

Learning Entity and Relation Embeddings with Entity Description for Knowledge Graph Completion

Shaozhi Dai¹, Yanchun Liang^{1,2}, Shuyan Liu¹, Ying Wang¹, Wenle Shao¹, Xixun Lin¹ and Xiaoyue Feng^{1,*}

¹Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, College of Computer Science and Technology, Jilin University, Changchun 130012, China

²Zhuhai Laboratory of Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Zhuhai College of Jilin University, Zhuhai 519041, China

*Corresponding author

Abstract—With the growth of existing knowledge graph, the completion of knowledge graph has become a crucial problem. In this paper, we propose a novel model based on description-embodied knowledge representation learning framework, which is able to take advantages of both fact triples and entity description. Specifically, the relation projection is combined with description-embodied representation learning to learn entity and relation embeddings. Convolutional neural network and TransR are adopted to get the description-based and structure-based representation of entity and relation, respectively. We employ FB15K dataset generated from a large knowledge graph freebase, to evaluate the performances of the proposed model. Experimental results show that our proposed model greatly outperforms other existing baseline models.

Keywords—knowledge graph completion; entity description; natural language processing

I. INTRODUCTION

Knowledge graphs (KGs) on a large scale such as NELL [1], Freebase [2], and WordNet [3] are important for applications in natural language processing fields, such as language modeling, web searching and so on. KGs store structured multi-relational data in the form of (head entity, relation, tail entity), which are called facts. However, despite massive facts that can be contained, most of KGs built semi-automatically or collaboratively are still far from complete and always suffered from sparseness [4]. The completion of KGs has become a critical work.

In recent years, embedding methods used in knowledge graphs completion have attractive more and more attentions [5]. Traditional embedding models embed both entities and their relations into a latent low-dimensional space. Among them, TransE [6] considers the relation as the translation between two entities and performs better than most of these models. However, TransE ignores different roles played in different entity. TransH [7] and TransR [8] are the variants of TransE, which improved TransE by setting a hyperplane for each specific relation or by modeling the relations and entities in distinct spaces, respectively. For these models, one disadvantage is that they only focus on the observed structured facts. They do not consider incorporating other information except for facts. There are many methods on entity textual description [9,10], entity types [11], rules [12] or even temporal

information [13]. Methods incorporated textual information like NTN [14] simply initialize the entity representation by averaging the word vectors.

In this paper, we propose a model which combined relation projection with description-embodied representation learning, named RDRL, to augment TransR model with entity textual information. The proposed model is based on the description-embodied knowledge representation learning (DKRL) [11]. Specifically, we set a relation-specific projection for each entity for the distinction of the entity and relation space. The advantages of our model are as follows: 1) could be used to further improve the classical model such as TransR by combining entity description, 2) could retain the merits of original embedding models and perform better on some occasions, such as dealing with the problem of data sparsity.

The remainder of this paper is organized as follows. Section II introduces the previous related work. Section III describes our proposed model. The training details are presented in Section IV. Section V presents the experiments and results. Finally, Section VI draws a conclusion.

II. RELATED WORK

Knowledge graph embedding aims to embed the relations and entities into a latent space. The learning process can be divided into three steps: (1) entity/relation representation. (2) scoring function definition, and (3) parameter estimation.

A. TransE and DKRL

The traditional model TransE embeds entities and relations into a same low-dimensional vector space. For any triple (e_i, r_k, e_j) which describe facts (head entity, relation, tail entity), TransE holds the assumption $e_i + r_k \approx e_j$ and defines the energy of a triple as follows:

$$E(e_i, r_k, e_j) = \|e_i + r_k - e_j\|_2^2 \quad (1)$$

The embedding e_i , r_k and e_j take values in R^d , which are trained by minimizing the following rank criterion that favors lower energy for the correct triplets than for the corrupted triplets, i.e.,

$$L = \sum_{(e_i, r_k, e_j) \in S} \sum_{(e'_i, r'_k, e'_j) \in S'} [\gamma + E(e_i + r_k, e_j) - E(e'_i + r'_k, e'_j)]_+ \quad (2)$$

where $[x]_+$ denotes the positive part of x , S is the set of correct triples, and S' is the set of corrupted triplets with head or tail entities replaced by any random entity. $\gamma > 0$ is a margin used to separate the correct triplets and the corrupted one.

DKRL takes advantages of TransE which is based on structure representation and also considers entity description [11]. So, given an entity, not only its fact triples but also its entity description can be modeled. In the DKRL model, continuous bag-of-words (CBOW) model and deep convolutional neural model [15] could be used as encoders to represent the semantics of specific entity descriptions. The energy function of DKRL is defined as

$$E = E_S + E_D \quad (3)$$

where E_S and E_D are the energy functions of structured-based representation like TransE and description-based representation, respectively. Since the representation of an entity can either be structure-based or description-based, E_D is defined as $E_D = E_{DS} + E_{SD} + E_{DD}$. DKRL minimizes the following rank criterion for training:

$$L = \sum_{(e_i, r_k, e_j) \in S} \sum_{(e'_i, r'_k, e'_j) \in S'} \max(0, E(e_i + r_k, e_j) - E(e'_i + r'_k, e'_j) + \gamma) \quad (4)$$

where S' is the set of corrupt triples with head entity, tail entity or relation replaced, $\gamma > 0$ is a hyper-parameter of margin that will separate corrupted triples and correct triples, and $E(e_i, r_k, e_j)$ is the energy of triple (e_i, r_k, e_j) for a dissimilarity measure E , which is L₁-norm or L₂-norm. The $\max(x, y)$ aims to return the maximum value between x and y .

B. Other Models

There have been other attempts on knowledge graph embedding based on entity description [14, 16]. In this paper, we take advantage of the entity textual information to impute the lack of useful information and improve the performance of knowledge graph completion.

III. THE PROPOSED METHOD

To solve the problems of TransE such as the problem of lower distinctiveness when handling complex relationship except for the one-to-one relation, we proposed a model, named RDRL, which combined relation projection with description-embodied representation learning by using TransR instead of TransE. Relation information can be beneficial if it can be used efficiently. Since the continuous bag-of-words model ignores word orders in text and is easier to be affected by the quality of extracted keywords when encoding semantics of entity descriptions, we use convolutional neural network to represent the description embedding.

Given any triple (e_i, r_k, e_j) indicating that there is a relationship between e_i and e_j , we follow the treatment in TransR and use a projection matrix M_{r_k} to map the head entity e_i and the tail entity e_j from their entity space to the specific relation space. The mapping can be described as:

$$e_{i_k} = M_{r_k} e_i, e_{j_k} = M_{r_k} e_j, \quad (5)$$

where $e_i, e_j \in R^k$, $r_k \in R^d$, and k and d are the dimensions of entity embedding and relation embedding, respectively. Using the mapping matrix, the score function can be defined as

$$E_S(e_i, r_k, e_j) = \|e_{i_k} + r_k - e_{j_k}\|_2^2 \quad (6)$$

Since the representation e_{i_k} and e_{j_k} can directly represent the entities in their vector space, the representation is based only on the structure. After getting the structured-based representations for the entities, we combine the structured-based representations with the description-based representations. Finally, we define the energy function as

$$E(e_i, r_k, e_j) = E_S(e_i, r_k, e_j) + E_D(e_i, r_k, e_j) \quad (7)$$

where E_S shares the same formulation as (6). To make the two representations of entities representations compatible, we define $E_D(e_i, r_k, e_j)$ as

$$E_D(e_i, r_k, e_j) = E_{DD}(e_i, r_k, e_j) + E_{DS}(e_i, r_k, e_j) + E_{SD}(e_i, r_k, e_j) \quad (8)$$

in which $E_{DD}(e_i, r_k, e_j)$ is totally based on the description representation. We also have

$$E_{DS}(e_i, r_k, e_j) = \|e_{i_d} + r_k - e_{j_s}\| \quad (9)$$

and

$$E_{SD}(e_i, r_k, e_j) = \|e_{i_s} + r_k - e_{j_d}\| \quad (10)$$

where e_i is the description-based representation and e_j is the structure-based representation. The relation representation is shared by all the energy functions and will contribute to the mutual promotion of the two different representations. Our model will finally project both the entity representations into the same vector space with the energy function defined.

IV. TRAINING DETAILS AND IMPLEMENTATION DETAILS

We train our RDRL as the procedure of DKRL, and adopt stochastic gradient descent to minimize the final margin-based

objective function which is defined as (4). We denote $S = \{(e_i, r_i, e_i) | i=1, 2, \dots, t\}$ as the set of corrected triples and S' is the set of corrupted triples, which is defined as

$$S' = \{(e'_i, r_i, e_i) \cup (e_i, r'_i, e_i) \cup (e_i, r_i, e'_i) | i = 1, 2, \dots, t\} \quad (11)$$

where each element of S' is constructed by randomly replacing an entity or relation. Note that our RDRL is not restricted to the type of entity representation.

In our training process, the convolutional kernels are initialized randomly. The vectors of words are pre-trained by GloVe [17] on Wikipedia. The CNN encoder takes the word vectors as inputs and outputs of the corresponding entity description-based embeddings. And we use a multi-thread version of CNN to learn the representations for a better efficiency. Structure-based embeddings of entities and relations are initialized with results of TransE.

V. EXPERIMENTS

A. Data Sets

We adopt FB15K which is built with a commonly used knowledge graph Freebase [2] to evaluate the proposed methods. FB15K contains 1345 relations and 14951 entities which provides massive information of the world stored in structure e.g., /medicine/disease/risk_factors.

Entities with no descriptive words or shorter than 3 words are not included to ensure that all entities should have description in the test period. Unlike DKRL, we use the word embeddings trained by GloVe as the representation of entities. Table I shows the final statistics of the data sets we used, where the number of the relation types, entity types, and triples has already been split into training, validation and testing sets.

TABLE I. STATISTICS OF THE DATA SETS USED IN OUR EXPERIMENT

Dataset	Relation	Entity	Train	Validation	Test
FB15K	1,341	14,904	483,142	48,991	57,803

B. Knowledge Graph Completion

Knowledge graph completion aims to predict the missing entity or relation in a corrupted triple (e_i, r_k, e_j) with e_i, r_k or e_j missing.

• Parameter Settings

We compared our RDRL with some related existing models. The configurations for our experiments are given as follows: entity/relation dimension n in $\{20, 50, 100\}$, the fixed learning rate α among $\{0.001, 0.002, 0.01\}$, and margin γ among $\{0.5, 1.0, 1.5\}$. For the CNN encoder, the window size k among $\{1, 2, 3\}$, the word embedding dimension w among $\{50, 80, 100\}$, and the feature map dimension f among $\{50, 100, 150\}$. The optimal configuration on the valid set are: $n=100, \alpha=0.001, \gamma=1.5, k=2, w=100, f=100$.

• Evaluation Measurement

We evaluate our model on the two sub-tasks: entity prediction and relation prediction. For the entity prediction, we replace the head entity with every entity existing in the entity data set and use the score function to compute its score. Later we rank their scores in a descending order. For the tail entity, we repeat the same procedure as the head entity. And for the relation prediction, we use the same method by replacing the relation with each relation in the relation data set. As used in [8], we employ the same metrics: (1) the mean of the predicted ranks; (2) the hits@10 for entity and the hits@1 for relation which respectively means the proportion of the correct entities ranked in top 10 and the proportion of the correct relations ranked in top 1. There exist many entities in the same incomplete triple which are not sure to be implausible. And if they rank before the replaced entity or relation, they will cause results that are not ideal. Thus, we also follow the evaluation setting Raw and Filter. Filter indicates that the method removes all the other plausible entities or relations in the descent ranking. Raw means that the method does not consider the problem above.

• Results

We list the evaluation results of entity prediction in Table II and relation prediction in Table III. By comparing with the listed models, it can be seen from Table II that our RDRL greatly outperforms the original DKRL, and it shows consistent and significant improvement over all the following listed models on each metric in Table III. TranE, TransH(unif), TransH(bern), TransR(unif), DKRL(CBOW), and DKRL(CNN) represent methods from [6, 7, 7, 8, 8, 11]. DKRL(CNN)+TransE(ours) represent DKRL methods based on CNN and TransE. RDRL represents our proposed model.

TABLE II. COMPARED RESULTS ON ENTITY PREDICTION

Metric	Mean Rank		Hits@10(%)	
	Raw	Filter	Raw	Filter
TransE	243	125	34.9	47.1
TransH(unif)	211	84	42.5	58.5
TransH(bern)	212	87	45.7	64.4
TransR(unif)	226	78	43.8	65.5
DKRL(CBOW)	236	151	38.3	51.8
DKRL(CNN)	200	113	44.3	57.6
DKRL(CNN)+TransE(ours)	184	96	45.4	60.4
RDRL	180	89	45.9	62.3

TABLE III. COMPARED RESULTS ON RELATION PREDICTION

Metric	Mean Rank		Hits@1(%)	
	Raw	Filter	Raw	Filter
TransE	2.91	2.53	69.5	90.2
DKRL(CBOW)	2.85	2.51	65.3	82.7
DKRL(CNN)	2.91	2.55	69.8	89.0
DKRL(CNN)+TransE	2.41	2.03	69.8	90.8
DKRL(CNN)+TransE(ours)	2.34	1.97	70.4	90.3
RDRL	2.09	1.68	71.3	93.6

C. Triple Classification

Triple classification has already been explored by [14] which is a binary classification problem. The task aims to judge whether a given triple is true or not. In the experiment, we evaluate our model on the benchmark dataset FB15K. For

the task of binary classification, we need negative triples. Since negative triples are not included in the released FB15K dataset, we produce negative samples following the same process in [14]. To perform the task, we set different thresholds $\{\delta_r\}$ for different relations. Given a triple (e_i, r_k, e_j) , if its score $E(e_i, r_k, e_j)$ is below the threshold δ_r , we classify the triple as positive, otherwise negative. The threshold δ_r is set by maximizing the classification accuracy on the validation set.

TABLE IV. COMPARED RESULTS ON TRIPLE CLASSIFICATION (%)

Method	Result(FB15K)
TransE(unif)	77.8
TransE(bern)	85.3
TransH(unif)	78.4
TransH(bern)	85.8
TransR(unif)	79.2
TransR(bern)	87.0
DKRL(cnn)+TransE(ours)	79.3
RDRL	89.7

We compare our model with some existing work using the results showed in [8] and use the code of DKRL which is released in [11] to complete the task. We adopt the accuracy of the classification as the evaluation metric. Table IV shows the results of triple classification on FB15K dataset. From Table IV, we can see that RDRL performs best on FB15K dataset. Specifically, RDRL improves the performance of TransR(unif) by 13.3% and improves the performance of DKRL by 13.1%.

TABLE V. STATISTICS OF FB15K DATA SETS

Frequency	1-3	4-15	16-50	51-300	>300
Number of relation	738	252	165	142	48
Number of triples	585	1941	4810	17468	32999
Mean Rank of DKRL(Filter)	116	109	111	106	173
Mean Rank of RDRL(Filter)	105	98	102	101	103
Improvement (%)	9.5	10.1	8.1	4.7	40.5

D. Discussions

In the first section it is pointed out that, although the massive facts FB15K can be contained, the KG still suffers from the problem of data sparseness and thus some relations may not be captured correctly. So, for further analysis, we divided the relations into five groups according to the frequency, as shown in Table V. The frequency was obtained according to the times they occur in the test set. We compare MeanRank(filter) of DKRL and RDRL. Table V reports the results and the improvement from DKRL to RDRL. From the results, we can see that: (1) for the relation whose frequency is 1 to 3, the improvement from RDRL to DKRL is 9.48%, which indicates that separating the entity space and the relation space does better when handling the data sparsity problem, (2) the most significant improvement occurs when handling the relation with frequency greater than 300, which shows that RDRL performs better for most triples in the KG.

VI. CONCLUSIONS

Jointly learning the embeddings of KGs with the multi-meta information has gained increasingly interests. In this paper, we propose the RDRL model by combining the relation

projection with description-embodied representation learning for better performances. To evaluate the performance of our proposed model, experiments are performed on the three tasks. The compared results show that our proposed model outperforms state-of-the-art models.

ACKNOWLEDGEMENTS

The authors are grateful for the support of the National Natural Science Foundation of China (61602207, 61572228, and 61472158), the Zhuhai Premier Discipline Enhancement Scheme, the Guangdong Premier Key-Discipline Enhancement Scheme, and the Educational Commission of Jilin Province.

REFERENCES

- [1] A. Carlson, J. Betteridge, B. Kisiel, B. Settles, E.R. Hruschka Jr, and T.M. Mitchell, "Toward an Architecture for Never-Ending Language Learning," in AAAI, 2010, pp. 1306-1313.
- [2] K. Bollacker, C. Evans, P. Paritosh, T. Sturge, J. Taylor, "Freebase: A Collaboratively Created Graph Database for Structuring Human Knowledge," in Proc. 2008 ACM SIGMOD Int. Conf. Manag. Data, ACM, New York, NY, USA, 2008, pp. 1247-1250.
- [3] G.A. Miller, "WordNet: A Lexical Database for English," Commun ACM, vol. 38, pp. 39-41, 1995.
- [4] R. West, E. Gabrilovich, K. Murphy, S. Sun, R. Gupta, and D. Lin, "Knowledge Base Completion via Search-based Question Answering," in Proc. 23rd Int. Conf. World Wide Web, ACM, New York, NY, USA, 2014, pp. 515-526.
- [5] Q. Wang, Z. Mao, B. Wang, and L. Guo, "Knowledge Graph Embedding: A Survey of Approaches and Applications," IEEE Trans. Knowl. Data Eng. vol 29, pp. 2724-2743, 2017.
- [6] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko, "Translating Embeddings for Modeling Multi-relational Data," in Adv. Neural Inf. Process. Syst. 26, Curran Associates, Inc., 2013, pp. 2787-2795.
- [7] Z. Wang, J. Zhang, J. Feng, and Z. Chen, "Knowledge Graph Embedding by Translating on Hyperplanes," in AAAI, 2014, pp. 1112-1119.
- [8] Y. Lin, Z. Liu, M. Sun, Y. Liu, and X. Zhu, "Learning Entity and Relation Embeddings for Knowledge Graph Completion," in AAAI, 2015, pp. 2181-2187.
- [9] H. Zhong, J. Zhang, Z. Wang, H. Wan, and Z. Chen, "Aligning Knowledge and Text Embeddings by Entity Descriptions," in ENMLP, 2015, pp. 267-272.
- [10] J. Xu, K. Chen, X. Qiu, and X. Huang, "Knowledge Graph Representation with Jointly Structural and Textual Encoding," ArXiv161108661 Cs. (2016).
- [11] R. Xie, Z. Liu, J. Jia, H. Luan, and M. Sun, "Representation Learning of Knowledge Graphs with Entity Descriptions," in AAAI, 2016, pp. 2659-2665.
- [12] S. Guo, Q. Wang, L. Wang, B. Wang, and L. Guo, "Jointly Embedding Knowledge Graphs and Logical Rules," in EMNLP, 2016, pp. 192-202.
- [13] T. Jiang, T. Liu, T. Ge, L. Sha, S. Li, B. Chang, and Z. Sui, "Encoding temporal information for time-aware link prediction," in EMNLP, 2016, pp. 2350-2354.
- [14] R. Socher, D. Chen, C.D. Manning, and A. Ng, "Reasoning With Neural Tensor Networks for Knowledge Base Completion," in Adv. Neural Inf. Process. Syst. 26, Curran Associates, Inc., 2013, pp. 926-934.
- [15] D. Ciregan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," in 2012 IEEE Conf. Comput. Vis. Pattern Recognit., 2012, pp. 3642-3649.
- [16] Z. Wang and J. Li, "Text-enhanced Representation Learning for Knowledge Graph," in AAAI, 2016, pp. 1293-1299.
- [17] J. Pennington, R. Socher, and C.D. Manning, "Glove: Global Vectors for Word Representation," in EMNLP, 2014, pp. 1532-1543.