

# A Convolutional Neural Network-based Model for Knowledge Base Completion and Its Application to Search Personalization

Dai Quoc Nguyen<sup>a,\*</sup>, Dat Quoc Nguyen<sup>b</sup>, Tu Dinh Nguyen<sup>a</sup> and Dinh Phung<sup>a</sup>

<sup>a</sup> Monash University, Australia

E-mails: dai.nguyen@monash.edu, tu.dinh.nguyen@monash.edu, dinh.phung@monash.edu

<sup>b</sup> The University of Melbourne, Australia

E-mail: dqnguyen@unimelb.edu.au

**Editors:** Luis Espinosa Anke, Cardiff University, UK; Thierry Declerck, DFKI GmbH, Germany; Dagmar Gromann, Technical University Dresden, Germany

**Solicited reviews:** Sergio Oramas, Pompeu Fabra University, Spain; two anonymous reviewers

**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 obtains better link prediction and triple classification results than previous state-of-the-art models on benchmark datasets WN18RR, FB15k-237, WN11 and FB13. We further apply our ConvKB to a search personalization problem which aims to tailor the search results to each specific user based on the user's personal interests and preferences. In particular, we model the potential relationship between the submitted query, the user and the search result (i.e., document) as a triple (*query, user, document*) on which the ConvKB is able to work. Experimental results on query logs from a commercial web search engine show that ConvKB achieves better performances than the standard ranker as well as strong search personalization baselines.

**Keywords:** Knowledge base completion, Convolutional neural network, ConvKB, Link prediction, Triple classification, Search personalization

## 1. Introduction

Large-scale knowledge bases (KBs), such as YAGO [41], Freebase [3] and DBpedia [25], 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 [21, 38, 60], question answering [17, 66] and machine reading [61]. However, the KBs are still incomplete, i.e., missing a lot of valid triples [40, 55]. 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 [4].

Many embedding models have proposed to learn vector or matrix representations for entities and relations, obtaining state-of-the-art (SOTA) link prediction

---

\*Corresponding author. E-mail: dai.nguyen@monash.edu.

results [35]. In these embedding models, valid triples obtain lower implausibility scores than invalid triples. Let us take the well-known embedding model TransE [5] 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.  $\mathbf{v}_h + \mathbf{v}_r \approx \mathbf{v}_t$  (here,  $\mathbf{v}_h, \mathbf{v}_r$  and  $\mathbf{v}_t$  are embeddings of  $h, r$  and  $t$  respectively). That is, a TransE score  $\|\mathbf{v}_h + \mathbf{v}_r - \mathbf{v}_t\|_p$  of the valid triple  $(h, r, t)$  should be close to 0 and smaller than a score  $\|\mathbf{v}_{h'} + \mathbf{v}_{r'} - \mathbf{v}_{t'}\|_p$  of an invalid triple  $(h', r', t')$ . The transitional characteristic in TransE also implies the global relationships among same dimensional entries of  $\mathbf{v}_h, \mathbf{v}_r$  and  $\mathbf{v}_t$ . Other transition-based models extend TransE to use additional projection vectors or matrices to translate head and tail embeddings into the relation vector space, such as: TransH [54], TransR [27], TransD [19], STransE [33] and TransSparse [20]. Furthermore, DISTMULT [62] and ComplEx [48] 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 [16, 26, 29, 32, 47]. See other embedding models for KB completion in Nguyen [31].

Recently, convolutional neural networks (CNNs), originally designed for computer vision [24], have significantly received research attention in natural language processing [9, 22]. 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. [10] proposed ConvE—the first model applying CNN for KB completion. In ConvE, only  $\mathbf{v}_h$  and  $\mathbf{v}_r$  are reshaped and then concatenated into an input matrix which is fed to the convolution layer. Different filters of the same  $3 \times 3$  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  $\mathbf{v}_t$  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  $\mathbf{v}_h$  or  $\mathbf{v}_r$ , i.e., *ConvE does not observe the global relationships among same dimensional entries of an embedding triple  $(\mathbf{v}_h, \mathbf{v}_r, \mathbf{v}_t)$* , so that ConvE ignores the transi-

tional characteristic in transition-based models, which is one of the most useful intuitions for KB completion.

In this paper, we present ConvKB—a novel embedding model which proposes a novel use of CNN for the KB completion task. In ConvKB, each entity or relation is associated with a unique  $k$ -dimensional embedding. Let  $\mathbf{v}_h, \mathbf{v}_r$  and  $\mathbf{v}_t$  denote  $k$ -dimensional embeddings of  $h, r$  and  $t$ , respectively. For each triple  $(h, r, t)$ , the corresponding triple of  $k$ -dimensional embeddings  $(\mathbf{v}_h, \mathbf{v}_r, \mathbf{v}_t)$  is represented as a  $k \times 3$  input matrix. This input matrix is fed to the convolution layer where different filters of the same  $1 \times 3$  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 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 [10] and FB15k-237 [45], and 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.
- We also evaluate ConvKB for triple classification on two benchmark datasets WN11 and FB13 [40]. The goal is to classify whether a given triple is valid or not. ConvKB also does better than previous SOTA models, obtaining the best and second best accuracies on WN11 and FB13, respectively.
- We adapt ConvKB to search personalization where the search results for a query from a user are driven toward the personal needs of that user by exploiting historical interactions (e.g., submitted queries, clicked documents) between the user and the system [1, 18, 42, 43, 56]. Our general aim is to re-rank the documents returned by the search system in such a way that the more rel-

evant documents are ranked higher. More specially, we train our ConvKB to measure a score for each triple (*query, user, document*), and to reward higher plausibility scores for more relevant documents. We then verify this application of ConvKB on the query logs of a commercial web search engine. Experimental results show that ConvKB significantly improves the ranking quality over the strong baselines.

The paper is organized as follows. We provide related work in Section 2. We then describe our proposed model ConvKB in Section 3. We evaluate and compare ConvKB with previous models on the link prediction and triple classification tasks in Section 4. The application of ConvKB to search personalization is introduced in Section 5. The conclusion is finally presented in Section 6.

## 2. Related work

TransH [54] extends TransE to allow entities playing different roles in different relations. Each relation  $r$  is associated with a relation-specific hyperplane  $\mathbf{w}_r$ , and then the embeddings of  $h$  and  $t$  are projected to this hyperplane. The TransH score function is defined as:

$$f_{\text{TransH}}(h, r, t) = \|\mathbf{v}_{h\perp} + \mathbf{v}_r - \mathbf{v}_{t\perp}\|_p$$

where  $\mathbf{v}_{h\perp} = \mathbf{v}_h - \mathbf{w}_r^\top \mathbf{v}_h \mathbf{w}_r$  and  $\mathbf{v}_{t\perp} = \mathbf{v}_t - \mathbf{w}_r^\top \mathbf{v}_t \mathbf{w}_r$  are the projected embeddings of  $h$  and  $t$  on  $\mathbf{w}_r$  respectively.

TransR [27] extends TransH to perform projections where each relation  $r$  is associated with a projection matrix  $\mathbf{W}_r$  which is used to map entity embeddings into the vector space of relations:

$$f_{\text{TransR}}(h, r, t) = \|\mathbf{W}_r \mathbf{v}_h + \mathbf{v}_r - \mathbf{W}_r \mathbf{v}_t\|_p$$

Both TransH and TransR use only one vector or matrix to perform projections, ignoring the fact that head and tail entities have different properties to each relation. Therefore, head and tail entities should be associated with their own projection vectors or matrices as presented in direct extensions of TransH and TransR such as TransD [19], STransE [33], lpp-TransD [65], TransR-FT [12], TranSparse [20] and ITransF[59]. The transitional characteristics in these transition-based models can be intuitively defined as: if  $(h, r, t)$  is a valid fact, the *projected* embedding of  $h$  plus the embedding of  $r$  is close to the *projected* embedding of  $t$ . This reflects global relationships among same dimensional entries of *projected* entity embeddings with relation embedding. In our ConvKB model,

Table 1

The score functions in previous models and in our ConvKB model.  $\|\mathbf{v}\|_p$  denotes the  $p$ -norm of  $\mathbf{v}$ .  $\langle \mathbf{v}_h, \mathbf{v}_r, \mathbf{v}_t \rangle = \sum_i v_{h_i} v_{r_i} v_{t_i}$  denotes a tri-linear dot product.  $\bar{\mathbf{v}} = \text{Re}(\mathbf{v}) - i\text{Im}(\mathbf{v})$  with  $\text{Re}(\mathbf{v})$  and  $\text{Im}(\mathbf{v})$  corresponding to the real and imaginary parts of the complex-valued vector  $\mathbf{v}$ , and  $i$  denoting the square root of -1. In addition,  $g$  denotes a non-linear function.  $*$  denotes a convolution operator.  $\cdot$  denotes a dot product.  $\text{concat}$  denotes a concatenation operator.  $\hat{\mathbf{v}}$  denotes a 2D reshaping of  $\mathbf{v}$ .  $\Omega$  denotes a set of filters.

Model	The score function $f(h, r, t)$
TransE	$\ \mathbf{v}_h + \mathbf{v}_r - \mathbf{v}_t\ _p$
DISTMULT	$\langle \mathbf{v}_h, \mathbf{v}_r, \mathbf{v}_t \rangle$
ComplEx	$\text{Re}(\langle \mathbf{v}_h, \mathbf{v}_r, \bar{\mathbf{v}}_t \rangle)$
ConvE	$g(\text{vec}(g(\text{concat}(\hat{\mathbf{v}}_h, \hat{\mathbf{v}}_r) * \Omega)) \mathbf{W}) \cdot \mathbf{v}_t$
Our ConvKB	$\text{concat}(g([\mathbf{v}_h, \mathbf{v}_r, \mathbf{v}_t] * \Omega)) \cdot \mathbf{w}$

each filter with the shape of  $1 \times 3$  is responsible for mapping head, tail and relation embeddings to the relation vector space. So ConvKB can generalize the transitional characteristics in the transition-based models.

DISTMULT [62] and ComplEx [48] use a tri-linear dot product to compute the score for each triple. See formal definitions of DISTMULT and ComplEx in Table 1. In addition, NTN [40] uses a bilinear tensor operator into a neural network to compute the triple score. Recent approaches also show that using relation paths between entities in the KBs could help to get contextual information for improving the KB completion performance [16, 26, 32, 47]. For example, PTransE-ADD [26] and TransE-COMP [16] represent each path by summing up the embeddings of all relations in the path, while Bilinear-COMP [16] and PRUNED-PATHS [47] represent each relation in the path by a matrix and directly use matrix multiplication to modeling relation path. Some approaches also incorporate textual mentions derived from a large external corpus for improving the performance [14, 45, 46, 53]. See other methods for learning from KBs in [31, 35].

## 3. 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.

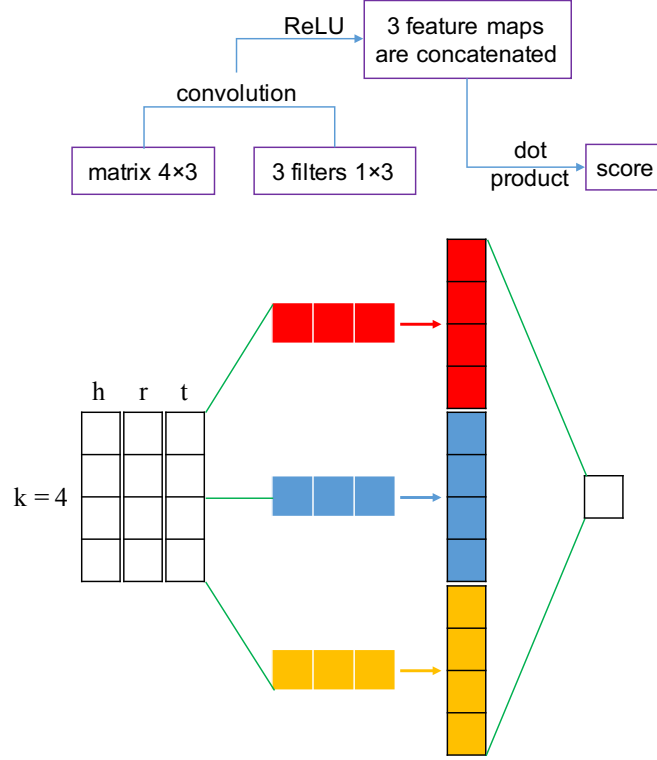


Figure 1. Process involved in the proposed ConvKB (with  $k = 4$  and  $\tau = 3$  for illustration purpose).

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

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

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

Our ConvKB uses different filters  $\in \mathbb{R}^{1 \times 3}$  to generate different feature maps. Let  $\Omega$  and  $\tau$  denote the set of filters and the number of filters, respectively, i.e.  $\tau = |\Omega|$ , resulting in  $\tau$  feature maps. These  $\tau$  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:

$$f(h, r, t) = \text{concat}(g([v_h, v_r, v_t] * \Omega)) \cdot \mathbf{w}$$

where  $\Omega$  and  $\mathbf{w}$  are shared parameters, independent of  $h, r$  and  $t$ ;  $*$  denotes a convolution operator; and  $\text{concat}$  denotes a concatenation operator.

If we only use one filter  $\omega$  (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  $\omega = [1, 1, -1]$  and  $\mathbf{w} = \mathbf{1}$  during training, ConvKB reduces to the plain TransE model [5]. So our ConvKB model can be viewed as an extension of TransE to further model global relationships.

We use the Adam optimizer [23] to train ConvKB by minimizing the loss function  $\mathcal{L}$  [48] with  $L_2$  regularization on the weight vector  $\mathbf{w}$  of the model:

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

**Input:** knowledge base  $\mathcal{G}$ , set of entities  $\mathcal{E}$ , set of relations  $\mathcal{R}$ , set of filters  $\Omega$ , weight vector  $\mathbf{w}$ , batch size  $batch\_size$ ,  $L_2$ -regularizer  $\lambda$ , boolean  $init\_filter\_normal$ , pre-trained entity and relation  $k$ -dimensional embeddings produced by TransE [5].

```

// Initialize variables.
for  $e \in \mathcal{E}$  do
   $\mathbf{v}_e \leftarrow \text{embeddings\_by\_TransE}(e)$ 
for  $r \in \mathcal{R}$  do
   $\mathbf{v}_r \leftarrow \text{embeddings\_by\_TransE}(r)$ 
// Using a truncated normal distribution with  $init\_filter\_normal = True$  while
  using  $[0.1, 0.1, -0.1]$  with  $init\_filter\_normal = False$ .
for  $\omega \in \Omega$  do
   $\text{Initialize}(\omega, init\_filter\_normal)$ 
 $\mathbf{w} \leftarrow \text{uniform} \left( -\frac{\sqrt{6}}{\sqrt{k \times |\Omega| + 1}}, \frac{\sqrt{6}}{\sqrt{k \times |\Omega| + 1}} \right)$ 
// Optimization step.
for  $epoch = 1, 2, \dots, 200$  do
  for  $i = 1, 2, \dots, \frac{|\mathcal{G}|}{batch\_size} + 1$  do
    Sample a valid_batch of  $batch\_size$  triples  $(h, r, t)$  from  $\mathcal{G}$ .
    invalid_batch = {}
    for  $(h, r, t) \in \text{valid\_batch}$  do
       $\text{invalid\_batch} \leftarrow \text{invalid\_batch} \cup \text{sample\_invalid\_triple}(h, r, t)$ 
    batch  $\leftarrow \text{valid\_batch} \cup \text{invalid\_batch}$ 
    for  $(h, r, t) \in \text{batch}$  do
      Compute  $f(h, r, t) = \text{concat}(g([\mathbf{v}_h, \mathbf{v}_r, \mathbf{v}_t] * \Omega)) \cdot \mathbf{w}$ 
      if  $(h, r, t)$  is valid then
         $l_{(h,r,t)} = 1$ 
      else
         $l_{(h,r,t)} = -1$ 
      Compute gradient  $\nabla \mathcal{L}_{batch}$  w.r.t batch:  $\nabla \sum_{(h,r,t) \in \text{batch}} \log(1 + \exp(l_{(h,r,t)} \cdot f(h, r, t))) + \frac{\lambda}{2} \|\mathbf{w}\|_2^2$ 
      Update entity and relation embeddings, weight vector  $\mathbf{w}$  and filters w.r.t  $\nabla \mathcal{L}_{batch}$ .

```

**Algorithm 1:** Parameter optimization for ConvKB in the KB completion.

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  $\mathcal{G}$ . We use the common Bernoulli trick [27, 54] to generate the head or tail entities for invalid triples. For each relation  $r$ , let  $\eta_h$  denote the averaged number of head entities per tail entity whilst  $\eta_t$  denote the averaged number of tail entities per head entity. Given a valid triple  $(h, r, t)$  of relation  $r$ , we then generate a new head entity  $h'$  with probability  $\frac{\eta_t}{\eta_t + \eta_h}$  to form an invalid triple  $(h', r, t)$  and

a new tail entity  $t'$  with probability  $\frac{\eta_h}{\eta_t + \eta_h}$  to form an invalid triple  $(h, r, t')$ . Algorithm 1 details the learning process of our ConvKB model.

#### 4. KB completion evaluation

We evaluate ConvKB on two KB completion tasks: the link prediction task [5] and the triple classification task [40]. We use benchmark datasets WN18RR [10] and FB15k-237 [45] for link prediction, and datasets FB13 and WN11 [40] for triple classification.

Table 2

Statistics of the experimental datasets. In both WN11 and FB13, each validation and test set also contains the same number of incorrect triples as the number of correct triples. It is to note that the FB13 test set is filtered to only contain 7 relations taken from 13 relations appearing in the FB13 training set.

Dataset	$ \mathcal{E} $	$ \mathcal{R} $	#Triples in train/valid/test		
FB15k-237	14,541	237	272,115	17,535	20,466
WN18RR	40,943	11	86,835	3,034	3,134
FB13	75,043	13	316,232	5,908	23,733
WN11	38,696	11	112,581	2,609	10,544

WN18RR and FB15k-237 are subsets of two common datasets WN18 and FB15k [5], respectively. As noted by Toutanova and Chen [45], 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. [10]. 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. It is also worth noting that when constructing datasets FB13 and WN11, Socher et al. [40] filtered out triples from the test set if either or both of their head and tail entities also appear in the training set in a different relation type or order. Table 2 gives statistics of the experimental datasets.

#### 4.1. Link prediction

##### 4.1.1. Task description

In the KB completion or link prediction task [5], 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. [5], 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 [5], 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. We report the fi-

nal scores on the test set for the model obtaining the highest Hits@10 score on the validation set.<sup>1</sup>

##### 4.1.2. Training protocol

We use the common Bernoulli trick [27, 54] 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.<sup>2</sup> We train TransE 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  $\omega$  and the weight vector  $\mathbf{w}$ , we use Adam [23] 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  $\lambda$  at 0.001 in our objective function. The filters  $\omega$  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.

##### 4.1.3. 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 improve-

<sup>1</sup>Some previous works also reported Hits@1. However, the formulas of MRR and Hits@1 show a strong correlation between these two scores. So using Hits@1 does not really reveal any additional information for this task.

<sup>2</sup>We employ a TransE implementation available at: <https://github.com/datquocnguyen/STransE>

Table 3

Link prediction 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. [10] 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 underline.

Method	WN18RR			FB15k-237		
	MR	MRR	H@10	MR	MRR	H@10
IRN [39]	–	–	–	<u>211</u>	–	46.4
KBGAN [7]	–	0.213	48.1	–	0.278	45.8
DISTMULT [62] [★]	5110	0.43	49	254	0.241	41.9
ComplEx [48] [★]	5261	<u>0.44</u>	<u>51</u>	339	0.247	42.8
ConvE [10]	5277	<b>0.46</b>	48	246	<u>0.316</u>	49.1
TransE [5] (our results)	<u>3384</u>	0.226	50.1	347	0.294	46.5
Our ConvKB model	<b>2554</b>	0.248	<b>52.5</b>	257	<b>0.396</b>	<b>51.7</b>
$KB_{LRN}$ [14]	–	–	–	<b>209</b>	0.309	<u>49.3</u>
R-GCN+ [37]	–	–	–	–	0.249	41.7
Neural LP [63]	–	–	–	–	0.240	36.2
Node+LinkFeat [45]	–	–	–	–	0.293	46.2

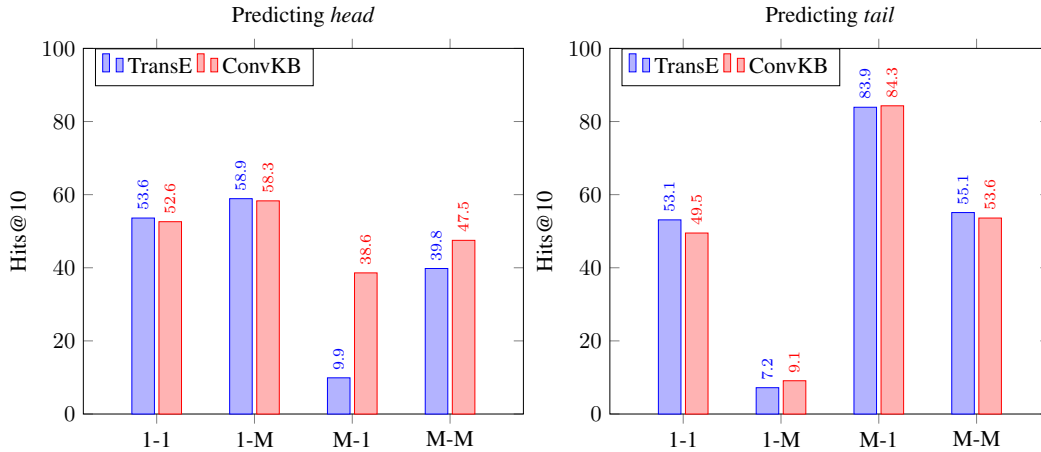


Figure 2. Hits@10 (in %) on the FB15k-237 test set w.r.t each relation category.

ments 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 [26, 32, 36, 48]. However, when comparing the CNN-based embedding model ConvE with other embedding models, Dettmers et al. [10] 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,  $KB_{LRN}$

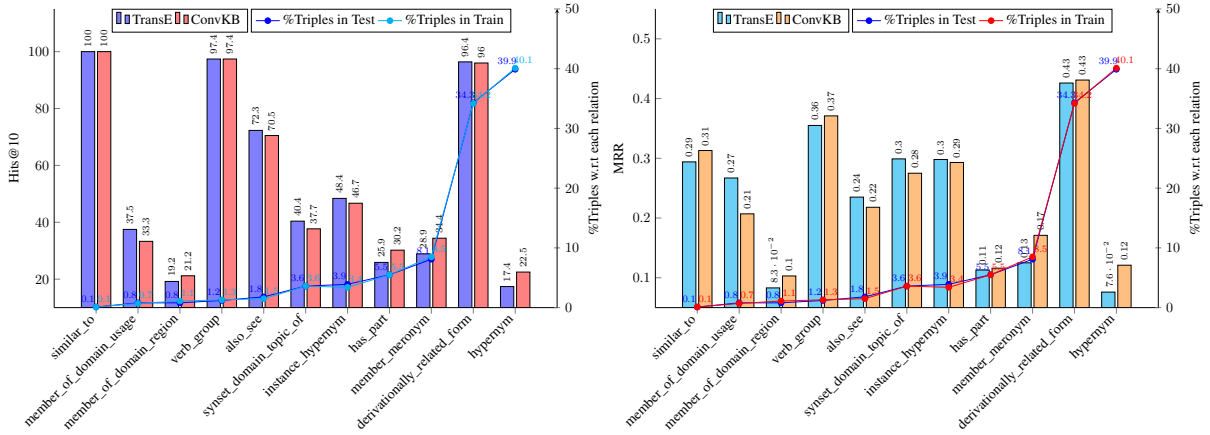


Figure 3. Hits@10 and MRR on the WN18RR test set w.r.t each relation. The right y-axis is the percentage of triples corresponding to relations.

gives better MR than ConvKB. We plan to extend ConvKB with relation path information to obtain better link prediction performance in future work.

Following Bordes et al. [5], we explore the Hits@10 results on the FB15k-237 test set corresponding to the relation categories. For each relation  $r$ , we calculate the averaged number  $\eta_h$  of heads per tail and the averaged number  $\eta_t$  of tails per head. If  $\eta_h < 1.5$  and  $\eta_t < 1.5$ ,  $r$  is classified as one-to-one (**1-1**). If  $\eta_h < 1.5$  and  $\eta_t \geq 1.5$ ,  $r$  is classified as one-to-many (**1-M**). If  $\eta_h \geq 1.5$  and  $\eta_t < 1.5$ ,  $r$  is classified as many-to-one (**M-1**). If  $\eta_h \geq 1.5$  and  $\eta_t \geq 1.5$ ,  $r$  is classified as many-to-many (**M-M**). We find that 17, 26, 81 and 113 relations are classified as 1-1, 1-M, M-1 and M-M, respectively. And 0.9%, 6.3%, 20.5% and 72.3% of the FB15k-237 test triples have their relations classified as 1-1, 1-M, M-1 and M-M, respectively.

Figure 2 shows the Hits@10 results for separately predicting head and tail entities on the FB15k-237 test set with respect to (w.r.t.) each relation category. We find that ConvKB is outperformed by TransE in 1-1 as 1-1 relations are relatively rare. We also find that both TransE and ConvKB are easier to predict entities on the relational “side 1” triples (i.e., predicting head entities in 1-1 and 1-M, and predicting tail entities in 1-1 and M-1). However, TransE is not good at predicting head entities in M-1 and M-M where TransE obtains the Hits@10 scores of 9.9% and 39.8%, while ConvKB is better in achieving the Hits@10 scores of 38.6% and 47.5%, respectively. A reason is probably that ConvKB could bring a generalization of projecting the embedding triples into the vector space of relations rather than TransE. Hence, this helps ConvKB to better modeling M-1 and M-M relations.

For a more concrete example, Figure 3 presents Hits@10 and MRR scores on WN18RR w.r.t each relation type. *member\_meronym* and *hypernym* are 1-M and M-1 relations, respectively. We find that TransE encounters a difficulty when dealing with these relation types. E.g., for 1,251 triples containing the relation *hypernym* from 3,134 test triples in the WN18RR test set, TransE only obtains the Hits@10 and MRR scores of 17.4% and 0.076 respectively, while ConvKB performs better than TransE and gets the Hits@10 and MRR scores of 22.5% and 0.121 respectively. In summary, figures 2 and 3 show that ConvKB are better at modeling 1-M, M-1 and M-M relations than TransE.

## 4.2. Triple classification

### 4.2.1. Task description

The triple classification task aims to predict whether a given triple  $(h, r, t)$  is valid or not [40]. Each relation  $r$  is associated with a threshold  $\theta_r$ . For an unseen test triple  $(h, r, t)$ , if its score is below  $\theta_r$  then it will be classified as valid, otherwise invalid. Following Socher et al. [40], the relation-specific threshold  $\theta_r$  is obtained by maximizing the micro-averaged classification accuracy on the validation set.

### 4.2.2. Training protocol

Similar to the training protocol in Section 4.1.2, we sample invalid triples using the common Bernoulli trick and also train TransE to produce entity and relation embeddings for initializing embeddings in ConvKB. The best accuracies obtained by TransE on the validation set are when using  $l_1$ -norm, learning rate at 0.01,  $\gamma = 7$  and  $k = 50$  for WN11, and using  $l_2$ -norm, learning rate at 0.01,  $\gamma = 1$  and  $k = 100$  for FB13. We



Table 4

Accuracy results (in %) on the WN11 and FB13 test sets. The last 4 rows report accuracies of the models that use relation paths or incorporate with a large external corpus. The best score is in **bold** while the second best score is in underline. “Avg.” denotes the averaged accuracy over two datasets.

Method	WN11	FB13	Avg.
NTN [40]	70.6	87.2	78.9
TransH [54]	78.8	83.3	81.1
TransR [27]	85.9	82.5	84.2
TransD [19]	86.4	<b>89.1</b>	<u>87.8</u>
TransR-FT [12]	86.6	82.9	84.8
TransSparse-S [20]	86.4	88.2	87.3
TransSparse-US [20]	86.8	87.5	87.2
ManifoldE [57]	<u>87.5</u>	87.2	87.4
TransG [58]	87.4	87.3	87.4
lppTransD [65]	86.2	88.6	87.4
TransE [5] (our results)	86.5	87.5	87.0
Our ConvKB model	<b>87.6</b>	<u>88.8</u>	<b>88.2</b>
TransE-NMM [32]	86.8	88.6	87.7
TEKE_H [53]	84.8	84.2	84.5
Bilinear-COMP [16]	77.6	86.1	81.9
TransE-COMP [16]	80.3	87.6	84.0

then use a grid search to choose the hyper-parameter for ConvKB. We monitor the accuracy after each training epoch, and obtain the best accuracies on validation set when using  $k = 50$ ,  $\tau = 200$ , the truncated normal distribution for filter initialization, and the initial learning rate at  $5e^{-4}$  on WN11; and  $k = 100$ ,  $\tau = 200$ , also the truncated normal distribution for filter initialization, and the initial learning rate at  $5e^{-5}$  on FB13.

#### 4.2.3. Main results

Table 4 presents the accuracy results of our ConvKB model and previous published results on the WN11 and FB13 datasets. On WN11, ConvKB obtains an accuracy of 87.6% which outperforms all other models. On FB13, ConvKB gains a second highest accuracy of 88.8% which is 0.3% outperformed by TransD. Compared to TransE, ConvKB absolutely improves by 1.1% on WN11 and 1.3% on FB13. Overall, ConvKB yields the best performance averaged over these two benchmark datasets. This also indicates the generalization of ConvKB over different datasets.

Regarding to TransE, Table 4 demonstrates that we obtain very competitive accuracies of 86.5% and 87.5% on WN11 and FB13 respectively. On WN11,

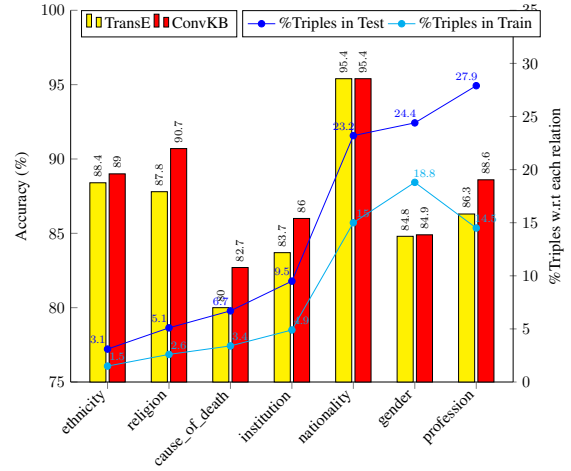


Figure 4. Accuracy results on the FB13 test set w.r.t each relation. The right y-axis is the number of triples corresponding to relations.

TransE is comparable with TransD, TransR-FT and TransSparse-S while it scores better than lppTransD and TransE-COMP. On FB13, TransE performs slightly better than ManifoldE and TransG while it achieves similar scores in comparison with TransE-COMP and TransSparse-US. Note that TransR, TransSparse-S/US and TransD also perform the embedding initialization using TransE outputs (but these models do not report their TransE accuracy results). Hence, these models might get better results when using our TransE results as shown in this paper.

Figure 4 visualizes the accuracy results of different relations on FB13 for TransE and ConvKB. Relations *institution* and *profession* can be categorized as M-M where ConvKB is about 2.3% absolute higher accuracy than TransE, while remaining relations can be categorized as M-1. In short, Figure 4 shows that ConvKB performs equal to or better than TransE for all 7 relations in the FB13 test set.

## 5. Application for search personalization

Search personalization, an important feature of commercial search engines, has been recently attracted much attention from both academia [6, 8, 11, 28, 30, 49, 51, 64] and industry (e.g., Bing, Google, Airbnb [15]). Unlike classical searching methods, personalized search systems utilize the historical interactions such as submitted queries and clicked documents between a user and the systems to tailor returned search results to the needs of that user [1, 18, 42, 43, 56].

Table 5

Basic statistics of the dataset [51].

# users	106
#distinct queries	6,632
#SAT clicks	8,052
#distinct documents	33,591

That historical information can be used to build the *user* profile, which is crucial to effective personalization [42, 43, 50, 52].

Given a *user*, a submitted *query* and the *documents* returned by a search system for that query, our approach is to re-rank the returned documents so that the more relevant documents should be ranked higher. Following Vu et al. [51], we represent the relationship between the submitted query, the user and the returned document as a (h, r, t)-like triple (*query*, *user*, *document*). The triple captures how much interest a user puts on a document given a query. Therefore, we can also evaluate the effectiveness of our ConvKB model for the search personalization task.

We evaluate ConvKB using the search results returned by a commercial search engine. We use the same dataset of query logs of 106 anonymous users from Vu et al. [51]. A log entity consists of a user identifier, a query, top-10 returned documents ranked by the search engine and clicked documents along with the user’s dwell time. Vu et al. [51] employed the SAT criteria [13] to identify whether or not a clicked document is relevant from the query logs (i.e., a SAT click). They then assigned a *relevant* label to a returned document if it is a SAT click and also assigned *irrelevant* labels to the remaining top-10 documents. The rank position of the *relevant* labeled documents is used as the ground truth to evaluate the search performance before and after re-ranking. As a result, the dataset contains 8,052 valid triples (*query*, *user*, *relevant document*) in which 5,658, 1,184 and 1,210 valid triples are used for training, validation and test, respectively. Table 5 presents the dataset statistics.

### 5.1. Evaluation protocol

Our ConvKB model is used to re-rank the original list of top-10 documents returned by the commercial search engine as follows: (1) We train ConvKB and use the trained model to calculate a score for each triple (*question*, *user*, *document*). (2) We then sort the scores in the ascending order to achieve a new ranked list. To

evaluate the performance, we use two common metrics in document ranking: MRR and Hits@1.<sup>3</sup>

### 5.2. Training protocol

#### 5.2.1. Query and document embedding initialization

We initialize query and document embeddings for ConvKB and the baseline TransE, then fix query and document embeddings (i.e. not updating these embeddings) during training.

To initialize document embeddings, we follow Vu et al. [51] to train a LDA topic model [2] with 200 topics only on the *relevant* documents (i.e., SAT clicks) extracted from the query logs. We then use the trained LDA model to infer the probability distribution over topics for each document. We use the topic proportion vector of each document as its document embedding (i.e.  $k = 200$ ). In particular, the  $z^{\text{th}}$  element ( $z = 1, 2, \dots, k$ ) of the vector embedding for document  $d$  is:  $v_{d,z} = P(z | d)$  where  $P(z | d)$  is the probability of the topic  $z$  given the document  $d$ .

We also represent each query by a probability distribution vector over topics. Let  $\mathcal{D}_q = \{d_1, d_2, \dots, d_n\}$  be the set of top  $n$  ranked documents returned for a query  $q$  (here,  $n = 10$ ). The  $z^{\text{th}}$  element of the vector embedding for query  $q$  is defined as in [51]:  $v_{q,z} = \sum_{i=1}^n \lambda_i P(z | d_i)$ , where  $\lambda_i = \frac{\delta^{i-1}}{\sum_{j=1}^n \delta^{j-1}}$  is the exponential decay function of  $i$  which is the rank of  $d_i$  in  $\mathcal{D}_q$ . And  $\delta$  is the decay hyper-parameter ( $0 < \delta < 1$ ).

#### 5.2.2. Hyper-parameter tuning

Similar to the training protocol presented in Section 4.1.2, we run model up to 200 epochs and perform a grid search to choose optimal hyper-parameters on the validation set. Following Vu et al. [51], we use  $\delta = 0.8$ . We also monitor the MRR score after each training epoch and obtain the highest MRR score on the validation set when using the *margin* at 5,  $l_1$ -norm and learning rate at  $5e^{-3}$  for TransE; and using  $\tau = 500$ , the truncated normal distribution for filter initialization, and the initial learning rate at  $5e^{-4}$  for ConvKB.

### 5.3. Results

Table 6 presents the experimental results of ConvKB, TransE and the previous published results of other strong baselines, in which ConvKB obtains highest MRR and Hits@1 scores. In particular, ConvKB

<sup>3</sup>We re-rank the list of top-10 documents returned by the search engine, so all models obtain the same Hits@10 scores.

Table 6

Experimental results on the test set. ★ denotes the results reported in Vu et al. [51]. **SE**: The original rank is returned by the search engine. **CI**: This baseline use a personalized navigation method based on previously clicking returned documents [44]. **SP**: A search personalization method makes use of the short-term profiles [1, 50]. The subscripts denote the relative improvement over the baseline **TransE**.

Model	MRR	Hits@1 (%)
<b>SE</b> [★]	0.559	38.5
<b>CI</b> [44] [★]	0.597	41.6
<b>SP</b> [1, 50] [★]	0.631	45.2
<b>TransE</b> [5] [★]	0.645	48.1
<b>STransE</b> [33] [★]	0.656	50.1
<b>TransE</b> (our results)	0.669	50.9
Our <b>ConvKB</b> model	0.750 <sub>+12.1%</sub>	59.9 <sub>+17.7%</sub>

does significantly better than TransE with relative improvements at 12.1% for MRR and 17.7% for Hits@1. It is probably because our model not only can capture richer relational characteristics within the triple but also generalize the transitional relationships between embeddings of user queries and relevant documents for user profiles. We also obtain higher TransE results than those reported in Vu et al. [51]. The reason is that for each valid triple, rather than using only one invalid triple as in [51], we take into account its all invalid triples to train TransE (each valid or invalid triple contains a relevant- or irrelevant-labelled document, respectively).

## 6. 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 ConvKB model outperforms other state-of-the-art models on two benchmark datasets WN18RR and FB15k-237 for the link prediction task, and on two other benchmark datasets WN11 and FB13 for the triple classification task. ConvKB obtains the best mean rank and the highest Hits@10 on WN18RR and obtains the highest mean reciprocal rank and Hits@10 on FB15k-237. In addition, ConvKB produces the best accuracy on WN11 and the second best accuracy on FB13. Moreover, we show the

effectiveness of ConvKB for search personalization, in which ConvKB outperforms the strong baselines on the query logs of a commercial web search engine.

In the future work, we plan to extend ConvKB with relation path information to achieve better performance. We will also adapt ConvKB to other personalization tasks where we can model each task as a triple relationship, e.g. in personalized query suggestion or auto-completion.

Our ConvKB implementation is available at: <https://github.com/daiquocnguyen/ConvKB>.

## Bibliographic note

This paper extends our paper [34] published in Proceedings of NAACL-HLT 2018 (Volume 2: Short Papers). We first add a significantly improved analysis to the link prediction task. Then we conduct new extensive empirical study on the triple classification and search personalization tasks.

## Acknowledgement

This research was partially supported by the Australian Research Council (ARC) DP150100031 and DP160103934.

## References

- [1] P. N. Bennett, R. W. White, W. Chu, S. T. Dumais, P. Bailey, F. Borisyuk, and X. Cui. Modeling the impact of short- and long-term behavior on search personalization. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '12*, pages 185–194, Portland, Oregon, USA, 2012. ACM. URL <http://dx.doi.org/10.1145/2348283.2348312>.
- [2] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. In T. G. Dietterich, S. Becker, and Z. Ghahramani, editors, *Advances in Neural Information Processing Systems 14*, pages 601–608. MIT Press, 2002. URL <http://papers.nips.cc/paper/2070-latent-dirichlet-allocation.pdf>.
- [3] K. Bollacker, C. Evans, P. Paritosh, T. Sturge, and J. Taylor. Freebase: A collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data, SIGMOD '08*, pages 1247–1250, Vancouver, Canada, 2008. ACM. ISBN 978-1-60558-102-6. URL <https://doi.org/10.1145/1376616.1376746>.
- [4] A. Bordes, J. Weston, R. Collobert, and Y. Bengio. Learning Structured Embeddings of Knowledge Bases. In *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence, AAAI'11*, pages 301–306, San Francisco, California, 2011. AAAI Press. URL <https://www.aaai.org/ocs/index.php/AAAI/AAAI11/paper/view/3659>.

- [5] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko. Translating embeddings for modeling multi-relational data. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 2787–2795. Curran Associates, Inc., 2013.
- [6] F. Cai, S. Wang, and M. de Rijke. Behavior-based personalization in web search. *Journal of the Association for Information Science and Technology*, 68(4):855–868, 2017. URL <https://doi.org/10.1002/asi.23735>.
- [7] L. Cai and W. Y. Wang. KBGAN: Adversarial learning for knowledge graph embeddings. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1470–1480, New Orleans, Louisiana, 2018. Association for Computational Linguistics. URL <http://www.aclweb.org/anthology/N18-1133>.
- [8] Z. Cheng, S. Jialie, and S. C. Hoi. On effective personalized music retrieval by exploring online user behaviors. In *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '16*, pages 125–134, Pisa, Italy, 2016. ACM. URL <http://doi.org/10.1145/2911451.2911491>.
- [9] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa. Natural language processing (almost) from scratch. *Journal of Machine Learning Research*, 12:2493–2537, 2011. URL [www.jmlr.org/papers/volume12/collobert11a/collobert11a.pdf](http://www.jmlr.org/papers/volume12/collobert11a/collobert11a.pdf).
- [10] T. Dettmers, P. Minervini, P. Stenetorp, and S. Riedel. Convolutional 2D knowledge graph embeddings. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence, AAAI'18*, pages 1811–1818. AAAI Press, 2018. URL <https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/17366>.
- [11] Z. Dou, R. Song, and J.-R. Wen. A large-scale evaluation and analysis of personalized search strategies. In *Proceedings of the 16th International Conference on World Wide Web, WWW '07*, pages 581–590, Banff, Alberta, Canada, 2007. International World Wide Web Conferences Steering Committee. URL <http://doi.org/10.1145/1242572.1242651>.
- [12] J. Feng, M. Huang, M. Wang, M. Zhou, Y. Hao, and X. Zhu. Knowledge graph embedding by flexible translation. In *Proceedings of the Fifteenth International Conference on Principles of Knowledge Representation and Reasoning, KR'16*, pages 557–560, Cape Town, South Africa, 2016. AAAI Press. URL <https://www.aaai.org/ocs/index.php/KR/KR16/paper/view/12887>.
- [13] S. Fox, K. Karnawat, M. Mydland, S. Dumais, and T. White. Evaluating Implicit Measures to Improve Web Search. *ACM Transactions on Information Systems*, 23(2):147–168, 2005. URL <http://doi.org/10.1145/1059981.1059982>.
- [14] A. García-Durán and M. Niepert. KBLRN: End-to-End Learning of Knowledge Base Representations with Latent, Relational, and Numerical Features. *arXiv preprint abs/1709.04676*, 2017. URL <https://arxiv.org/abs/1709.04676>.
- [15] M. Grbovic. Search ranking and personalization at Airbnb. In *Proceedings of the Eleventh ACM Conference on Recommender Systems, RecSys '17*, pages 339–340, Como, Italy, 2017. ACM. URL <http://doi.org/10.1145/3109859.3109920>.
- [16] K. Guu, J. Miller, and P. Liang. Traversing knowledge graphs in vector space. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 318–327, Lisbon, Portugal, 2015. Association for Computational Linguistics. URL <http://aclweb.org/anthology/D15-1038>.
- [17] Y. Hao, Y. Zhang, K. Liu, S. He, Z. Liu, H. Wu, and J. Zhao. 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, Vancouver, Canada, 2017. Association for Computational Linguistics. URL <http://aclweb.org/anthology/P17-1021>.
- [18] M. Harvey, F. Crestani, and M. J. Carman. Building user profiles from topic models for personalised search. In *Proceedings of the 22Nd ACM International Conference on Information & Knowledge Management, CIKM '13*, pages 2309–2314, San Francisco, California, USA, 2013. ACM. URL <http://doi.org/10.1145/2505515.2505642>.
- [19] G. Ji, S. He, L. Xu, K. Liu, and J. Zhao. 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, Beijing, China, 2015. Association for Computational Linguistics. URL <http://www.aclweb.org/anthology/P15-1067>.
- [20] G. Ji, K. Liu, S. He, and J. Zhao. Knowledge graph completion with adaptive sparse transfer matrix. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, AAAI'16*, pages 985–991, Phoenix, Arizona, 2016. AAAI Press. URL <https://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/view/11982>.
- [21] G. Kasneci, F. M. Suchanek, G. Ifrim, M. Ramanath, and G. Weikum. Naga: Searching and ranking knowledge. In *Proceedings of the 2008 IEEE 24th International Conference on Data Engineering, ICDE '08*, pages 953–962, Washington, DC, USA, 2008. IEEE Computer Society. URL <https://doi.org/10.1109/ICDE.2008.4497504>.
- [22] Y. Kim. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1746–1751, Doha, Qatar, 2014. Association for Computational Linguistics. URL <http://www.aclweb.org/anthology/D14-1181>.
- [23] D. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. URL <https://arxiv.org/abs/1412.6980>.
- [24] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. URL <http://doi.org/10.1109/5.726791>.
- [25] J. Lehmann, R. Isele, M. Jakob, A. Jentzsch, D. Kontokostas, P. N. Mendes, S. Hellmann, M. Morse, P. Van Kleef, S. Auer, et al. DBpedia – a large-scale, multilingual knowledge base extracted from Wikipedia. *Semantic Web*, 6:167–195, 2015. URL <https://doi.org/10.3233/SW-140134>.
- [26] Y. Lin, Z. Liu, H. Luan, M. Sun, S. Rao, and S. Liu. 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, Lisbon, Portugal, 2015. Association for Computational Linguistics. URL <http://aclweb.org/anthology/D15-1082>.

- [27] Y. Lin, Z. Liu, M. Sun, Y. Liu, and X. Zhu. Learning entity and relation embeddings for knowledge graph completion. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, AAAI'15, pages 2181–2187, Austin, Texas, 2015. AAAI Press. URL <https://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/view/9571>.
- [28] X. Liu. Modeling users' dynamic preference for personalized recommendation. In *Proceedings of the 24th International Conference on Artificial Intelligence*, IJCAI'15, pages 1785–1791, Buenos Aires, Argentina, 2015. AAAI Press. URL <https://www.ijcai.org/Proceedings/15/Papers/254.pdf>.
- [29] Y. Luo, Q. Wang, B. Wang, and L. Guo. Context-dependent knowledge graph embedding. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1656–1661, Lisbon, Portugal, 2015. Association for Computational Linguistics. URL <http://aclweb.org/anthology/D15-1191>.
- [30] N. Nanas, M. Vavalis, and A. De Roeck. A network-based model for high-dimensional information filtering. In *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '10, pages 202–209, Geneva, Switzerland, 2010. ACM. URL <http://doi.org/10.1145/1835449.1835485>.
- [31] D. Q. Nguyen. An overview of embedding models of entities and relationships for knowledge base completion. *arXiv preprint*, arXiv:1703.08098, 2017. URL <https://arxiv.org/abs/1703.08098>.
- [32] D. Q. Nguyen, K. Sirts, L. Qu, and M. Johnson. Neighborhood mixture model for knowledge base completion. In *Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning*, pages 40–50, Berlin, Germany, 2016. Association for Computational Linguistics. URL <http://www.aclweb.org/anthology/K16-1005>.
- [33] D. Q. Nguyen, K. Sirts, L. Qu, and M. Johnson. 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, San Diego, California, 2016. Association for Computational Linguistics. URL <http://www.aclweb.org/anthology/N16-1054>.
- [34] D. Q. Nguyen, T. D. Nguyen, D. Q. Nguyen, and D. Phung. A novel embedding model for knowledge base completion based on convolutional neural network. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 327–333, New Orleans, Louisiana, 2018. Association for Computational Linguistics. URL <http://www.aclweb.org/anthology/N18-2053>.
- [35] M. Nickel, K. Murphy, V. Tresp, and E. Gabrilovich. A Review of Relational Machine Learning for Knowledge Graphs. *Proceedings of the IEEE*, 104(1):11–33, 2016. URL <https://doi.org/10.1109/JPROC.2015.2483592>.
- [36] M. Nickel, L. Rosasco, and T. Poggio. Holographic embeddings of knowledge graphs. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, AAAI'16, pages 1955–1961, Phoenix, Arizona, 2016. AAAI Press. URL <https://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/view/12484>.
- [37] M. Schlichtkrull, T. Kipf, P. Bloem, R. v. d. Berg, I. Titov, and M. Welling. Modeling relational data with graph convolutional networks. *arXiv preprint arXiv:1703.06103*, 2017. URL <https://arxiv.org/abs/1703.06103>.
- [38] M. Schuhmacher and S. P. Ponzetto. Knowledge-based graph document modeling. In *Proceedings of the 7th ACM International Conference on Web Search and Data Mining*, WSDM '14, pages 543–552, New York, New York, USA, 2014. ACM. URL <http://doi.org/10.1145/2556195.2556250>.
- [39] Y. Shen, P. Huang, M. Chang, and J. Gao. Traversing knowledge graph in vector space without symbolic space guidance. *arXiv preprint arXiv:1611.04642v4*, 2017. URL <https://arxiv.org/abs/1611.04642v4>.
- [40] R. Socher, D. Chen, C. D. Manning, and A. Ng. Reasoning with neural tensor networks for knowledge base completion. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 926–934. Curran Associates, Inc., 2013.
- [41] F. M. Suchanek, G. Kasneci, and G. Weikum. Yago: A core of semantic knowledge. In *Proceedings of the 16th International Conference on World Wide Web*, WWW '07, pages 697–706, Banff, Alberta, Canada, 2007. International World Wide Web Conferences Steering Committee. URL <http://doi.org/10.1145/1242572.1242667>.
- [42] J. Teevan, S. T. Dumais, and E. Horvitz. Personalizing search via automated analysis of interests and activities. In *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '05, pages 449–456, Salvador, Brazil, 2005. ACM. URL <http://doi.org/10.1145/1076034.1076111>.
- [43] J. Teevan, M. R. Morris, and S. Bush. Discovering and using groups to improve personalized search. In *Proceedings of the Second ACM International Conference on Web Search and Data Mining*, WSDM '09, pages 15–24, Barcelona, Spain, 2009. ACM. URL <http://doi.org/10.1145/1498759.1498786>.
- [44] J. Teevan, D. J. Liebling, and G. Ravichandran Geetha. Understanding and predicting personal navigation. In *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, WSDM '11, pages 85–94, Hong Kong, China, 2011. ACM. URL <http://doi.org/10.1145/1935826.1935848>.
- [45] K. Toutanova and D. Chen. 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, Beijing, China, 2015. Association for Computational Linguistics. URL <http://www.aclweb.org/anthology/W15-4007>.
- [46] K. Toutanova, D. Chen, P. Pantel, H. Poon, P. Choudhury, and M. Gamon. Representing text for joint embedding of text and knowledge bases. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1499–1509, Lisbon, Portugal, 2015. Association for Computational Linguistics. URL <http://aclweb.org/anthology/D15-1174>.
- [47] K. Toutanova, V. Lin, W.-t. Yih, H. Poon, and C. Quirk. Compositional learning of embeddings for relation paths in knowledge base and text. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1434–1444, Berlin, Germany, 2016. Association for Computational Linguistics. URL <http://www.aclweb.org/anthology/P16-1136>.

- [48] T. Trouillon, J. Welbl, S. Riedel, E. Gaussier, and G. Bouchard. Complex embeddings for simple link prediction. In *Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48*, ICML'16, pages 2071–2080, New York, NY, USA, 2016. JMLR.org. URL <http://proceedings.mlr.press/v48/trouillon16.pdf>.
- [49] T. Vu, D. Song, A. Willis, S. N. Tran, and J. Li. Improving search personalisation with dynamic group formation. In *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval*, SIGIR '14, pages 951–954, Gold Coast, Queensland, Australia, 2014. ACM. URL <http://doi.org/10.1145/2600428.2609482>.
- [50] T. Vu, A. Willis, S. N. Tran, and D. Song. Temporal latent topic user profiles for search personalisation. In *Proceedings of the European Conference on Information Retrieval*, pages 605–616, Vienna, Austria, 2015. Springer International Publishing. URL [https://doi.org/10.1007/978-3-319-16354-3\\_67](https://doi.org/10.1007/978-3-319-16354-3_67).
- [51] T. Vu, D. Q. Nguyen, M. Johnson, D. Song, and A. Willis. Search Personalization with Embeddings. In *Proceedings of the European Conference on Information Retrieval*, pages 598–604, Aberdeen, Scotland, 2017. Springer International Publishing. URL [https://doi.org/10.1007/978-3-319-56608-5\\_54](https://doi.org/10.1007/978-3-319-56608-5_54).
- [52] T. Vu, A. Willis, U. Kruschwitz, and D. Song. Personalised query suggestion for intranet search with temporal user profiling. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*, CHIIR '17, pages 265–268, Oslo, Norway, 2017. ACM. URL <http://doi.org/10.1145/3020165.3022129>.
- [53] Z. Wang and J. Li. Text-enhanced representation learning for knowledge graph. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, IJCAI'16, pages 1293–1299, New York, New York, USA, 2016. AAAI Press. URL <https://www.ijcai.org/Proceedings/16/Papers/187.pdf>.
- [54] Z. Wang, J. Zhang, J. Feng, and Z. Chen. Knowledge graph embedding by translating on hyperplanes. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, AAAI'14, pages 1112–1119, Quebec City, Quebec, Canada, 2014. AAAI Press. URL <https://www.aaai.org/ocs/index.php/AAAI/AAAI14/paper/view/8531>.
- [55] R. West, E. Gabrilovich, K. Murphy, S. Sun, R. Gupta, and D. Lin. Knowledge base completion via search-based question answering. In *Proceedings of the 23rd International Conference on World Wide Web*, WWW '14, pages 515–526, Seoul, Korea, 2014. International World Wide Web Conferences Steering Committee. URL <http://doi.org/10.1145/2566486.2568032>.
- [56] R. W. White, W. Chu, A. Hassan, X. He, Y. Song, and H. Wang. Enhancing personalized search by mining and modeling task behavior. In *Proceedings of the 22nd International Conference on World Wide Web*, WWW '13, pages 1411–1420, Rio de Janeiro, Brazil, 2013. International World Wide Web Conferences Steering Committee. URL <http://doi.org/10.1145/2488388.2488511>.
- [57] H. Xiao, M. Huang, and X. Zhu. From one point to a manifold: Knowledge graph embedding for precise link prediction. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, IJCAI'16, pages 1315–1321, New York, New York, USA, 2016. AAAI Press. URL <https://www.ijcai.org/Proceedings/16/Papers/190.pdf>.
- [58] H. Xiao, M. Huang, and X. Zhu. Transg : A generative model for knowledge graph embedding. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2316–2325, Berlin, Germany, 2016. Association for Computational Linguistics. URL <http://www.aclweb.org/anthology/P16-1219>.
- [59] Q. Xie, X. Ma, Z. Dai, and E. Hovy. An interpretable knowledge transfer model for knowledge base completion. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 950–962, Vancouver, Canada, 2017. Association for Computational Linguistics. URL <http://aclweb.org/anthology/P17-1088>.
- [60] C. Xiong, R. Power, and J. Callan. Explicit semantic ranking for academic search via knowledge graph embedding. In *Proceedings of the 26th International Conference on World Wide Web*, WWW '17, pages 1271–1279, Perth, Australia, 2017. International World Wide Web Conferences Steering Committee. URL <https://doi.org/10.1145/3038912.3052558>.
- [61] B. Yang and T. Mitchell. 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, Vancouver, Canada, 2017. Association for Computational Linguistics. URL <http://aclweb.org/anthology/P17-1132>.
- [62] B. Yang, W.-t. Yih, X. He, J. Gao, and L. Deng. Embedding Entities and Relations for Learning and Inference in Knowledge Bases. In *Proceedings of the International Conference on Learning Representations*, 2015. URL <https://arxiv.org/abs/1412.6575>.
- [63] F. Yang, Z. Yang, and W. W. Cohen. Differentiable learning of logical rules for knowledge base reasoning. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 2319–2328. Curran Associates, Inc., 2017.
- [64] L. Yang, Q. Guo, Y. Song, S. Meng, M. Shokouhi, K. McDonald, and W. B. Croft. Modeling User Interests for Zero-Query Ranking. In *Proceedings of the European Conference on Information Retrieval*, pages 171–184. Springer International Publishing, 2016. URL [https://doi.org/10.1007/978-3-319-30671-1\\_13](https://doi.org/10.1007/978-3-319-30671-1_13).
- [65] H.-G. Yoon, H.-J. Song, S.-B. Park, and S.-Y. Park. A translation-based knowledge graph embedding preserving logical property of relations. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 907–916, San Diego, California, 2016. Association for Computational Linguistics. URL <http://www.aclweb.org/anthology/N16-1105>.
- [66] Y. Zhang, K. Liu, S. He, G. Ji, Z. Liu, H. Wu, and J. Zhao. Question Answering over Knowledge Base with Neural Attention Combining Global Knowledge Information. *arXiv preprint arXiv:1606.00979*, 2016. URL <https://arxiv.org/abs/1606.00979>.