Relation Path Modeling with Entity Description for Knowledge Graph Completion

Qi Luo and Xingwei Hao

***Abstract*—A large number of fact triples are contained in knowledge graph. However, the knowledge is not completed in large-scale knowledge graph at present, which calls for knowledge graph completion urgently. Most proposed approaches embed entities and relationships into a continuous low-dimensional vector space through a single factual triple information, the knowledge graph completion is achieved through vector calculation. Besides, features are extracted from the relation paths between entities, which are used for training classifiers to predict relationships between entities. In this work, we designed a model, which not only utilize the information of relation paths among fact triples and entities in knowledge graph, but also the description text information of entities in knowledge graph were utilized. Through these information of entities and the distributed expression of the relationships, the task of knowledge graph completion was achieved. Through extensive experimental evaluations, the proposed model proves to be with better retrieval performance compared to state-of-the-arts.**

***Index Terms*—Knowledge graph, knowledge graph completion, link prediction, relation path.**

I. Introduction

Similar to knowledge bases such as WordNet [1], Freebase [2] and DBpedia [3], which are used to express complex structural information in the real world. Knowledge graph [4] was proposed by Google in order to improve the search quality. Besides, knowledge graph also play an important role in question answering [5], named entity recognition [6], information extraction [7]. Knowledge graph aims to describe the concepts in the real world and the relationships among them. The nodes in knowledge graphs represent entities, and the label between two nodes represent the relationship between them. The knowledge graphs provide structured knowledge. A typical knowledge can be expressed as a factual tuple consisting of a head entity, a relationship and a tail entity(denoted as *(h, r, t)*). For instance, we know that the capital of China is Beijing, which can be represented as (Beijing, Capital of China) in the form of fact triple. Although the amount of data in the knowledge graph is very large, but it is still not complete, it calls for existed fact triples and the specific rules for knowledge graph completion. Recently, most proposed approaches embed entities and relationships into a continuous low-dimensional vector space through a single factual triplet information. The knowledge graph completion is achieved through vector calculation. For instance, TransE [8], TransH [9], TransR [10] and TransD [11], and these kinds of methods performs well for link prediction in large knowledge graph. Compared to these proposed approaches, relation paths modeling based methods have better performance, for examples, PRA [12], PCRA [13] and CPRA [14] take advantage of the relationship information contained in relation paths between entities in the knowledge graphs. The feature vector of relation classification is constructed by summing up the large number of relation paths between entities, and then relation classifiers were constructed for relation extraction, finally achieving the purpose of knowledge graph completion. The proposed method makes full use of the characteristics of the knowledge graph, which not only provide high prediction accuracy, but also provide more reasonable prediction explanations.

Knowledge represent methods can be a good way to capture intrinsic characteristics of the data, it is efficient and easy to manipulate, while these purely data-driven methods cannot provide reasonable explanations for final predictions. However relationship path modeling based methods can provide better explanations for link prediction, there still exists serious problems of rare relations dealing with graphs with low connectivity. In sparse graphs, the information of relation path is relatively less, which has a very serious effect on the prediction result. In the knowledge graph with more relation paths, the number of the relationship paths increases exponentially and the operation of extracting path features could be time-consuming. None of these methods take the description text of the entity in the knowledge graph into account. While these description texts contains rich semantic information, in our methods, the information of each description text is considered for encoding to represent an entity, combined with the path between the entities, vectors representation between entities and relations were studied, entities and relations are embedded into a continuous low-dimensional vector space, in which space link prediction are conducted. Our model not only uses the efficient representation of the knowledge representation methods, but also utilizes the explicit characteristics of the relationship path between entities, which provide more reasonable explanation for predict results, at the same time, description text information were also taken into account, thus the model makes full use of the diversification of knowledge base to carry out the task of knowledge graph completion.

II. Related Work

In recent years, data on the internet has grown rapidly, and in order to obtain better search results from the vast ocean of data, the knowledge graph has been proposed. In order to

Manuscript received September 23, 2018; revised November 12, 2018.

The authors are with the Computer science and technology Department, Shandong University, Shandong, China (e-mail: luoqi0918@gmail.com, hxw@sdu.edu.cn).

doi: 10.18178/ijke.2018.4.2.104

complete the knowledge graph, the task of knowledge graph complement has attracted increasing research interest. The purpose of this task is to infer the missing triples through the exist fact triple. There are many earlier methods including structured embedding based method (SE) [15], semantic matching energy (SME) [16], and so on. There are also some other well-known methods such as knowledge representation based translate based method, relation path modeling based method, besides, other methods that add the path relationship between entities to knowledge expression, for instance, PTransE model [17], which take advantage of the superiority of both knowledge expression and relation path, also have shown good performance.

In SE method, the entities are embedded into d-dimension vector space, which are represented by two d*d-dimension matrixes, the head entity *h* and tail entity *t* are projected to related space, Then distances between projected vectors are calculated, the semantic relevance between these two entities under relation *R* are represented by this distance. However, the method is less synergistic due to the use of two different matrix projections, the semantic relevance between these two entities are not described precisely.

More complex operation are proposed by SME method to seek for the semantic relevance between entities. Each entity and vector is represented by a low-dimension vector, multiple projection matrices are defined by SME, which are used to describe the intrinsic relationship between the entity and the relationship.

Knowledge represent model based on translation (TransE Model), which is motivated by word2vec [18] treat the relationship between entities in the knowledge base as some sort of translation from entities. The relation *r* can be treated as a translation of the head entity vector *h* to the tail entity vector. For each fact triple in knowledge graphs, it is hoped that $h + r \approx t$ in TransE method, and the scoring function is represented as :

$$E(h, r, t) = \|r - h - t\|_{L_1 / L_2} \tag{1}$$

From this formulation, we can conclude that the existence of fact triple *(h，r，t)* will reduce the value of the scoring function, on the other hand, the value of the scoring function will be high without fact triples. There are less parameters in this model, which contributes to better performance in large-scale knowledge graph. While this model find it hard to deal with complex relationships such as one-to-many and many-to-one relations. Based on TransE model, other expanded models such as TransH model, TransR model and TransD model perform better in dealing with complex relationships.

Another kind of method, which is talk about a lot, is called relation path modeling based method. Lao and Cohen, et al, have proposed PRA algorithm, the main idea of PRA is to extract the relationship path that connect two entities as a feature to predict the relationship between the two entities. A typical PRA model is consist of three main steps: feature extraction, feature computation, relational classifier. Although these kinds of methods have achieved good efficiency, due to the reason that they are all relationship statistics based methods, there are only satisfied on connectivity diagrams, there still exists serious problems of rare relations dealing with graphs with low frequency relationships. In view of these problems, there have being many researches that improve in these algorithms.

III. Methodolgy

A. Problem Formulation

Suppose that given a set of fact triples in a given knowledge graph: $S = \{(h, r, t)\}$, in which $h, t \in E$ represent entities, $r \in R$ represent relations, E，R represent the sets of entities and relationships, *S^'* is the set of negative samples of *S*.

$$\begin{aligned} S' = & \{(h', r, t) | h' \in E\} \\ & \cup \{(h, r', t) | r' \in R\} \\ & \cup \{(h, r, t') | t' \in E\} \end{aligned} \tag{2}$$

A negative sample is a head entity, and a tail entity or relationship can be arbitrarily replaced by another entity or relationship in *S*, if the triple has been in *S*, it cannot be used as a negative sample.

We assume that h_e, t_e are the entity vectors obtained by the TransE model, h_h, t_h are the entity vectors obtained by the TransH model, h_p, t_p are the entity vectors obtained by the entity description text.

B. DPTransE&DPTransH

Suppose that there are multiple connected paths between the two entities *h* and *t*, and the relationship path is expressed as $p = (r_1, r_2, \ldots\ldots, r_l)$, according to the path-constraint resource allocation algorithm, for the existed fact triple *(h, r, t)*, The scoring function about path can be expressed as:

$$E_{path}(h, p, t) = \frac{1}{Z} \sum_{p \in P(h,t)} R(p \mid h, t) E(h, p, t) \tag{3}$$

where S_i is the set of entities from relation r_i to entity *n*. For the relation path *p* and head node *h*, $R_p(h) = 1$, and we can get fact triple *(h, p, t)*: $R(p | h, t) = R_p(t)$.

A brief description is supplied for each entity in the knowledge base, according to the method of representing a text as a vector, which was proposed by [19], we got a vector called Paragraph Vector, which was used to represent the corresponding entity of the description text. There are two kinds of training methods: Distributed Memory version of Paragraph Vector (PV-DM) and Bag of Words version of Paragraph Vector (PV-BOW). After changing the description text of the entity to a vector, the scoring function based on the TransE model is represented as:

$$\begin{aligned} E_{EP}(h, r, t) = & \|h_e + r - t_e\| + \|h_e + r - t_p\| \\ & + \|h_p + r - t_e\| + \|h_p + r - t_p\| \end{aligned} \tag{4}$$

where h_e , t_e are the entity vectors obtained by the TransE model, and h_p , t_p are the entity vectors obtained by doc2vec.

On the basis of the TransH model, the entity is projected onto the hyperplane with the normal vector w_r , the projection vectors of *h* and *t* on the hyperplane are $h_\perp$, $t_\perp$, the head vector and tail vector of the solid projection on the hyper plane are connected by d_r with a lower error, by limiting $\|w_r\|^2 = 1$, the score function can be defined as:

$$E_H(h, r, t) = \|h_\perp + d_\perp - t_\perp\| \tag{5}$$

Which is:

$$E_H(h, r, t) = \left\|\left(h - w_r^\perp h w_r\right) + d_r - \left(t - w_r^\perp t w_r\right)\right\| \tag{6}$$

Apply to the entity vector of the text representation and the TransH training entity vector:

$$\begin{aligned} E_{HP}(h, r, t) = {} & E_H(h_e, r, t_e) \\ & + E_H(h_e, t, t_p) + E_H(h_p, r, t_e) \\ & + E_H(h_p, r, t_p) \end{aligned} \tag{7}$$

We utilize the margin-based optimization method, the loss function is defined as:

$$L = \sum_{(h,r,t)\in S}\left[L(h, r, t) + \frac{1}{Z}\sum_{p\in P(h,t)} R(p \mid h, t) L(h, p, t)\right] \tag{8}$$

Among which:

$$L = \sum_{(h',r',t')\in S^-}\left[\gamma + E_{path}(h, p, t) - E_{path}(h', p, t')\right] \tag{9}$$

Based on TransE model,

$$L = \sum_{(h',r',t')\in S^-}\left[\gamma + E_{EP}(h, r, t) - E_{EP}(h', r', t')\right] \tag{10}$$

Based on TransH model,

$$L = \sum_{(h',r',t')\in S^-}\left[\gamma + E_{HP}(h, r, t) - E_{HP}(h', r', t')\right] \tag{11}$$

In the loss function $[x]_+ = \max(x, 0)$, γ is the margin, and S is the fact triple present in the knowledge graph, S^- is the counterexample in the knowledge graph, we construct a counter-example by randomly replacing the existing entity in the existing triplet or tail entity.

In the optimization process, the stochastic gradient descent method is utilized, during the process, $\|h\| \le 1$, $\|w_r\| \le 1$, $\|t\| \le 1$, the path length is selected as 2-step or 3-step, in freebase, if 4-step is selected, and there will be billions of paths between the two entities, the calculation consumption will increase beyond our imagination.

IV. Experiments

A. Dataset and Experiments Setting

FB15K was utilized as the database for experiments, which is a subset of a large-scale knowledge graph Freebase, the data statistics in FB15K is shown below in Table I, we evaluate the performance of our model and other models in the task of knowledge graph complement on this database.

TABLE I: Dataset Statistics

Dataset	#Relation	#Entity	#Train	#Valid	#Test
FB15k	1341	114904	472860	48991	57803

B. Experiment

Two experimental schemes are adopted to evaluate the performance of the model: (1) entity prediction, (2) relationship prediction.

Entity prediction means that when the head entity or tail entity in the test tuple is lost, we can reason out the possible head entity or tail entity, for each test tuple *(h, r, t)*, we replace its head entity or tail entity with each entity in the dataset, according to the score from the scoring function *E (h, r, t)* {score function} and the score in ascending order, the possible candidates in the data set are arranged.

Two indicators are adopted as the evaluation index, which are the mean Rank of all entities and the proportion of valid entities in the top 10 entities. This method is feasible but flawed, for example, given a fact triple (Obama, President Of, USA), If the head node of Obama is missing, the head node of Lincoln may be treated as an invalid prediction, but it is useful in the knowledge graph, we will filter out these valid entities when we rank these entities. The unfiltered predictions are called "Raw" and otherwise are called as "Filter". Here we use a reordering method: firstly, all candidate entities are ranked by the score of TransE, and then the top 500 candidate entities are rearranged by the score of PTransE, the result is as follows:

TABLE II: The Evaluation Results of the Entities

	Mean Rank		Hits@10(%)	
	Raw	Filter	Raw	Filter
SE	273	162	28.8	44.1
SME	274	154	30.7	40.8
TransE	243	125	34.9	47.1
TransH	212	87	45.7	64.4
PTransE	200	54	51.8	83.4
DPTransE	187	46	56.7	87.3
DPTransH	192	42	58.8	89.2

The experimental results are consistent with our expectation, the trend of raw and filter is the same. From TABLE II we can see, compared to the TransE, TransH and TransR models, our model have obtained more significant performance in Mean Rank, on the Hits@10, DPTransH model is better DPTransE model, which means that DPTransH model performs better than DPTransE model in dealing with complex relationships.

The goal of relational prediction is to predict the relationships that may exist between two given entities. For each test tuple *(h, r, t)*, we replace the tuple with each possible relation in the knowledge graph, and then the score of scoring function was calculated, all the candidate relationships in the dataset are ranked by their ranking in ascending order. Two evaluation criteria are utilized in our experiments: Mean Rank and Hits@1.

The results of the relationship prediction are shown in TABLE III, from the table we can see that DPTransH performs better than other models on Mean Rank and Hits@1, the results of TransE, TransH, TransR models are similar. This indicates that the way that we consider the relationship between the path and the entity description text for the entity and the representation of the relationship is valid. The relationship in the knowledge map is divided into one-to-one, one-to-many, many-to-many and many-to-many, we evaluated prediction performance under these four.

TABLE III: THE RESULT OF RELATIONSHIP PREDICTION

	Mean Rank		Hits@1(%)	
	Raw	Filter	Raw	Filter
TransE	243	125	34.9	47.1
TransH	212	87	45.7	64.4
PTransE	200	54	51.8	83.4
DPTransE	187	46	56.7	87.3
DPTransH	192	42	58.8	89.2

TABLE IV: THE RESULT OF RELATIONSHIP PREDICTION

	Predict Head Entities(Hits@10)				Predict Tail Entities(Hits@10)			
	1 to 1	1 to N	N to 1	N to N	1 to 1	1 to N	N to 1	N to N
SE	35.6	62.6	17.2	37.2	34.9	14.6	68.3	41.3
SME	35.1	53.7	19.0	40.3	32.7	14.9	61.6	43.3
TransE	43.7	65.7	18.2	47.2	43.7	19.7	66.7	50.0
TransH	66.8	87.6	28.7	64.5	65.5	39.8	83.3	67.2
PTransE	91.0	92.8	60.9	83.8	91.2	74.0	88.9	86.4
DPTransE	92.1	94.2	63.5	85.2	92.5	76.3	90.1	88.5
DPTransH	91.7	93.7	59.6	88.5	92.6	72.4	89.3	89.9

V. CONCLUSION AND FUTURE WORK

In this paper, we propose two models, which are called DPTransE and DPTransH, they both combine the relationship path and the entity description text. The DPTransE model makes good use of the relationship between the information text and the entity in the knowledge graph. DPTransH further extends the ability of the model to handle many-to-one and one-to-many complex relationships on DPTransE. We evaluated our model with tasks such as knowledge graph completion and entity prediction. The experimental results show that our model have achieved good performance in dealing with problems in these scenes. The result indicate that we can use the relationship between the entity path and the entity's description text to carry out the task of knowledge graph completion more effectively.

In future work, we will apply the relationship path and entity description text to TransR, TransD and other models, on the other hand, we will also explore how to achieve tasks of knowledge graph representation learning and knowledge graph completion by combining more effective information in the knowledge graph.

ACKNOWLEDGEMENT

The completion of this paper is inseparable from the help of teachers and other friends, the discussion with them is indispensable and thanks for their help.

REFERENCES

[1] G. A. Miller, "WordNet: A lexical database for the English language," *Contemporary Review*, vol. 241, no. 1, pp. 206-208, 2002.

[2] K. Bollacker, "Freebase: A collaboratively created graph database for structuring human knowledge," *SIGMOD 2008*, pp. 1247-1250, 2008.

[3] J. Lehmann, "DBpedia: A large-scale, multilingual knowledge base extracted from Wikipedia," *Semantic Web 6.2*, pp. 167-195, 2015.

[4] A. S, *Introducing the Knowledge Graph*, America: Official Blog of Google, 2012.

[5] A. Bordes, J. Weston, and N. Usunier, "Open question answering with weakly supervised embedding models," in *Proc. European Conf. on Machine Learning and Knowledge Discovery in Databases Springer-Verlag New York, Inc.*, 2014, pp. 165-180.

[6] H. Tanev, *et al.*, "A WordNet-based approach to named entities recognition," in *Proc. the Workshop on Building and Using Semantic Networks Association for Computational Linguistics*, 2002, pp. 1-7.

[7] R. Hoffmann, "Knowledge-based weak supervision for information extraction of overlapping relations," in *Proc. the Association for Computational Linguistics: Human Language Technologies Association for Computational Linguistics*, 2011, pp. 541-550.

[8] A. Bordes, "Translating embeddings for modeling multi-relational data," in *Proc. International Neural Information Processing Systems Curran Associates Inc.*, 2013, pp. 2787-2795.

[9] J. L. Feng, *Knowledge Graph Embedding by Translating on Hyperplanes, AAAI 2014*.

[10] Y. K. Lin, "Learning entity and relation embeddings for knowledge graph completion," in *Proc. the Twenty-Ninth AAAI Conf. on Artificial Intelligence AAAI Press*, 2015, pp. 2181-2187.

[11] G. L. Ji, *et al.*, "Knowledge graph embedding via dynamic mapping matrix," in *Proc. the Association for Computational Linguistics and the International Joint Proceedings of Natural Language Processing*, 2015, pp. 687-696.

[12] N. Lao and W. W. Cohen, *Relational Retrieval Using A Combination of Path-Constrained Random Walks*, Kluwer Academic Publishers, 2010.

[13] N. Lao, T. Mitchell, and W. W. Cohen, "Random walk inference and learning in a large scale knowledge base," in *Proc. Empirical Methods in Natural Language Processing (EMNLP 2011), A Meeting of Sigdat, A Special Interest Group of the ACL DBLP*, UK: John Mcintyre Conference Centre, Edinburgh, 2011, pp. 529-539.

[14] Q. Wang, "Knowledge base completion via coupled path ranking," in *Proc. the Association for Computational Linguistics*, 2016, pp. 1308-1318.

[15] A. Bordes, "Learning structured embeddings of knowledge bases," in *Proc. Artificial Intelligence*, AAAI 2011, San Francisco, California, USA: August DBLP, 2011.
[16] A. Bordes, "A semantic matching energy function for learning with multi-relational data," *Machine Learning*, vol. 94, no. 2, pp. 233-259, 2014.
[17] Y. K. Lin, "Modeling relation paths for representation learning of knowledge bases," *Computer Science*, 2015.
[18] T. Mikolov, "Efficient estimation of word representations in vector space," *Computer Science*, 2013.
[19] Q. V. Le and T. Mikolov, *Distributed Representations of Sentences and Documents*, vol. 4, 2014.

Qi Luo was borned in China. He earned his four-year bachelor degree of science and technology in North East University at Qinhuangdao (NEUQ), Qinhuangdao, China. Currently, he works as a postgraduate of Shandong University (SDU).

Xingwei Hao earned his doctor degree in Shandong University. He is a professor of Shandong University.