continuous Vector Space Models and their Compositionality*, pages 57–66.

Kristina Toutanova, Xi Victoria Lin, Wen tau Yih, Hoifung Poon, and Chris Quirk. 2016. Compositional Learning of Embeddings for Relation Paths in Knowledge Bases and Text. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pages 1434–1444.

Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. 2016. Complex Embeddings for Simple Link Prediction. In *Proceedings of the 33nd International Conference on Machine Learning*, pages 2071–2080.

Thanh Vu, Dat Quoc Nguyen, Mark Johnson, Dawei Song, and Alistair Willis. 2017. Search Personalization with Embeddings. In *Proceedings of the 39th European Conference on Information Retrieval*, pages 598–604.

Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. 2014. Knowledge Graph Embedding by Translating on Hyperplanes. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, pages 1112–1119.

Robert West, Evgeniy Gabrilovich, Kevin Murphy, Shaohua Sun, Rahul Gupta, and Dekang Lin. 2014. Knowledge Base Completion via Search-based Question Answering. In *Proceedings of the 23rd International Conference on World Wide Web*, pages 515–526.

Chenyan Xiong, Russell Power, and Jamie Callan. 2017. Explicit semantic ranking for academic search via knowledge graph embedding. In *Proceedings of the 26th International Conference on World Wide Web*, pages 1271–1279.

Bishan Yang and Tom Mitchell. 2017. Leveraging Knowledge Bases in LSTMs for Improving Machine Reading. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1436–1446.

Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. 2015. Embedding Entities and Relations for Learning and Inference in Knowledge Bases. In *Proceedings of the International Conference on Learning Representations*.

Fan Yang, Zhilin Yang, and William W Cohen. 2017. Differentiable Learning of Logical Rules for Knowledge Base Reasoning. In *Advances in Neural Information Processing Systems 30*, pages 2316–2325.

Yuanzhe Zhang, Kang Liu, Shizhu He, Guoliang Ji, Zhanyi Liu, Hua Wu, and Jun Zhao. 2016. Question answering over knowledge base with neural attention combining global knowledge information. *arXiv preprint arXiv:1606.00979*.