IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received September 19, 2019, accepted October 24, 2019, date of publication October 29, 2019, date of current version November 7, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2950015

Knowledge Graph Embedding With Interactive Guidance From Entity Descriptions

WEN'AN ZHOU, SHIRUI WANG, AND CHAO JIANG
School of Computer Science, Beijing University of Posts and Telecommunications, Beijing 100876, China

Corresponding author: Shirui Wang (wangshirui@bupt.edu.cn)

This work was supported by the Fundamental Research Funds for the Central Universities under Grant 2019XD-A20.

ABSTRACT Knowledge Graph (KG) embedding aims to represent both entities and relations into a continuous low-dimensional vector space. Most previous attempts perform the embedding task using only knowledge triples to indicate relations between entities. Entity descriptions, although containing rich background information, have not been well utilized in these methods. In this paper, we propose Entity Descriptions-Guided Embedding (EDGE), a novel method for learning the knowledge graph representations with semantic guidance from entity descriptions. EDGE enables an embedding model to learn simultaneously from 1) knowledge triples that have been directly observed in a given KG, and 2) entity descriptions which have rich semantic information about these entities. In the learning process, EDGE encodes the semantics of entity descriptions to enhance the learning of knowledge graph embedding, and integrates such learned KG embedding to constraint their corresponding word embeddings in entity descriptions. Through this interactive procedure, semantics of entity descriptions may be better transferred into the learned KG embedding. We evaluate EDGE in link prediction and entity classification on Freebase and WordNet. Experimental results show that: 1) with entity descriptions injected, EDGE achieves significant and consistent improvements over state-of-the-art baselines; and 2) compared to those one-time injection schemes studied before, the interactive guidance strategy maximizes the utility of entity descriptions for KG embedding, and indeed achieves substantially better performance.

INDEX TERMS Knowledge graph embedding, entity descriptions, interactive guidance.

I. INTRODUCTION

Knowledge graphs (KGs) such as Freebase [1], DBpedia [2] and YAGO [3] provide a structured representation of world knowledge and are extremely useful and crucial resources for several artificial intelligent related applications including question answering [4]–[7] and recommendation systems [8]–[11]. A typical KG is represented as a multi-relational graph with entities as nodes and relations as different types of edges, and expresses knowledge as triple facts in the form of (head entity, relation, tail entity) or (h, r, t), indicates the specific relation between two entities.

The symbolic representation of KGs with triples is effective in representing structured data, however, with the increased size of KGs, computation inefficiency and data sparsity become serious in various applications related with KGs that people designed in a graph-based method. Recently, a new approach named knowledge graph embedding has been proposed to embed knowledge triples which include entities and relations into a continuous low-dimensional vector space. The embedding from such representation methods contain rich semantic information and can significantly promote a broad range of downstream tasks such as knowledge acquisition and inference [12]–[14].

Most previous representation methods solely learn from fact triples observed in a KG [15]–[24]. In fact, the vast majority of KGs store knowledge acquired in a text-based form, and the construction of KGs often stems from text-based knowledge extraction. So it can be said that the entity descriptions contain rich and important knowledge information, and it is also one of the multi-source information that can interact with the knowledge base. For example, Figure 1 shows descriptions of two entities extracted from the Freebase, in which the content contains more useful and rich semantic knowledge related with the entities.

Considering the power of entity descriptions in knowledge acquisition and inference, combining knowledge graph embedding models with the entity description became a

The associate editor coordinating the review of this manuscript and approving it for publication was Noel Crespi.

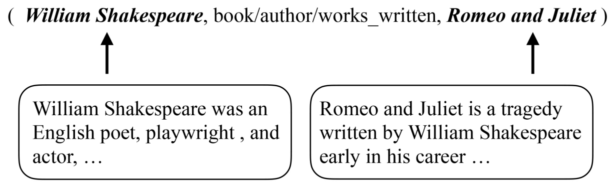

FIGURE 1. An example of entity descriptions in Freebase.

growing interest area recently. Socher et al. proposed NTN [25], a method which used the average representation of each word contained in the entity name to represent the entity and share the same word embedding with different and similar entities. Zhang *et al.* [26] followed a similar approach, representing entity as the average of word embedding in its corresponding text descriptions. Xie *et al.* [27] encoded the descriptions with the Bag-of-Words encoder and Convolutional Neural Network (CNN) encoder to obtain the semantics related with entities. Xiao *et al.* [28] proposed the semantic space projection (SSP) model which jointly learns from the symbolic triples and textual descriptions. Although these models are able to learn better representations after encoding and integrating descriptions, they still have their drawbacks and restrictions.

First of all, these models cannot take into account the context and word order well when they encoding and extracting the semantic information from descriptions. However, for the natural language, especially for the sentence and document, it is quite important to fully consider the context of each word for text representation and semantic extraction. Previous methods fail to efficiently obtain the semantic information from entity descriptions when presenting the models to improve KG embedding. Furthermore, these models only made a one-time injection of entity descriptions into the learning of KG embedding. We argue that there are also strong relationships between the entities of KG and their corresponding words of descriptions. Through the iterative guidance from the KG embedding and word embedding of descriptions, the semantic information of entity embedding can be improved. Yet, despite this merit, the iterative guidance between the KG embedding and description representations have not been well studied in previous methods.

This paper proposes Entity Descriptions-Guided Embedding (EDGE), a novel method for learning KG embedding with semantic guidance from entity descriptions. EDGE enables an embedding model to learn simultaneously from 1) knowledge triples that have been directly observed in a given KG, and 2) entity descriptions which have rich semantic information about these entities. Specifically, we propose a hierarchical Bi-directional Long Short-Term Memory (BiLSTM) max pooling encoder to encode entity descriptions and enhance the learning of knowledge graph embedding. Then we integrate such learned entity representation to constraint each word embedding of entity descriptions in an iterative guidance model. Through this iterative procedure, semantics of entity descriptions may be better transferred into the learned embedding. We evaluate the effectiveness of EDGE in link prediction and entity classification on Freebase and YAGO. Experimental results reveal that: 1) with entity descriptions injected, EDGE achieves significant improvements over state-of-the-art baselines; and 2) compared to those one-time injection schemes studied before, the iterative guidance strategy maximizes the utility of entity descriptions for KG embedding, and indeed achieves substantially better performance.

II. RELATED WORK

The representational basis for a broad range of downstream related with KG is knowledge graph embedding, which embed knowledge triples include entities and relations into a continuous low-dimensional vector space. In recent years there are a variety of methods that learn such representations from fact triples observed in a KG directly. Bordes *et al.* [31] proposed TransE, a translation-based model which indicates that relation embedding is a translation from head entity embedding to tail entity embedding. Wang *et al.* [15] proposed a TransE-based model TransH, which makes the same entity have different representations in the hyperplane specified by different relationships and alleviates the multi-mapping attribute relationship problems. Lin *et al.* [16] proposed TransR, which establishes representations of entities and relationships in separate physical and relational spaces. The entities are mapped to the relational space through the corresponding mapping matrix. Ji *et al.* [41] introduced TransD that considering the interaction between entities and relationships in a more granular way. Ji *et al.* [42] proposed TranSparse, replacing the general transfer matrix with an adaptive sparse transfer matrix.

Considering the power of entity descriptions in knowledge acquisition and inference, there are also several methods using the textual information from the internet and database to enhance the performance of knowledge graph embedding. Socher et al. proposed NTN [25], a method which used the average representation of each word contained in the entity name to represent the entity and share the same word embedding with different and similar entities. Zhang *et al.* [26] followed a similar approach, representing entity as the average of word embedding in its corresponding text descriptions. Xie *et al.* [27] encoded the descriptions with the Bag-of-Words encoder and Convolutional Neural Network (CNN) encoder to obtain the semantics related with entities. Xiao *et al.* [28] proposed the semantic space projection (SSP) model which jointly learns from the symbolic triples and textual descriptions. Ringsquandl *et al.* [29] proposed an embedding model that using event logs, the entity sequence that may occur in the KG, to improve the performance of knowledge graph embedding. Jiang *et al.* [30] presented recursive neural knowledge network (RNKN), which combines semantic information of knowledge graph with recursive neural network for multi-disease diagnosis. Guan *et al.* [44] proposed a model named

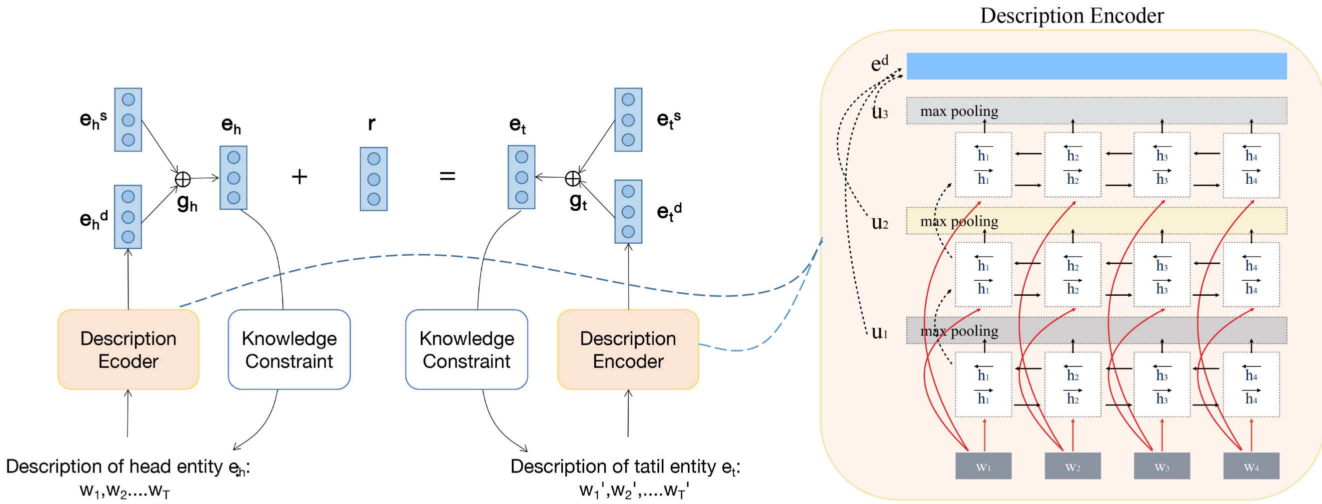

FIGURE 2. The overall architecture of our model (EDGE) and the structure of entity description encoder.

KEC, which jointly embeds entities and their concepts into a semantic space.

III. METHODS

This section introduces Entity Descriptions-Guided Embedding (EDGE), a novel paradigm for learning the knowledge graph embedding with semantic guidance from entity descriptions. To utilize both fact triples and rich textural information contained in entity descriptions, we first propose a Hierarchical Bi-directional Long Short-Term (BiLSTM) Max Pooling Memory Encoder to better represent and obtain the semantic information of entity descriptions, and then integrates such learned KG representations to constraint each word embedding of entity descriptions. The overall framework of our model EDGE is shown in Figure 2, the encoded embedding $\boldsymbol{e}_h^d$ and $\boldsymbol{e}_t^d$ are obtained through the descriptions encoder, and then integrate the structural embedding $\boldsymbol{e}_h^s$ and $\boldsymbol{e}_t^s$ respectively through the gate $\boldsymbol{g}_h$ and $\boldsymbol{g}_t$ to represent the head entity embedding $\boldsymbol{e}_h$ and tail entity embedding $\boldsymbol{e}_t$. In what follows, we first introduce our basic embedding model, and then prescribe the hierarchical BiLSTM max pooling encoder, the knowledge constraint method, and the optimization process.

A. THE BASIC EMBEDDING MODEL

To model triples, we follow TransE [31], a simple and efficient translation model which can achieve state-of-the-art predictive performance on many KG related tasks [4]–[11]. Specifically, given a set of triples $O = \{(e_i, r_k, e_j)\}$ in a knowledge graph, each triple contains two entities $e_i, e_j \in E$ and the relation $r_k \in R$. While E is the set of entities and R is the set of relations in a given KG. Then we assume each head entity e_h, tail entity e_t and relation r have a corresponding embedding, i.e., head entity embedding $\boldsymbol{e}_h$, tail entity embedding $\boldsymbol{e}_t$, and relation embedding $\boldsymbol{r}$. The energy function is then defined as

$$E(e_h, r, e_t) = \|\boldsymbol{e}_h + \boldsymbol{r} - \boldsymbol{e}_t\| \tag{1}$$

Which indicates that relation embedding $\boldsymbol{r}$ is a translation from $\boldsymbol{e}_h$ to $\boldsymbol{e}_t$, the tail representation $\boldsymbol{e}_t$ should be the nearest neighbour of $\boldsymbol{e}_h + \boldsymbol{r}$. TransE is a state-of-the-art KG embedding method which performs well in most KG related tasks such knowledge graph completions and entity classification.

B. HIERARCHICAL BILSTM MAX POOLING ENCODER

In fact, the vast majority of KGs store knowledge acquired in a text-based form, and the construction of KGs often stems from text-based knowledge extraction. So it can be said that the entity descriptions contain rich and important knowledge information, and it is also one of the multi-source information that can interact with the knowledge base. To encode the rich semantic representation of a given entity description, we need to translate the sentence into a fixed-length embedding in the vector space. There are many methods used in text or sentence representations. These methods generally first translate words into word embeddings through a projection layer and then combine such embeddings with different architectures such as Recurrent Neural Network (RNN) [32]–[34], Convolutional Neural Network (CNN) [35], [36]. In this paper, to better encode the semantics of entity description, we focus on the sentence embedding approach. Consider the powerful performance of BiLSTM architecture, we use Hierarchical BiLSTM with Max Pooling to encode the entity descriptions.

Given a sentence of entity description, we first embed the individual words with pre-trained word embedding through Word2vec [37], a toolkit developed by Google for learning word embeddings that can quickly and effectively express a word into a vector-based on a given corpus. Then the sequence of these word embeddings are inputted into our hierarchical BiLSTM max pooling encoder.

Given a sentence of words $(w_1, \ldots, w_T)$, the output of the bi-directional LSTM is a set of vectors $(h_1, \ldots, h_T)$, where each h_t of $(h_1, \ldots, h_T)$ is the concatenation

$$h_t = \left[\overrightarrow{h_t}, \overleftarrow{h_t}\right] \tag{2}$$

of a forward $\overrightarrow{h_t}$ and a backward $\overleftarrow{h_t}$

$$\overrightarrow{h_t} = \overrightarrow{LSTM}_t(w_1, \ldots w_T) \tag{3}$$

$$\overleftarrow{h_t} = \overleftarrow{LSTM}_t(w_1, \ldots w_T) \tag{4}$$

Then the max pooling layer is used to produce the maximum value for each dimension over the hidden units $(h_1, \ldots, h_T)$, while it also have the same dimensionality as h_t. Following the strong results of the BiLSTM max pooling network by Conneau *et al.* [38], we take the hierarchical BiLSTM max pooling as our encoder to improve the neural network's ability to encode and memory each words. Specifically, we assume each layer of neural network can re-read the input description.

We take three layers of BiLSTM max pooling networks as our hierarchical structure. we initialize the initial hidden state and the cell state of the second and the third BiLSTM layer with the final hidden and cell state of previous layer. Also, through the max pooling after each BiLSTM layer, we take the max values over each dimension of the hidden units. The final output embedding $\boldsymbol{e_d}$ is the average of all three max pooling vector $\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}$.

C. KNOWLEDGE CONSTRAINT

We argue that entity descriptions can better enhance KG embedding, however, in an interactive manner. Given the learned entity embedding, its corresponding word embedding and entity description, the learned entity embedding can be used to refine its corresponding word embedding and description representation. Then, the newly rectified representations of words and sentence, in turn, will help to learn better KG embeddings. To achieve such interactive guidance from entity descriptions, we take the constraint model and use the Euclidean distance to define the distance between a pair of vectors. As shown in Figure 3, we want the embedding of word *Romeo and Juliet* (grey) in entity description to be closed with the embedding of entity *Romeo and Juliet* (blue) and its neighbours in the given knowledge graph. Specifically, we want the word embedding w_i related with the entity in the description to be closed to the learned entity embedding e_i and its neighbor e_j, $\forall j$ in the given KG while the relation $(e_i, e_j) \in R$, the following optimization problem becomes

$$\min \sum_{i=1}^{|V|} \left[\alpha_i \|w_i - e_i\|^2 + \sum_{(e_i,e_j)\in R} \beta_{ij} \left\|w_i - e_j\right\|^2\right] \tag{5}$$

where V is the vocabulary of words, α and β are adjustable values used to control the relative strengths. This optimization problem is convex, and its solution can be found by solving a system of linear equations. To retrofit the word embedding w_i, the updating procedures are

$$(\alpha_i + \sum_{(e_i,e_j)\in R} \beta_{ij})w_i - \alpha_i e_i - \sum_{(e_i,e_j)\in R} \beta_{ij} e_j = 0 \tag{6}$$

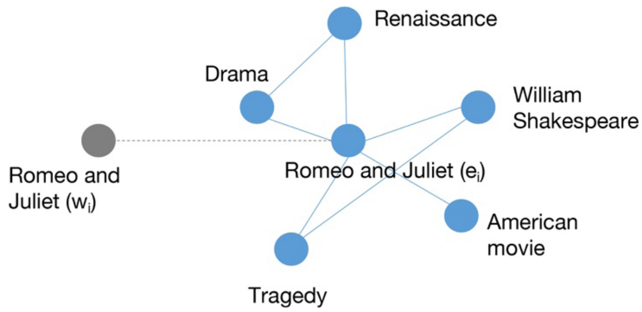

FIGURE 3. An example shows the entity and its neighbors in a given KG (blue), and its corresponding word in entity description (grey).

Then the updated word embedding w_i is

$$w_i = \frac{\sum_{(e_i,e_j)\in R} \beta_{ij} e_j + \alpha_i e_i}{\sum_{(e_i,e_j)\in R} \beta_{ij} + \alpha_i} \tag{7}$$

This retrofitting approach described above is modular which means that it can be applied to word embeddings obtained from any model.

D. TRAINING

Since both structure and entity descriptions can provide useful and rich semantic information on KG embedding learning, we integrate both of them into a joint representation. For an entity e, we propose two kinds of embedding to represent it include e^s, the structural embedding of e, and e^d, the encoded representation from entity description. To combine such two kind of information, we apply a gating mechanism when combining the structural representation of entity and entity representation from descriptions. The final joint entity embedding e is obtained through the combination between the e^s and e^d.

$$e = g_e \odot e^s + (1 - g_e) \odot e^d \tag{8}$$

where g_e is an element-wise multiplication, $\odot$ is an element-wise multiplication, each parameters of g_e are in [0, 1] to balance these two kinds of entity representations. Specifically, when the gate is close to 1, the embedding of entity description will be ignored. Consider the difference between information contained in each dimensions, we also set g_e as a vector. To constrain the value of each element of g_e is in [0, 1], we compute the gate through the logistic sigmoid function to

$$g_e = \sigma(\tilde{g_e}) \tag{9}$$

where $\tilde{g_e} \in \mathbb{R}^d$ is a real-value vector and once $\tilde{g_e}$ is learned, it keeps unchanged. The objective function we used for training is to minimize the following score function

$$L = \sum_{(e_h,r,e_t)\in O} \sum_{(e_h',r',e_t')\in O'} \max(\gamma + d(\boldsymbol{e_h} + \boldsymbol{r}, \boldsymbol{e_t}) - \boldsymbol{d}(\boldsymbol{e_h'} + \boldsymbol{r'}, \boldsymbol{e_t'}), 0) \tag{10}$$

where $\gamma > 0$ is a margin hyper-parameter, $d(\boldsymbol{e_h} + \boldsymbol{r}, \boldsymbol{e_t})$ is defined as a distance function to measure the distance

TABLE 1. Statistics of datasets used in experiments.

Dataset	#Rel	#Ent	#Train	#Valid	#Test
FB15k	1,341	14,904	472,860	48,991	57,803
WN18	18	40,837	141,322	5,000	5,000

between $\boldsymbol{e_h} + \boldsymbol{r}$ and $\boldsymbol{e_t}$. In this paper we use L1-norm as our distance function. O' is the negative sampling set of O, we have

$$O' = \left\{\left(e_h', r, e_t\right) | e_h' \in E\right\} \cup \left\{\left(e_h, r, e_t'\right) | e_t' \in E\right\} \cup \left\{\left(e_h, r', e_t\right) | r' \in R\right\} \quad (11)$$

E. OPTIMIZATION

The parameters of hierarchical BiLSTM max pooling are randomly initialized, each word embeddings of entity descriptions is pre-trained by a state-of-the-art word embedding model, Word2Vec [37], and the structural embeddings of KG triples could either be pre-trained with existing basic KG embedding models or initialized randomly. Then we use stochastic gradient descent (SGD) as the standard back propagation in optimization. The optimization procedure is applied top-down through our EDGE model until the basic layer to rectify parameters of each layer.

IV. EXPERIMENTS AND ANALSIS

A. DATASETS

In this paper, we adopt two datasets: FB15K [31] and WN18 [39] to evaluate EDGE on link prediction entity classification. The former dataset is a subgraph of a typical large-scale knowledge graph Freebase [1] containing 1,345 relations, 14,951 entities and 592,213 triples in total. The entity descriptions we use is extracted from the Wiki, and to confirm that each entities have the corresponding description embedding, we remove 47 entities from the dataset which have no descriptions, and we also remove all the triples contained these entities. The final dataset we used contains 1,341 relations and 472,860 triples, and test set has 57,803 triples. The latter dataset is a subset extracted from the WordNet [40] containing 18 relations and 40,837 entities and the test set has 5000 triples. As shown in Table 1, we list the statistical data of WN18 and FB15k.

B. EXPERIMENT SETTINGS

The KG entity/relation dimension n we trained is among {50,100,200}, the two learning rates λ_s, λ_d among {0.001, 0.002, 001}, the regularization η among {0, 1E−5, 1E−6}, and the margin γ among {1.0, 1.5, 2.0} to learn the parameters of structure and entity description encoding. The distance function can be set both L1-norm or L2-norm. The embeddings of each word contains in entity description are learned through Word2Vec [1]. The final optimal configurations we used after training and testing are:$\gamma = 2, n = 100, \eta = 1E - 5$, $\lambda_s = 0.001$, $\lambda_d = 0.002$, and L1 distance.

We compare EDGE with several state-of-the-art basic embedding models, including TransE [31], TransH [15], TransD [41], TransR [16], TranSparse [42], STransE [43]. These basic models rely only on triples observed in a KG and use no entity descriptions. We further take DKRL [27], SSP [28], KEC [44] as additional baselines which learn KG embedding integrated with external information such as entity descriptions. And to better evaluate the performance of EDGE, we choose the best-selected parameters presented in their original papers. In contrast, EDGE integrates the entity description information through the hierarchical BiLSTM max pooling encoder, and combine such knowledge into KG embedding in an interactive manner.

C. LINK PREDICTION

Link prediction is a kind of KG related task which aims to complete the triples given in a KG, i.e., to predict tail entity when given head entity and relation, or to predict head entity when given tail entity and relation. To evaluate the performance in link prediction, we follow the standard protocol used in [31], using the following two evaluation metrics: 1) the Mean Rank, the average rank of results which are predicted correctly, 2) HITS@N, the proportion of valid entities or relations ranked in top N predictions. We also follow the two evaluation settings, the original (possibly flawed) result is termed raw, and the newer one (which removed the false predicted triples included in the train, validation and test datasets before evaluation) as filter [31].

The experimental results of link prediction task on both WN18 and FB15k are shown in Table 2. From the Table 2, we can observe that proposed model significantly outperforms even all the baseline metrics on both datasets, which indicates that KG representations can be improved better through the external information such as entity descriptions, particularly in the interactive guidance manner. Compared to the best performing baseline KEC, EDGE achieves an improvement from 143 to 105 (filter) in Mean Rank and 94.3 to 95.5 (filter) in HITS@10 on WN18, and an improvement from 71 to 70 (filter) in Mean Rank and 78.5 to 82.1 in HITS@10 on FB15k (filter).

Besides, our model performs compatible with other state-of-the-art methods, while our raw result of HITS@10 performs a little bit worse than the method KEC (0.2 point below) on WN18 and a bit worse than the results of SSP and STransE on FB15k. The reason we consider is that our model is based on the basic translation model, TransE, although we add entity descriptions to enhance the knowledge graph embedding, the basic score function still has certain defects when comparing with the other three functions. We believe that the effectiveness and performance of our model could be improved effectively by introducing the other state-of-the-art score function, such as SSP and other concept space projection methods for KG embedding.

TABLE 2. Results on link prediction task.

Dataset	WN18				FB15k			
Metric	Mean Rank		HITS@10		Mean Rank		HITS@10	
	Raw	Filter	Raw	Filter	Raw	Filter	Raw	Filter
TransE [31]	263	251	75.4	89.2	243	125	34.9	47.1
TransH [15]	318	303	75.4	86.7	212	87	45.7	64.4
TransD [41]	224	212	79.6	92.2	194	91	53.4	73.3
TransR [16]	238	225	79.8	92.0	198	77	48.2	68.7
TranSparse [42]	223	221	80.1	93.2	190	82	53.7	79.9
STransE [43]	-	206	-	93.4	-	**69**	-	79.9
DKRL [27]	-	-	-	-	181	91	49.6	67.4
SSP [28]	168	156	81.2	93.2	154	77	**57.2**	79.0
KEC [44]	158	143	**82.8**	94.3	165	71	53.9	78.5
EDGE	**137**	**105**	**82.6**	**95.5**	**151**	**70**	**55.8**	**82.1**

D. ENTITY CLASSIFICATION

Entity classification is a crucial task related with many natural language processing (NLP) applications, which aims to predict and confirm the entities types in a triple (h, r, t). Since each entity in a triple has its corresponding types, e.g., the entity William Shakespeare has the types of person/author/English author and person/English person. It is noticed that the entity classification is a multi-label task. Following Neelakantan and Chang [45], we use the mean average precision (MAP) as our evaluation metric in the entity classification task.

TABLE 3. Results on entity classification task.

Dataset	WN18	FB15K
Metric	MAP(%)	MAP(%)
TransE [31]	80.3	87.8
TransH [15]	-	88.2
TransD [41]	-	88.0
TransR [16]	-	82.1
TranSparse [42]	-	88.5
STransE [43]	-	-
DKRL [27]	85.8	90.1
SSP [28]	-	94.4
KEC [44]	-	**95.0**
EDGE	**93.2**	**95.7**

The experimental results on entity classification task on both WN18 and FB15k are shown in Table 3. From Table 3, we can observe that EDGE outperforms the results of all baselines in entity classification task, which illustrates the importance of external knowledge information on KG embedding. On FB15k, our model achieves 95.7% while 93.2% on WN18 dataset, performs better than that of state-of-the-art methods. Overall, the experiment results on entity classification indicate that our interactive guidance manner is effective in improving the performance of KG embedding.

V. CONCLUSIONS

This paper proposes a novel method that learns knowledge graph embeddings with interactive guidance from entity descriptions, referred to as EDGE. It enables an embedding model to learn simultaneously from knowledge triples that have been directly observed in a given KG, and entity descriptions which have rich semantic information about these entities. EDGE encodes semantics of entity descriptions to enhance the learning of knowledge graph embedding, and integrates such learned representation to constraint each word embedding of entity descriptions. Through this interactive procedure, semantics of entity descriptions may be better transferred into the learned embedding. Link prediction and entity classification results on Freebase and WordNet show that EDGE achieves significant performance over some state-of-the-art baselines. Moreover, EDGE demonstrates the usefulness of interactive guidance between the KG embeddings and entity descriptions.

In the future, we would like to explore the following research directions to improve the performance of KG embedding: (1) EDGE model only considers text information related with entity descriptions into KG embedding, while there is also other information could be integrated to our model such as path information existed in KG. We will explore the integration of such kind of information with knowledge graph representation. (2) More description information such as database information could be explored to express rich semantic information, while the description encoder could be designed more complicated for improving KG embedding.

REFERENCES

[1] K. Bollacker, C. Evans, P. Paritosh, T. Sturge, and J. Taylor, "Freebase: A collaboratively created graph database for structuring human knowledge," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, Jun. 2008, pp. 1247–1250.

[2] C. Bizer, J. Lehmann, G. Kobilarov, S. Auer, C. Becker, R. Cyganiak, and S. Hellmann, "DBpedia—A crystallization point for the Web of data," *J. Web Semantics*, vol. 7, no. 3, pp. 154–165, Sep. 2009.

[3] F. M. Suchanek, G. Kasneci, and G. Weikum, "YAGO: A core of semantic knowledge," in *Proc. WWW*, May 2007, pp. 697–706.

[4] X. Huang, J. Zhang, D. Li, and P. Li, "Knowledge graph embedding based question answering," in *Proc. 12th ACM Int. Conf. Web Search Data Mining*, Feb. 2019, pp. 105–113.

[5] Z. Jia, A. Abujabal, and R. R. Saha, "TEQUILA: Temporal question answering over knowledge bases," in *Proc. 27th ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2018, pp. 1807–1810.

[6] Y. Li, Q. Miao, and J. Geng, "Question answering for technical customer support," in *Proc. CCF Int. Conf. Natural Lang. Process. Chin. Comput.* Cham, Switzerland: Springer, 2018, pp. 3–15.

[7] H. Jin, Y. Luo, C. Gao, X. Tang, and P. Yuan, "ComQA: Question answering over knowledge base via semantic matching," *IEEE Access*, vol. 7, pp. 75235–75246, 2019.
[8] C. Yan, Y. Chen, and L. Zhou. "Differentiated fashion recommendation using knowledge graph and data augmentation," *IEEE Access*, vol. 7, pp. 102239–102248, 2019.
[9] H. Wang, F. Zhang, and J. Wang, M. Zhao, W. Li, X. Xie, and M. Guo, "Ripplenet: Propagating user preferences on the knowledge graph for recommender systems," in *Proc. 27th ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2018, pp. 417–426.
[10] J. K. Tarus, Z. Niu, and G. Mustafa, "Knowledge-based recommendation: A review of ontology-based recommender systems for e-learning," *Artif. Intell. Rev.*, vol. 50, no. 1, pp. 21–48, Jun. 2018.
[11] H. Wang, F. Zhang, M. Zhao, W. Li, X. Xie, and M. Guo, "Multi-task feature learning for knowledge graph enhanced recommendation," in*Proc. World Wide Web Conf.*, May 2019, pp. 2000–2010.
[12] T. R. Goodwin and S. M. Harabagiu, "Knowledge representations and inference techniques for medical question answering," *ACM Trans. Intell. Syst. Technol.*, vol. 9, no. 2, Jan. 2018, Jan. 2018.
[13] S. Guo, Q. Wang, and L. Wang, "Knowledge graph embedding with iterative guidance from soft rules," in *Proc. 32nd AAAI Conf. Artif. Intell.*, Apr. 2018, pp. 4816–4823.
[14] K. Jahr and A. Borrmann, "Semi-automated site equipment selection and configuration through formal knowledge representation and inference," *Adv. Eng. Informat.*, vol. 38, pp. 488–500, Oct. 2018.
[15] Z. Wang, J. Zhang, J. Feng, and Z. Chen, "Knowledge graph embedding by translating on hyperplanes," in *Proc. 28th AAAI*, Jun. 2014, pp. 1112–1119.
[16] Y. Lin, Z. Liu, X. Zhu, X. Zhu, and X. Zhu, "Learning entity and relation embeddings for knowledge graph completion," in *Proc. AAAI*, Feb. 2015, pp. 2181–2187.
[17] B. Yang, W.-T. Yih, X. He, J. Gao, and L. Deng, "Embedding entities and relations for learning and inference in knowledge bases," in *Proc. ICLR*, Dec. 2015, *arXiv:1412.6575*. [Online]. Available: https://arxiv.org/abs/1412.6575
[18] C. Xiong, R. Power, and J. Callan, "Explicit semantic ranking for academic search via knowledge graph embedding," in *Proc. WWW*, Apr. 2017, pp. 1271–1279.
[19] F. Zhang, N. J. Yuan, D. Lian, X. Xie, and W.-Y. Ma, "Collaborative knowledge base embedding for recommender systems," in *Proc. SIGKDD*, Aug. 2016, pp. 353–362.
[20] W. Guo, Y. Dai, Y. Chen, X. Chen, and N. Xiong, "Distributed representation learning of knowledge graph with diverse information," in *Proc. 9th Int. Symp. Parallel Archit., Algorithms Program. (PAAP)*, Dec. 2018, pp. 227–234.
[21] M. L. Nickel Rosasco and T. Poggio, "Holographic embeddings of knowledge graphs," in *Proc. AAAI*, Mar. 2016, pp. 1955–1961.
[22] Z. Hu, X. Ma, Z. Liu, E. Hovy, and E. Xing, "Harnessing deep neural networks with logic rules," in *Proc. ACL*, 2016, pp. 2410–2420.
[23] Q. Wang, Z. Mao, and B. Wang, "Knowledge graph embedding: A survey of approaches and applications," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 12, pp. 2724–2743, Dec. 2017.
[24] Z. Wei, J. Zhao, K. Liu, Z. Qi, Z. Sun, and G. Tian, "Large-scale knowledge base completion: Inferring via grounding network sampling over selected instances," in *Proc. CIKM*, Oct. 2015, pp. 1331–1340.
[25] R. Socher, D. Chen, C. D. Manning, and A. Ng, "Reasoning with neural tensor networks for knowledge base completion," in *Proc. NIPS*, 2013, pp. 926–934.
[26] D. Zhang, B. Yuan, D. Wang, and R. Liu, "Joint semantic relevance learning with text data and graph knowledge," in *Proc. ACL-IJCNLP*, 2015, pp. 32–40.
[27] R. Xie, Z. Liu, J. Jia, H. Luan, and M. Sun, "Representation learning of knowledge graphs with entity descriptions," in *Proc. 13th AAAI Conf. Artif. Intell.*, Mar. 2016, pp. 2659–2665.
[28] H. Xiao, M. Huang, L. Meng, and X. Zhu, "SSP: Semantic space projection for knowledge graph embedding with text descriptions," in *Proc. 31st AAAI Conf. Artif. Intell.*, Feb. 2017.
[29] M. Ringsquandl, E. Kharlamov, D. Stepanova, M. Hildebrandt, S. Lamparter, R. Lepratti, I. Horrocks, and P. Kröger, "Event-enhanced learning for KG completion," in *Proc. Eur. Semantic Web Conf.* Cham, Switzerland: Springer, 2018, pp. 541–559.
[30] J. Jiang, H. Wang, J. Xie, X. Guo, Y. Guan, and Q. Yu, "Medical knowledge embedding based on recursive neural network for multi-disease diagnosis," Sep. 2018, *arXiv:1809.08422*. [Online]. Available: https://arxiv.org/abs/1809.08422
[31] A. N. A. J. Bordes Usunier Garcia-Duran Weston and O. Yakhnenko, "Translating embeddings for modeling multi-relational data," in *Proc. NIPS*, 2013, pp. 2787–2795.
[32] A. Talman, A. Yli-Jyrë, and J. Tiedemann, "Sentence embeddings in NLI with iterative refinement encoders," Aug. 2018, *arXiv:1808.08762*. [Online]. Available: https://arxiv.org/abs/1808.08762
[33] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 3104–3112.
[34] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, "Deep contextualized word representations," Feb. 2018, *arXiv:1802.05365*. [Online]. Available: https://arxiv.org/abs/1802.05365
[35] B. Hu, Z. Lu, and H. Li, "Convolutional neural network architectures for matching natural language sentences," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2042–2050.
[36] S. Sudholt and G. A. Fink, "Evaluating word string embeddings and loss functions for CNN-based word spotting," in *Proc. 14th IAPR Int. Conf. Document Anal. Recognit. (ICDAR)*, vol. 1, Nov. 2017, pp. 493–498.
[37] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 3111–3119.
[38] A. Conneau, D. Kiela, H. Schwenk, L. Barrault, and A. Bordes, "Supervised learning of universal sentence representations from natural language inference data," May 2017, *arXiv:1705.02364*. [Online]. Available: https://arxiv.org/abs/1705.02364
[39] A. Bordes, X. Glorot, J. Weston, and Y. Bengio, "A semantic matching energy function for learning with multi-relational data," *Mach. Learn.*, vol. 94, no. 2, pp. 233–259, Feb. 2014.
[40] G. A. Miller, "WordNet: A lexical database for English," *Commun. ACM*, vol. 38, no. 11, pp. 39–41, 1995.
[41] G. Ji, S. He, L. Xu, K. Liu, and J. Zhao, "Knowledge graph embedding via dynamic mapping matrix," in *Proc. IJCNLP*, Jul. 2015, pp. 687–696.
[42] G. Ji, K. Liu, S. He, and J. Zhao, "Knowledge graph completion with adaptive sparse transfer matrix," in *Proc. 13th AAAI Conf.*, Feb. 2016, pp. 985–991.
[43] D. Q. Nguyen, K. Sirts, L. Qu, and M. Johnson, "STransE: A novel embedding model of entities and relationships in knowledge bases," in *Proc. HLT-NAACL*, Jun. 2016, pp. 460–466.
[44] N. Guan, D. Song, and L. Liao, "Knowledge graph embedding with concepts," *Knowl.-Based Syst.*, vol. 164, pp. 38–44, Jan. 2019.
[45] A. Neelakantan and M.-W. Chang, "Inferring missing entity type instances for knowledge base completion: New dataset and methods," in *Proc. NAACL*, May/Jun. 2015, pp. 515–525.

WEN'AN ZHOU received the Ph.D. degree in electrical engineering from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2002.

Before joining the Faculty at BUPT, in 2002, she was with the China Academic of Telecommunications Technology (CATT), in Research and Development on IS-95 cellular system, from 1997 to 1999. She had studied broadband wireless communication technology from the University of California at San Diego (UCSD), as a Visiting Scholar, in 2007, which is supported by the National Young Teacher Program. She is currently an Associate Professor and the Director of the ICN&SEC Center, School of Computer Science, BUPT. She has published more than 100 articles in journals and conferences. She holds seven patents in mobile network area. Her research interests include wireless communication theory, radio resource management in HetNet, and machine learning. She has hosted two National Projects, including National Key Project and 863 Project. She has also participated in dozens of projects, such as National Key Project, 863 Project, and projects funded by National Natural Science Foundation of China.

SHIRUI WANG received the B.S. degree in software engineering from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2017, and the master's degree from the Beijing University of Posts and Telecommunications, Beijing, where she is currently pursuing the Ph.D. degree in computer science.

Her research interests include knowledge graph embedding, natural language inference, and machine reading comprehension.

CHAO JIANG received the B.S. degree in spatial information and digital technology from Wuhan University (WHU), Wuhan, China, in 2017. He is currently pursuing the master's degree in software engineering with the Beijing University of Posts and Telecommunications (BUPT), Beijing, China.

His research interests include machine learning knowledge graph embedding and network security.