



Modeling path information for knowledge graph completion

Ying Shen¹ · Dagang Li² · Du Nan¹

Received: 1 February 2021 / Accepted: 26 August 2021 / Published online: 18 September 2021
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2021

Abstract

Knowledge graphs (KGs) store real-world information in the form of graphs consisting of relationships between entities and have been widely used in the Semantic Web community since it is readable by machines. However, most KGs are known to be very incomplete. The issues of structure sparseness and noise paths in large-scale KGs create a substantial barrier to representation learning. In this paper, we propose an Attribute-embodied neural Relation Path Prediction (ARPP) model to predict missing relations between entities in a KG. The ARPP framework leverages both structural information and textual information from the KG to enrich the representation learning and aid in learning more valuable information from noise paths for relation prediction. To handle the overlooked equal path weight distribution issue which hinders the performance of KG completion, our method evaluates the information propagation for the path by mining neighboring nodes. In order to verify the benefits of incorporating structural information and textual information and the effectiveness of path weight re-distribution, we conduct experiments from various aspects to evaluate the quantitative results for link prediction and entity prediction task, the accuracy change caused by the ablation studies, the effectiveness of the entity attribute and entity/sequence attention, the applicability of the proposed method on Knowledge Graph Completion task, and case study. Results demonstrate that the ARPP model significantly outperforms the state-of-the-art methods.

Keywords Knowledge graph completion · Relation prediction · Representation learning · Path denoising · Information propagation

1 Introduction

In this paper, we introduce the knowledge graph completion problem, which aims to alleviate the incompleteness and sparseness issues caused by collaboratively or semi-automatically knowledge graph (KG) construction.

The major challenge of the knowledge graph completion task is how to predict missing or likely connections between entities based on existing facts in a given KG.

KGs are a type of semantic network that depicts how concepts are related to one another and illustrates how they are interconnected [23]. In the literature, missing relation prediction in the KG has been the subject of significant effort from researchers in the last decade [1, 4, 33]. Despite the effectiveness of previous studies, relation prediction still suffers from the following issues:

- (1) There is a problem of performing learning and inference on a large-scale KG with *incomplete knowledge coverage and structure sparseness*. The node degree, which is related to the betweenness centrality in the network, differs from entity to entity in a KG [5]. Compared with the entities of large degree, the entities of small degree are trained with less information, resulting in an inaccurate entity presentation [12].
- (2) Limited performance caused by *noise paths*. For example, to identify the relationship between a disease and a symptom, the path “disease→alias→disease→symptom” is more effective than

✉ Ying Shen
sheny76@mail.sysu.edu.cn

Dagang Li
dgli@must.edu.mo

Du Nan
nan@dawnlight.com

¹ School of Intelligent Systems Engineering, Sun Yat-Sen University, Guangzhou, China

² Next Generation Internet Research Institute, Macau University of Science and Technology, Macau, China

the path “disease→suitable for eating→food←not suitable for eating←symptoms.” Most existing work assumes that KG paths have equal weights, but in reality, there is a gap in weight distribution between different relationships and paths [20]. It is also time-consuming to adopt traditional depth-first search algorithms to explore a large number of paths in the KG, especially considering that some paths are noise paths, which cannot provide useful information [22].

In this paper, we present ARPP, an Attribute-embodied neural Relation Path Prediction model for relation prediction. Specifically, we first propose an entity/relation representation model that leverages the structure information and textual information to alleviate the limitation of structure sparseness. The learned structural and textual representations make the KG essentially computable and have proved to be helpful for relation prediction. Afterward, we perform the information propagation for the path by mining the linear combination of neighboring nodes to effectively alleviate the impact of noisy and redundant paths on the relation prediction task. We validate the effectiveness of ARPP on a public dataset.

The main contributions of this paper are summarized as follows:

- (1) We expand the semantic structure of the KG by proposing a text-enhanced knowledge representation method considering both structural information and textual contexts in deep neural networks, thereby overcoming the barriers of KG sparseness.
- (2) To alleviate the impact of the noise paths and unbalanced relations on the relation prediction task, we consider the information propagation between path nodes to distinguish the subtle differences between paths and then emphasize those paths with rich information.
- (3) Experiments are conducted on a large benchmark dataset and indicate that our model is superior to the state-of-the-art methods on path prediction tasks.

We organize the rest of our article as follows. Section 2 reviews related works of knowledge graph completion methods. Section 3 describes our Attribute-embodied neural Relation Path Prediction model (ARPP). Section 4 provides experimental results and several model analyses. Finally, in Section 5, we make a conclusion of our work and point out several promising directions for future researches.

2 Related work

KGs typically suffer from missing relations [34], which give rise to the task of automatic knowledge base completion that requiring predicting whether a given triple is valid or not.

Initial KG completion methods mainly drew upon rule-based reasoning approaches [2] and statistical methods [6]. These methods use a first-order relational learning algorithm [17, 18] to learn the probability rules and substitute a specific entity to instantiate the rule to infer a new relation instance from other relation instances that have been learned.

Afterward, distributed representation-driven KG completion methods [23] attempt to learn fact triples for the vector representation of the KG and conduct the inference or prediction in a low-dimensional vector space. Current knowledge representation methods can be carried out by semi-structured data exploration and textual information extraction.

The semi-structured data exploration method has been introduced to automatically create or augment KGs with facts extracted from Wikipedia, which has led to the high accuracy and trustworthiness of facts in its automatically created KGs, including YAGO and DBpedia. Although semi-structured texts are informative, they cover only a fraction of the actual information expressed in the articles and thus cannot meet the demand of completeness in real-world applications.

Textual information extraction methods, which attempt to extract facts from the natural language text, can be grouped into two four approaches, i.e., (i) compositional, (ii) translational, (iii) neural network-based, and (iv) graph-based models.

Several studies have attempted to use graph structure learning methods, e.g., the Path Ranking Algorithm (PRA) [10], and predictive models with multi-task learning [30], to reason over discrete entities and relationships in KGs. However, current structure models do not take full advantage of the contextual information contained in KGs to find more relevant and predictive paths.

Inspired by the success of word embedding, translation-based embedding models attempt to interpret the relations among entities as translations from head entity to tail entity. Translation-based methods have become increasingly popular for their simplicity and effectiveness. Nevertheless, the heterogeneity and imbalance issues in KGs are ignored by previous translation models such as TransE [1], TransH [8], TransR [9], and RotatE [25].

Many studies have attempted to adopt neural network models to facilitate reasoning over relationships between two entities. Neelakantan et al. [15] construct

compositional vector representations for the paths and adopt Recursive Neural Networks to perform inference in the vector space. Shen et al. [19] perform multi-step path reasoning in the learned embedding space through shared memory and a controller. Socher et al. [24] introduce a semantic relationship prediction model based on a matrix-vector recursive neural network (RNN) to learn the compositional vector embeddings of phrases and sentences. Nathani et al. [14] propose a novel feature embedding based on the graph attention network (GAT) to capture both entity and relation features in a multi-hop neighborhood of a given entity. Jagvaral et al. [7] combine bidirectional long short-term memory (BiLSTM) and convolutional neural network (CNN) with an attention mechanism to perform path reasoning for link prediction tasks.

Graph-based models learn more expressive embeddings due to their parameter efficiency and consideration of complex relations. According to the transfer hypothesis, a score function is designed to measure the probability of a relation path connected by multiple triples [34]. Despite obtaining high precision and recall, these methods struggle to explore knowledge from the entities and relationships with small degrees in a KG [31].

In this study, we will explore the textual knowledge contained in the KG and show how to alleviate the impact of noise paths and unbalanced relations in such automatically extracted facts by considering the information propagation to better understand information differences.

3 Methodology

In this paper, we propose an Attribute-embodied neural Relation Path Prediction model (ARPP), a novel framework that leverages textual structure information to tackle relation prediction (see Fig. 1). First of all, we learn the textual embeddings of the entity attributes via a Bidirectional Long Short-Term Memory network (Bi-LSTM), and we learn the structural embeddings of the entity and relation included in the KG via TransE. Meanwhile, we project both entity embeddings and relation embeddings into the same representation space to generate multi-hop paths. Afterward, for the path denoising, we perform information propagation through a Graph Attention Network (GAT) network, which captures the semantic correlation between neighboring path nodes, and computes their linear combination using the normalized attention coefficients. Finally, we feed the path representation into a fully connected network to obtain the final relation prediction results. Next, we would elaborate the framework in detail.

3.1 Entity and relation representation

3.1.1 Structural embeddings of node and edge

Given a training set T of tuples (h, r, t) composed of two entity nodes $h, t \in E$ and a relationship edge $r \in R$, we transform each KG node and edge to a real-valued representation using the TransE model with the Freebase 15K database. The structures of the node and edge can be obtained from pre-trained word embedding matrix and are embedded as $v_{es} \in R^{dl}$ and $v_{rs} \in R^{dl}$, respectively, where dl is the vector dimension set by TransE.

3.1.2 Textual embeddings of node description

Each node has various attributes. Since the content of the node “description” attribute in KGs is much complete than that of other attributes, we encode the node “description” attribute as a textual representation:

The textual embedding consists of two parts, i.e., a Bi-LSTM network and a self-attention mechanism. The self-attention mechanism provides a set of summation weight vectors which are dotted with the hidden LSTM states. The resulting weighted hidden LSTM states are regarded as an embedding for the sequence.

“Description” Sequence Representation. We adopt TransE as the knowledge embedding method to generate initial embedding for the “Description” sequence. The input of the GloVe is the word sequence $W_l = [w_1, w_2, \dots, w_m]$, and the output is the word vector sequence $V_l = [v_1, v_2, \dots, v_m]$, where m is the fixed length of the word sequence, $v_i \in R^{d_w}$ is the word vector of d_w dimension, $i \in [1, m]$ denotes the index of the word sequence.

We adopt a Bi-LSTM to capture the “description” information. The output hw_t at time step t is computed by combining the output of two sub-networks for the past contexts ($V_l = [v_1, v_2, \dots, v_m]$) and future contexts ($V_{l_{reverse}} = [v_m, v_{m-1}, \dots, v_1]$), which is given by Eq. (1):

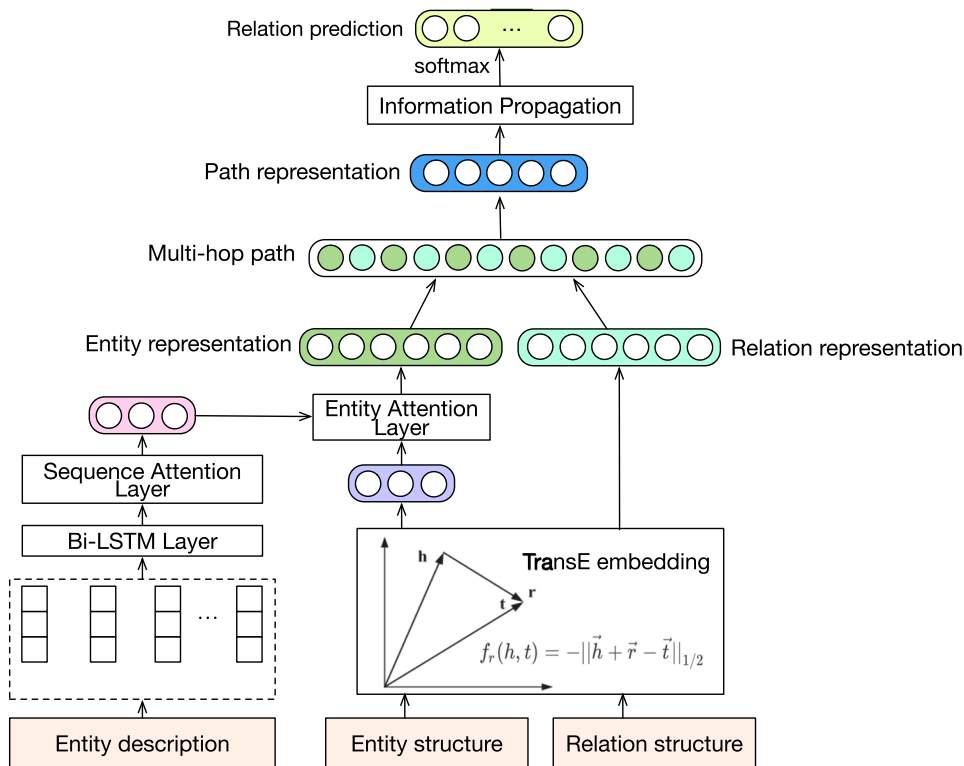
$$hw_t = hf_t + hb_t. \quad (1)$$

The vector sequence $Hw_l = [hw_1, hw_2, \dots, hw_m]$ refers to the output of the Bi-LSTM model at all time steps.

Sequence Reweighted by Attention. We apply a self-attention mechanism [13] to summarize the attention values and pay due attention to each word in sequence according to their interaction. The input of attention mechanism is all hidden LSTM states Hw , while the output is a vector of weights $\alpha_w \in R^m$, which is given by Eq. (3):

$$M_w = \tanh(W_{sw}H_w), \quad (2)$$

Fig. 1 Overview of ARPP model. Dark green, light green, and blue matrices denote entity representation, relation representation, and path representation, respectively



$$\alpha_w = \text{softmax}(w_w^T M_w), \tag{3}$$

where $M_w \in R^{d \times m}$ is the nonlinear mapping function, $W_{sw} \in R^{d \times d}$ and $w_w \in R^d$ are projection parameters. Accordingly, the textual embeddings of entity description v_{et} can be calculated by Eq. (4):

$$v_{et} = H_w \alpha_w^T. \tag{4}$$

3.1.3 Unifying encoding

Given the structural embeddings of the entity and relation as well as the textual embeddings of the entity description, we attempt to learn the node and edge representations and map them into the same representation space.

KG node representation using self-attention mechanism
The projection matrices, M_s and M_t , are adopted to project the node structural embeddings v_{es} and “description” textual embeddings v_{et} , respectively, into the same representation space, which are given by Eq. (5):

$$v_{etm} = M_t v_{et}, \quad v_{esm} = M_s v_{es}. \tag{5}$$

We adjust the attention weight of the node structural embeddings and textual embeddings; then, the node representation $v_e \in R^d$ is computed by merging both embeddings, which is given by Eq. (8):

$$M_e = \tanh(W_e [v_{etm}, v_{esm}]), \tag{6}$$

$$\alpha_e = \text{softmax}(w_e^T M_e), \tag{7}$$

$$v_e = v_{etm} \alpha_e^1 + v_{esm} \alpha_e^2, \tag{8}$$

where $W_e \in R^{d \times d}$ and $w_e \in R^d$ are projection parameters to be learned, $M_e \in R^{d \times 2}$ is the nonlinear mapping function, and α_e^1 and α_e^2 are the attention weights for the structural embeddings and textual embeddings, respectively.

KG Edge Representation. The edge representation is given by $v_r = M_s v_{rs}$. The edge representation is in the same vector space as the node representation because the structural embeddings of both are processed by TransE.

3.2 Path representation

The path representation $g_i = \{u_1, u_2, \dots, u_l\}$ is a multi-hop sequence, where u_{2i} and u_{2i+1} are the edge embedding v_r and the node embedding v_e , respectively.

Information Propagation. We adopt a variant of Graph Attention Network (GAT) [29] to propagate information among path nodes. The hidden states of input nodes are denoted as $g_i \in R^{2d \times 1}$, $i \in \{1, \dots, (m+n)\}$. A Multi-Layer Perceptron (MLP) is used to evaluate attention coefficients between a node i and its neighboring nodes j ($j \in \mathcal{N}_i$) at layer t as given by Eq. (9):

$$p_{ij}^{(t)} = W_a^{(t)}(\text{ReLU}(W_b^{(t)}[g_i^{(t-1)} \oplus g_j^{(t-1)} \oplus e_{ij}])), \tag{9}$$

where W_a^t and W_b^t are trainable parameters of the t -th layer, \oplus indicates the concatenation operation, and \mathcal{N}_i indicates the set of neighbors of node i . Afterwards, we apply the softmax function to normalize the coefficients as shown in Eq. (10):

$$\alpha_{ij}^{(t)} = \text{softmax}_j(p_{ij}^{(t)}) = \frac{\exp(p_{ij}^{(t)})}{\sum_{k \in \mathcal{N}_i} \exp(p_{ik}^{(t)})}. \tag{10}$$

Finally, a linear combination of the neighboring nodes is evaluated based on the normalized attention coefficients. The updated vector for node i at the t -th layer is formulated as in Eq. (11):

$$g_i^{(t)} = \sum_{j \in \mathcal{N}_i} \alpha_{ij}^{(t)} g_j^{(t-1)}. \tag{11}$$

Inspired by Transformer [28], we further apply a position-wise feed-forward (FFN) layer after each GAT layer. The path representation after propagation is represented as $g_i = g_i^{(2)}$.

3.3 Relation prediction

The path representation g_i from the Aggregation Graph is fed into a fully connected network to obtain the final relation prediction results as given by Eq. (12):

$$\tilde{y}_i = \text{softmax}(W_c g_i + b_c), \tag{12}$$

where W_c and b_c are trainable parameters of the predictor. The whole model is trained end-to-end by minimizing the cross-entropy loss, which is given by Eq. (13):

$$L = \text{CrossEntropy}(y, \tilde{y}). \tag{13}$$

4 Results and discussion

4.1 Experimental setup

We test our algorithm on a subset of Freebase 15K-237, whose paths were labeled in 2016 by Neelakantan’s group [15]. The dataset consists of triples (e_1, r, e_2) and the paths that connecting entity pairs (e_1, e_2) in the knowledge graph. Table 1 provides statistics of the dataset used.

We use the mean average precision (MAP) and mean rank (MR) as evaluation metrics to analyze the incorporation of structural and textual information and the effectiveness of path weight re-distribution, respectively.

For the word embeddings, we set the dimension to be 50 and pre-train them by GloVe¹. We feed initial word

representation into a Bi-LSTM with 230 hidden units. The batch size is set to be 32. The dropout rate and learning rate are set to 0.5 and 0.001, respectively. All other model parameters are randomly initialized from $[-0.1, 0.1]$. The parameters are regularized with a L_2 regularization strength of 0.0001. We use the depth-first search algorithm for path searching, setting the value of the path number to be 100 and the minimum and maximum lengths of the path to be 2 and 6, respectively. The threshold for path searching is set to 0.5. The score indicates the informative degree of the path pattern; the higher, the more important. For the GRU model implementation, we compute the cross-entropy as the loss function and use the ReLU activation function and AdaGrad optimizer. The dropout is adopted in the output layer to prevent overfitting.

For the baseline models, we use the implementation of the original paper.

4.2 Experimental results

We compare the performance of ARPP with the state-of-the-art methods, including the supervised relation extraction methods, translation-based methods, path-based methods, structural-based methods, and hybrid methods that utilize structural and textual information. Table 3 summarizes a comparison, introducing state-of-the-art methods in terms of the aim and approach.

To verify the effectiveness of various components of our model, we conduct the ablation test by removing the textual embeddings of “description” attributes (w/o textual description) and attention mechanism. The ablation test of the attention mechanism includes discarding the attention concerning the description sequence representation (w/o sequence attention) and entity representation (w/o entity attention). In addition, we process path representation through GRU rather than through path information propagation, which denotes (w/o path information propagation). The results of link prediction and entity prediction on the FB15K dataset are reported in Table 4. We observe the following:

- (1) ARPP is superior to the state-of-the-art results on FB15K. The reason may be that, compared with a model that only uses structural information or text information in the path, the path information obtained through information propagation has richer semantic knowledge. Therein demonstrating that the simultaneous consideration of structural information and textual description can effectively alleviate the training ineffectiveness on entities of low degree. In addition, unlike the knowledge representation model

¹ <http://nlp.stanford.edu/data/glove.6B.zip>.

Table 1 Notation list

Symbol	Description
(h, r, t)	Tuples containing the head and tail entity nodes h, t and the relationship edge r
v_{es}	Word embeddings of the node structure by TransE
v_{rs}	Word embeddings of the edge structure by TransE
W_l	Sequence embeddings of the entity description by TransE
V_l	Sequence vectors of the entity description by GloVE with input of W_l
Hw_l	Sequence vectors of the entity description by the Bi-LSTM model with inputs of V_l and its reversed vector
v_{et}	Textual embeddings of the entity description calculated by a self-attention mechanism
v_{esm}	Projected node representations from v_{es}
v_{etm}	Projected “description” textual representations from v_{et}
v_e	Reweighted node representations by merging v_{esm} and v_{etm} with a self-attention mechanism
v_r	Projected edge representations from v_{rs}
g_i	Path representation

Table 2 Statistics of our dataset

Statistical Information	Total Number/Length
Number of relations	27,791
Number of relation types (test set)	46
Number of entities	1.59M
Number of entities with “description” attribute	1.12M
Number of entity pairs	3.22M
Number of entity pairs (test set)	2M
Average length of paths	4.7
Minimum length of paths	1
Maximum length of paths	16
Number of paths	191M

that uses a single triple, the path information takes advantage of all the triple information in the path, which greatly improves the relation prediction task.

- (2) Our proposed model outperforms the TAPR model [21] which also incorporates structural information and textual description for knowledge graph completion tasks but does not conduct the path denoising with information propagation, indicating the advantage of combining the structural, textural, and path information included in the knowledge graph.
- (3) Generally, all factors contribute, and this results in a larger performance boost for relation prediction. The basic ARPP model (w/o textual description) cannot perform as well as the ARPP model, which shows that, for low-frequency entities, the description provides supplementary information for embedding; thus, the issue of sparsity in a knowledge base can be addressed properly. It is within our expectation that the adopted attention mechanisms significantly reduce noise and enhance the representation learning of description sequences and entities.

4.3 Model analysis

4.3.1 Utilization of attributes

To further analyze the performance of our model with respect to the entity attribute, we report the MAP results of ARPP and its basic model (w/o textual description) in Fig. 2. We can observe that in the MAP accuracy interval [0.6, 0.9], the ARPP achieves a significant improvement, indicating that textual information can largely alleviate the issue of sparsity when the structural information can provide certain but insufficient support.

4.3.2 Utilization of sequence attention

Due to space limitations, we randomly choose one relation, “_soccer_football_player_position_s,” from Freebase 15K to analyze the attention mechanism concerning the sequence representation.

Substantial amounts of positional information can aid in identifying the football player’s position, including

Table 3 Overall comparison between the state-of-the-art methods

Category	Method	Approach
Supervised relation extraction	NLFeat [26]	An observed feature model that utilizes text-augmented features
Translation-based	TransE [1]	An embedding method that represents the Euclidean distance between the head entity and the tail entity in low-dimensional continuous space
	TransH [8]	An embedding method that maps entities to hyperplanes corresponding to specific relationships
	TransR [9]	An embedding method that maps different relationships to different semantic spaces
	STransE [16]	An embedding method that represents each entity as a low-dimensional vector and each relation by two matrices and a translation vector
	PTransE [12]	An embedding method that considers relation paths as translations between entities and designs an algorithm to measure the reliability of relation paths
Path-based	PRA [35]	An algorithm that provides random paths in a graph
	RWR [11]	A general-purpose graph proximity measure aiming to solve the feature sparsity in link prediction
Structural-based	RNN-PATH [15]	An approach that composes the implications of a path using a recurrent neural network that takes as inputs vector embeddings of the binary relation in the path
	RNN chains of reasoning [4]	A model that combines the rich multi-step inference of symbolic logical reasoning with the generalization capabilities of neural networks
Hybrid methods that utilize structural and textual information	ALL-PATHS+NODES [27]	A compositional learning method of relation types and intermediate nodes for relation prediction on knowledge bases and text
	TEKE [32]	A novel KG representation learning method that takes advantage of the rich context information in a text corpus
	MINERVA [3]	A reinforcement learning model aiming to navigate a graph conditioned on the input query to reasoning over paths
	TAPR [21]	A Type-aware Attentive Path Reasoning framework to complete the knowledge graph by simultaneously considering KG structural information, textual information, and type information

forward, midfielder, defender, and goalkeeper. We classify the positive samples into two groups, i.e., the “position group,” involving the entities with positional information, and the “w/o position group,” containing the remaining entities.

Table 5 shows the accuracy and mean rank of different models for relation prediction. The accuracy increases when the mean rank decreases. Accordingly, we can observe that ARPP demonstrates robust superiority over its basic models, which are developed without sequence attention or positional information.

To predict the relation between “Marko Arnautović” and “Forward,” we visualize the attention scores assigned by ARPP in Fig. 3. The color depth indicates the importance degree of the words, darker representing greater importance. As one may expect, ARPP pays significantly more attention to those words that are contextually related to the pairwise entity, such as “as a forward for,” “Austrian footballer,” and “in race,” which verifies the effectiveness of path weight re-distribution and denoising.

4.3.3 Utilization of entity attention

The adopted attention mechanism can improve the entity representation by merging the structural information and attribute information according to their distributions. The abscissa in Fig. 4 is the proportion of entities with a degree less than or equal to 10 in the dataset, and the ordinate is the MAP.

We can see that when the entity training is sufficient, the entity attention cannot significantly improve the MAP. The increase in the MAP is more obvious when there are a large number of untrained entities in the dataset. The experimental results indicate that the entity attention can increase the entity degree and facilitate a better entity training.

4.4 Knowledge graph completion

The current KGs contain a large number of potential relations that have not been explored. In this section, we elaborate on several actual prediction results and show examples that highlight the strengths of the ARPP model.

Table 4 Result on FB15K Dataset (with ablation study)

Model	Link Prediction %MAP	Entity Prediction %Mean Rank	Entity Prediction %Hits@10
NLFeat (2015)	72.15	187	49.41
TransE (2013)	47.10	265	40.76
TransH (2014)	54.42	237	43.80
TransR (2015)	58.73	196	45.28
STransE (2016)	68.45	191	47.33
PTransE (2015)	71.26	184	47.33
PRA (2010)	64.43	192	47.12
RWR (2012)	64.93	191	47.00
RNN-PATH (2015)	68.43	177	46.28
RNN chains of reasoning (2017)	72.22	172	52.76
ALL-PATHS+NODES (2016)	74.57	170	53.33
TEKE (2016)	74.64	169	54.10
MINERVA (2018)	76.41	144	57.85
TAPR (2020)	78.52	136	64.80
ARPP	80.56	134	64.91
w/o textual description	75.21	148	58.16
w/o sequence attention	76.82	143	57.17
w/o entity attention	77.98	143	56.87
w/o path information propagation	77.10	140	56.43

Bold value indicates the best result in a column

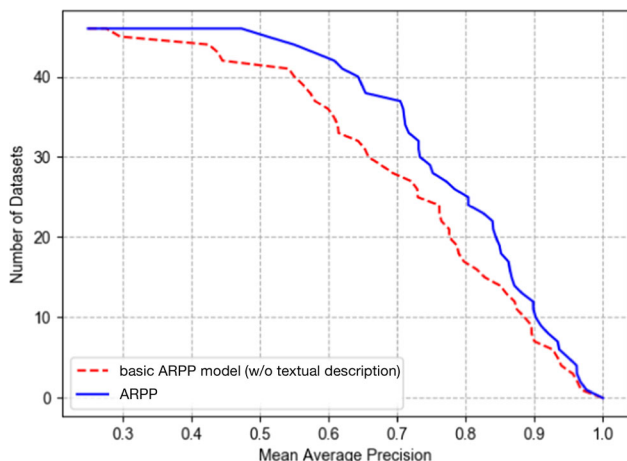


Fig. 2 Effect of textual description. The blue and red lines represent the MAP results for ARPP and its basic model (w/o textual description), respectively. The higher the MAP value is, the more accurate the method

Due to space limitations, we randomly choose one relation, i.e., the “disease-symptom” relationship, to present the performance of ARPP on the KG completion task (see Table 6). Considering the “head” entities (disease) as the starting point, we collect the “head–tail” entity pairs (disease–symptom) that satisfy the path pattern. For different paths within the “disease–syndrome” entity pairs, ARPP utilizes the entity attribute to further verify the authenticity of the specific relationship in an instance, and it selects paths with scores above the threshold (0.70) as candidate paths.

Table 6 shows that ARPP can discover certain diseases that are not directly connected to a certain symptom through an “alias” or “complication” relationship. For example, a birth defect, also known as a congenital disorder, can result in “disabilities that may be physical, intellectual, or developmental. Accordingly, we can associate the disease “birth defect” with the symptom “physical disability.”

Table 5 Comparison of models w/ and w/o sequence attention

	Position		w/o Position	
	Mean Rank (MR)	Accuracy	Mean Rank (MR)	Accuracy
ARPP	2746.08	0.7481	9272.68	0.2032
w/o sequence attention	6103.06	0.3146	8816.93	0.2653

Path {Marko Arnautović, _soccer_football_player_position_s, Forward}
 Marko Arnautović is an Austrian footballer who plays as a forward for Premier League club Stoke City and the Austrian national team.
 West Ham forward Marko Arnautović in race to prove fitness ahead of vital Watford clash.

Fig. 3 An example of the sequence attention visualization when predicting the relation between “Marko Arnautović” and “Forward.” The color depth indicates the importance degree of the words, darker representing greater importance

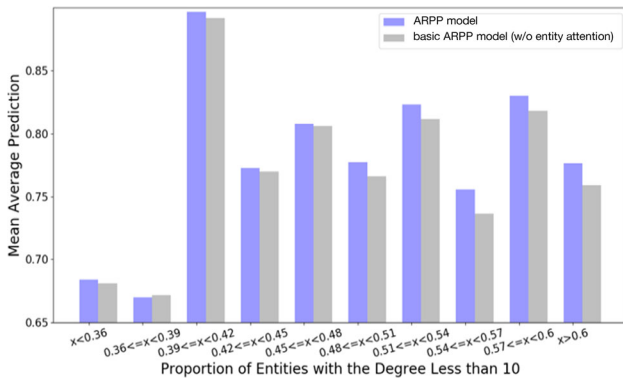


Fig. 4 Effect of entity attention. The abscissa is the proportion of entities with a degree less than or equal to 10 in the dataset, and the ordinate is the MAP. The entity attention can increase the entity degree and facilitate better entity training when there are a large number of untrained entities in the dataset

Another interesting result is the prediction given from the unrelated diseases and symptoms through the intermediate entity. For example, one of the lesions of the disease “stroke” is the “head,” whose related symptom may be “facial paralysis.” On the other hand, two different diseases, “lobular pneumonia” and “acute lung abscess,” that occur in the same body part, “lung,” may have the common symptom “empyema.” We also report noise instances in Table 6. The score of the “stroke→head←eczema” instance is low because the textual information of the “stroke” and “eczema” entities are

quite different, which indicates the benefits of incorporating textual information into predicting paths and ranking instances.

Note that ARPP can identify noise instances that require medical knowledge. Take the “bronchial asthma→lung←lobular pneumonia→shortness of breath” and “bronchial asthma →lung←lobular pneumonia→chills” instances as an example. The “shortness of breath” is the syndrome of the given two diseases, while “chills” is the specific symptom of lobular pneumonia. It is shown that ARPP is able to identify that the latter case is a noisy case.

5 Conclusions and future work

In this paper, we study the task of knowledge graph completion, which has recently attracted increased attention due to their broad applications in natural language processing and natural language generation. We present a novel and general relation prediction framework (ARPP) to predict unseen relationships between entities through reasoning inside a given KG. We greatly expand the KG even without external textual resources. The experimental results on public datasets show that our ARPP method significantly outperforms the state-of-the-art methods on path prediction tasks, which confirms that ARPP is an interpretable model that exploits the differences between paths and demonstrates the impact of each path on relation prediction tasks.

In the future, we will further alleviate the data sparsity problem by exploring the hidden relations beyond the context and other auxiliary data via BERT and transformer. In addition, a multi-view attention mechanism will be designed to interactively learn attentions from different graph components to improve overall representational learning.

Table 6 Sampled “disease–syndrome” path patterns discovered by ARPP

Path	Score	Instance
disease ^{alias} →disease ^{syndrome} →syndrome	0.7018	Birth defect → congenital disorder → physical disability
disease ^{complication} →disease ^{syndrome} →syndrome	0.8360	Diabetes → diabetic nephropathy → frequent urination
disease ^{lesion} →body parts ^{lesion} ←syndrome	0.7173	Stroke→head←facial paralysis
	0.1031	Stroke→head←eczema
disease ^{lesion} →body parts ^{lesion} ←disease ^{syndrome} →syndrome	0.7101	Bronchial asthma → lung ← lobular pneumonia → shortness of breath
	0.3045	Bronchial asthma → lung ← lobular pneumonia → chills

Acknowledgements This work was financially supported by the National Natural Science Foundation of China (No.61602013) and the Shenzhen General Research Project (No. JCYJ20190808182805919).

Declarations

Conflict of interest We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome. We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us.

References

- Bordes A, Usunier N, Garcia-Duran A, Weston J, Yakhnenko O (2013) Translating embeddings for modeling multi-relational data. In: *Advances in neural information processing systems*, pp. 2787–2795
- Bühmann L, Lehmann J (2013) Pattern based knowledge base enrichment. In: *International Semantic Web Conference*. Springer, pp. 33–48
- Das R, Dhuliawala S, Zaheer M, Vilnis L, Durugkar I, Krishnamurthy A, Smola A, McCallum A (2017) Go for a walk and arrive at the answer: reasoning over paths in knowledge bases using reinforcement learning. arXiv preprint [arXiv:1711.05851](https://arxiv.org/abs/1711.05851)
- Das R, Neelakantan A, Belanger D, McCallum A (2016) Chains of reasoning over entities, relations, and text using recurrent neural networks. arXiv preprint [arXiv:1607.01426](https://arxiv.org/abs/1607.01426)
- Deng Y, Li Y, Du N, Fan W, Shen Y, Lei K (2018) When truth discovery meets medical knowledge graph: estimating trustworthiness degree for medical knowledge condition. arXiv preprint [arXiv:1809.10404](https://arxiv.org/abs/1809.10404)
- Gardner M, Mitchell T (2015) Efficient and expressive knowledge base completion using subgraph feature extraction. In: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 1488–1498
- Jagvaral Batsalem, Lee Wan-Kon, Roh Jae-Seung, Kim Min-Sung, Park Young-Tack (2020) Path-based reasoning approach for knowledge graph completion using cnn-bilstm with attention mechanism. *Expert Syst Appl* 142:112960
- Ji G, He S, Xu L, Liu K, Zhao J (2015) Knowledge graph embedding via dynamic mapping matrix. In: *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, vol 1, pp. 687–696
- Guoliang J, Kang L, Shizhu H, Jun Z, Knowledge graph completion with adaptive sparse transfer matrix. In: *AAAI*, pp. 985–991
- Lao Ni, Cohen William W (2010) Relational retrieval using a combination of path constrained random walks. *Mach Learn* 81(1):53–67
- Lao N, Mitchell T, Cohen William W (2011) Random walk inference and learning in a large scale knowledge base. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, pp. 529–539
- Lin Y, Liu Z, Luan H, Sun M, Rao S, Liu S (2015) Modeling relation paths for representation learning of knowledge bases. arXiv preprint [arXiv:1506.00379](https://arxiv.org/abs/1506.00379)
- Lin Z, Feng M, dos Santos Cicero N, Yu M, Xiang B, Zhou B, Bengio Y (2017) A structured self-attentive sentence embedding. arXiv preprint [arXiv:1703.03130](https://arxiv.org/abs/1703.03130)
- Nathani D, Chauhan J, Sharma C, Kaul M (2019) Learning attention-based embeddings for relation prediction in knowledge graphs. In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 4710–4723
- Neelakantan A, Roth B, McCallum A (2015) Compositional vector space models for knowledge base completion. arXiv preprint [arXiv:1504.06662](https://arxiv.org/abs/1504.06662)
- Nguyen D Q, Sirts K, Qu L, Johnson M (2016) Stranse: a novel embedding model of entities and relationships in knowledge bases. arXiv preprint [arXiv:1606.08140](https://arxiv.org/abs/1606.08140)
- Adam P, Geoff S (2007) First order reasoning on a large ontology. *ESARLT* 257
- Pujara J, Miao H, Getoor L, Cohen W (2013) Knowledge graph identification. In: *International Semantic Web Conference*. Springer, pp. 542–557
- Shen Yelong, Huang P-S, Chang M-W, Gao J (2017) Modeling largescale structured relationships with shared memory for knowledge base completion. In: *Proceedings of the 2nd Workshop on Representation Learning for NLP*, pp. 57–68
- Shen Ying, Chen Daoyuan, Tang Buzhou, Yang Min, Lei Kai (2018) Eapb: entropy-aware pathbased metric for ontology quality. *J Biomed Semant* 9(1):20
- Shen Y, Ding N, Zheng H-T, Li Y, Yang M (2020) Modeling relation paths for knowledge graph completion. *IEEE Trans Knowled Data Eng*
- Shi Baoxu, Weninger Tim (2017) Proje: Embedding projection for knowledge graph completion. *AAAI* 17:1236–1242
- Socher R, Chen D, Manning C D, Ng A (2013) Reasoning with neural tensor networks for knowledge base completion. In: *Advances in neural information processing systems*, pp. 926–934
- Socher R, Huval B, Manning C D, Ng A Y (2012) Semantic compositionality through recursive matrixvector spaces. In: *Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning*. Association for Computational Linguistics, pp. 1201–1211
- Sun Z, Deng Z-H, Nie J-Y, Tang J (2019) Rotate: Knowledge graph embedding by relational rotation in complex space. In: *International Conference on Learning Representations*
- Toutanova K, Chen D, Pantel P, Poon H, Choudhury P, Gamon M (2015) Representing text for joint embedding of text and knowledge bases. In: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 1499–1509
- Toutanova K, Lin V, Yih W-T, Poon H, Quirk C (2016) 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)*, vol 1, pp. 1434–1444
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser Ł, Polosukhin I (2017) Attention is all you need. In: *Advances in Neural Information Processing Systems*, pp. 5998–6008
- Veličković P, Cucurull G, Casanova A, Romero A, Lio P, Bengio Y (2018) Graph attention networks. In: *Proceedings of the International Conference on Learning Representations (ICLR)*
- Wang Q, Liu J, Luo Y, Wang B, Lin C-Y (2016) Knowledge base completion via coupled path ranking. In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, vol 1, pp. 1308–1318
- Wang Q, Wang B, Guo L et al (2015) Knowledge base completion using embeddings and rules. In: *IJCAI*, pp. 1859–1866

32. Wang Z, Li J-Z (2016) Text-enhanced representation learning for knowledge graph. In: IJCAI, pp. 1293–1299
33. Wen D, Liu Y, Yuan K, Si S, Shen Y (2017) Attention-aware pathbased relation extraction for medical knowledge graph. In: International Conference on Smart Computing and Communication. Springer, pp. 321–331
34. West R, Gabrilovich E, Murphy K, Sun S, Gupta R, Lin D (2014) Knowledge base completion via search based question answering. In: Proceedings of the 23rd international conference on World wide web. ACM, pp. 515–526
35. Xu J, Chen K, Qiu X, Huang X (2016) Knowledge graph representation with jointly structural and textual encoding. arXiv preprint [arXiv:1611.08661](https://arxiv.org/abs/1611.08661)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.