Modeling Relation Paths for Knowledge Graph Completion

Ying Shen, Ning Ding, Hai-Tao Zheng, Yaliang Li, Min Yang*, *Member, IEEE*

Abstract— Knowledge graphs (KG) often encounter knowledge incompleteness. The path reasoning that predicts the unknown path relation between pairwise entities based on existing facts is one of the most promising approaches to the knowledge graph completion. However, most conventional path reasoning methods exclusively consider the entity description included in fact triples, ignoring both the type information of entities and the interaction between different semantic representations. In this study, we propose a novel method, Type-aware Attentive Path Reasoning (TAPR), to complete the knowledge graph by simultaneously considering KG structural information, textual information, and type information. More specifically, we first leverage types to enrich the representational learning of entities and relationships. Next, we describe a type-level attention to select the most relevant type of given entity in a specific triple without any predefined rules or patterns to reduce the impact of noisy types. After learning the distributed representation of all paths, path-level attention assigns different weights to paths, from which relations among entity pairs are calculated. We conduct a series of experiments on a real-world dataset to demonstrate the effectiveness of TAPR. Experimental results show that our method significantly outperforms all baselines on link prediction and entity prediction tasks.

Index Terms— Big data applications, Semantic Web, Data mining, Knowledge discovery

1 INTRODUCTION

A knowledge graph (KG) [1] is essentially a semantic network composed of entities and the relationship between entities, mostly in the form of triples, e.g., (Despicalble me, character_produced_by, Chris Meledandri), indicating the Chris Meledandri is the producer of the cartoon Despicalble me. Today, large-scale KGs such as Freebase [2], WordNet [3], DBpedia [4], OpenCyc [5], Wikidata [6] and YAGO [7] have had a profound impact on areas such as question answering [8], recommended systems [9], web search [10] and information retrieval [11].

Although typical KGs are large in size, they still encounter the knowledge incompleteness [12]. Take the entity / fact in Freebase as an example. The profession, nationality, spouse, siblings, ethnicity etc. are facts related to the entity "person". According to the statistics, 40% of entities with no facts while 56% of entities with less than 3 facts. The place of birth is missing for 71% of all people included in Freebase, even though this is a mandatory property of the schema [13].

There are two main methods to complete the knowledge graph: relation extraction and path reasoning. Significant progress has been made by relation extraction that explores entity relations from text corpora [14]. However, the supervised relation extraction that requires a large amount of labelled training data is usually labor-intensive and time-consuming, while distantly supervised relation extraction often suffers from the noisy labeling problem.

Many efforts have been devoted to path reasoning that aims to infer novel relational facts based on existing triples and completing missing links between pairwise entities [15]. Different from direct path, multi-hop paths are the paths via multi nodes between the source and destination nodes. When routing between two nodes, a multi-hop path that consists of a sequence of triples can provide more informative but noisy knowledge than a direct path. Through the intermediate node, two triples can be associated on the graph. Take Fig. 1 as an example. When one would like to know who first voiced Agnes on Despicable Me, one can obtain a path "Despicable Me→character→Agnes Gru→cast→Despicable Me 1→actor→Elsie Fisher" from the triples (Despicable Me, character, Agnes Gru), (Agnes Gru, cast, Despicable Me 1) and (Despicable Me 1, actor, Elsie Fisher) via the intermediate node / word conjunction. Afterwards, one can infer and learn that Elsie Fisher is the first actor to voice Agnes.

Despite the usefulness of the multi-hop path, path reasoning in knowledge graphs remains a challenge:

1. Limited performance due to the incomplete knowledge coverage and representation sparseness in the large-scale knowledge graph [16]. Even worse, structure information usually cannot distinguish the different meanings of relations and entities in different triples.
2. Most conventional methods exclusively consider the entity description included in fact triples, ignoring the type information of entities and the interaction between different semantic representations.

- *Y. Shen is with the School of Intelligent Systems Engineering, Sun Yat-Sen University. E-mail: shenying@pkusz.edu.cn.*
- *N. Ding and H. Zheng are with Graduate School at Shenzhen, Tsinghua University, Shenzhen 510100, P.R. China. E-mail: dingn18@mails.tsinghua.edu.cn, zheng.haitao@sz.tsinghua.edu.cn.*
- *Y. Li is with Alibaba Group, Bellevue, WA USA. E-mail: yaliang.li@alibaba-inc.com*
- *M. Yang (Corresponding author) is with the University of the Chinese Academy of Sciences, Shenzhen 510100, P.R. China. E-mail: min.yang@siat.ac.cn.*

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2020.2970044, IEEE Transactions on Knowledge and Data Engineering

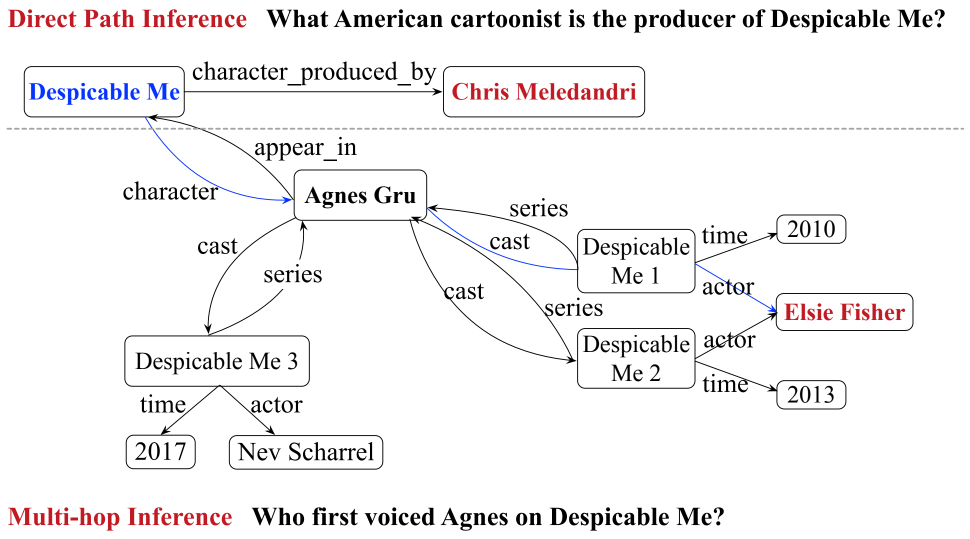


Fig. 1. Example of multi-hop path reasoning in knowledge graph.

3. In addition, the application of type information may bring noise. KGs often provide type information (e.g., Chris Meledandri is a producer, which is a person, which is a living thing) and type constraints (e.g., a person can only marry another person, not a thing). Unlike relation types, entities usually possess multiple types that have various representations in different scenarios [17]. For example, in a triple (Chris Meledandri, produce, Despicable Me) in which "Chris Meledandri" has different types, including cartoonist, writer, producer, animator, and voice actor, one must indicate the most significant roles via the relation and tail entity played in this triple.

To address the above issues, we propose a path reasoning approach to infer relations over multi-hop paths in large KGs. More specifically, we first propose an entity/relation attentive representation model to encode the KG structural and textual information so as to alleviate the limitation of structure sparseness and capture significant semantic information for type selection. Then we incorporate relation types to propose a constrained type attention so as to restrict the semantics of entities and pay less attention to the noisy entity types. Finally, we propose a knowledge-enhanced attention mechanism to reweight the relational paths, effectively reducing the impact of noisy paths.

The contributions of our work are as follows:

- We propose a novel knowledge-enhanced attentive path reasoning model that leverages structural information, textual information and type information to capture rich semantic knowledge and overcome the limitations of representation sparseness.
- We design an attentive learning model that fully captures knowledge graph information on word level, type level and path level, offering more flexibility in the path prediction process.
- The experimental results on the public dataset demonstrate that TAPR has remarkable applicability and robust superiority over competitors by a noticeable margin for link prediction, triple classification, and knowledge graph completion.

2 Related Work

2.1 Relation Path Reasoning

Recently, relation prediction in the graph has become an emerging research field because of the need for high-quality fact and edge predictions related to the relational data representation and modeling [18].

Many works have attempted to use graph structure learning methods, e.g., the Path Ranking Algorithm (PRA) [19], to reason about discrete entities and relationships in knowledge graph. PRA is a method for learning proximity measures on labeled graphs. Proximity is defined by a weighted combination of simple "path experts", each corresponding to following a particular sequence of labeled edges. However, the current structure models exclusively explore attribute similarity through entity clustering rather than consider the textual information pertaining to the entity in the graph [20].

Several studies process relational learning as a multi-relational representation learning problem by encoding both entities and relations in a low-dimensional space using Bayesian clustering [21][22], energy-based models [23], matrix factorization [24], etc. Among the existing models, translation-based methods have achieved state-of-the-art performance by converting entities and relation into vectors and regarding each relation as one translation from head entity to tail entity. Nevertheless, the current translation-based methods, including TransE [25], TransH[26], and TransR [27], assumes that the facts included in the knowledge graph are completely correct. The issues of fact error and conflict prevalent in real-world applications remain to be settled. In addition, all existing representation learning methods of knowledge bases use only direct relations between entities, ignoring rich information in relation paths.

Many studies have attempted to adopt neural network models to decompose a path as the sequence of relations in the path before modelling path representations between two entities. Neelakantan et al. [28] construct compositional vector representations for the paths and adopt Recursive Neural Networks to make an inference in the vector space. Shen et al. [29] perform a multi-step relation inference in an embedding neural space through the shared memory and controller. Socher et al. [30] introduce a relation prediction model based on an expressive neural tensor network that allows the mediated interaction of entity vectors via a tensor. Lin et al. [31] propose a path-based representation learning model by considering relation paths as translations between entities for representation learning, and representing relation paths via semantic composition of relation embeddings. However, in cases in which are various paths between a pairwise entities, these existing studies either select one path or utilize simple score pooling methods such as Top-K, LogSumExp, Average, etc. These simple score pooling methods cannot infer the relational path by considering multiple informative paths collectively according to the specific scenario [32].

2.2 Knowledge Graph Information Learning

Knowledge graph information learning studies methods for graph-structured data. There are two types of research to statistical relational models: the first is based on latent feature models such as tensor factorization [33] and multiway neural networks [34], while the second is based on mining observable patterns in the graph [35].

Information included in the knowledge graph, such as textual information and type information, is considered as supplements for the structured information embedded in triples and is also of great significance for the knowledge representation learning of KG. Wang et al. [36] propose a text-enhanced representation learning for KG by incorporating the textual contexts to each entity for and relation. Cao et al. [37] use low dimensional vectors to represent vertices appearing in a graph and integrate global structural information of the graph into the learning process.

The semantic information contained in the entity type is also useful for path reasoning, and it has only attracted attention recently. Das et al. [38] jointly reason about relations, entities, and entity-types and use neural attention modeling to incorporate multiple paths. Chen et al. [39] resolve the link prediction problem from a variational inference perspective. Xie et al. [40] propose a novel method to take advantages of entity types for the representation learning of knowledge graphs. Jiang et al. [41] propose a path-based relation inference model that learns entity pair representations with attentive path combination. They incorporate entity types by mapping each entity to the averaged representation of its types, which may neglect different attention values between informative types and noisy types and cannot select the most relevant type of a given entity in a specific triple. To the best of our knowledge, TAPR is the first effort to study entity type discrimination in path reasoning for knowledge graph completion.

2.3 Knowledge Graph Completion

KGs provide semantically structured information that is interpretable by computers — a property that is regarded as an important ingredient to build more intelligent machines [42]. Completeness, accuracy, and data quality are important parameters that determine the usefulness of KGs [43]. In addtion, a recent study [44] found that the growth of Wikipedia has been slowing down. Consequently, automatic knowledge base completion methods have been gaining more attention.

Collaborative knowledge graph completion, which was used to build Wikipedia and Freebase, leads to highly accurate results but suffers from its dependence on human experts [45]. Recently, knowledge representation has been used in the Semantic Web community with the purpose of creating a "web of data" that is readable by machines [46]. Current knowledge representation methods can be grouped into two main approaches, i.e., semi-structured data exploration and textual information extraction.

The first approach has been introduced recently to automatically create or augment KGs with facts extracted from Wikipedia, particularly its semi-structured data such as the infoboxes. This approach has led to the high accuracy and trustworthiness of facts in its automatically created KGs including YAGO and DBpedia [47]. 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 application. The second approach attemps to extract facts from the natural language text. For example, Annervaz et al. [48] develop a deep learning model that can extract KG relevant support facts from texts depending on the task using attention mechanism. Cannaviccio et al. [49] quantify the number of highly accurate facts that can be harvested with high precision from the text of Wikipedia articles using information extraction techniques bootstrapped from the entities and relations already in a KG. Example projects in this category include NELL [50] and the Knowledge Vault [10].

In this study, we will explore the textual knowledge contained in the KG and show how we reduce the level of "noise" in such automatically extracted facts by using the attention mechanism by simultaneously considering other KG information.

3 METHODOLOGY

Given a knowledge graph with a set of path sequences $P = \{p_1, p_2, \dots, \dots, p_n\}$, we define the j-th multi-hop path $p_j = \{e_1, r_1, e_2, r_2, \dots, e_i, r_i, \dots, e_m, r_m\}$, where e_i and r_i are the i-th intermediate entity vector and the relation vector, respectively. Our task of path prediction is to infer the unknown path relation between pairwise entities. In this context, we need to select a relationship r from the relationship set $\{r_k\}_{k=1}^{K}$ and give a confidence score for the query relation and the final representation of the path, to complete the missing relationship between the given pairwise entities in the knowledge graph. The overall architecture of the proposed TAPR is shown in Figure 2.

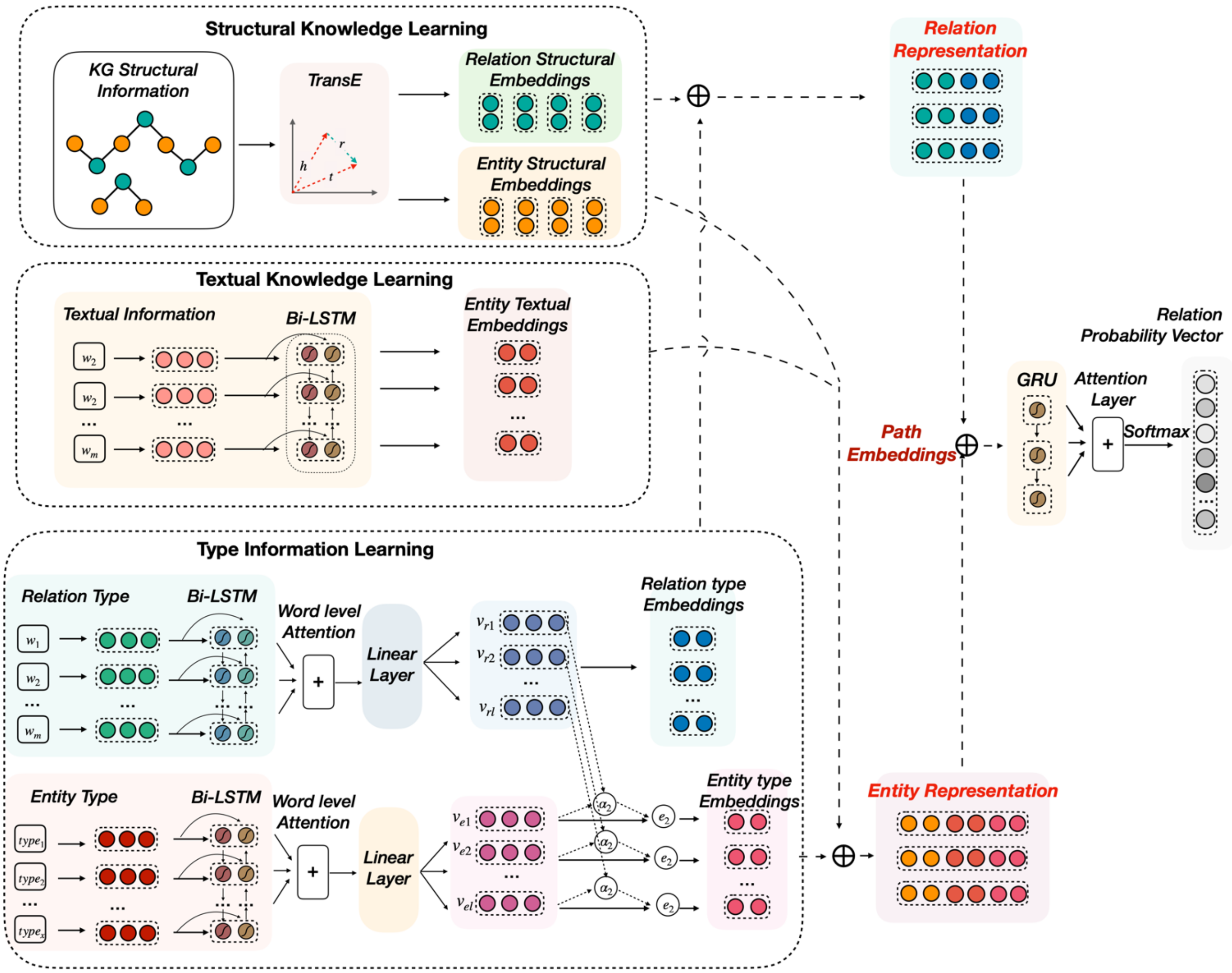


Fig. 2. Type-aware Attentive Path Reasoning framework for Knowledge Graph Completion.

3.1 Structural Knowledge Learning

We pre-train the TransE model with a subset of Freebase to obtain the vector representation of KG entity and relationship. The structures of the entity and relation are encoded as $v_{es} \in R^{dl}$ and $v_{rs} \in R^{dl}$, respectively, where dl is the vector dimension set by TransE.

3.2 Textual Knowledge Learning

Each entity has various attributes. We found that the content of entity "description" is relatively complete compared with other knowledge graph attributes, so we encode the entity "description" as vector representations.

Word representation. We adopt GloVE [1] to pre-train word embeddings. GloVe is an unsupervised learning algorithm for distributed word representation. This is achieved by mapping words into a meaningful space using the global matrix factorization and local context window methods. Given the entity "description" word sequence $W_l = [w_1, w_2, \dots, w_m]$, we obtain its word vector sequence $V_l = [v_1, v_2, \dots, v_m]$, where m is the fixed length of the word sequence and $v_i = R^{dw}$, $i \in [1, m]$ is the word vector of the dw dimension.

Sequence representation. A Bi-directional Long short-term memory (Bi-LSTM) model learns bidirectional long-term dependencies between time steps of time series or sequence data. Specifically, we run a forward LSTM and a backward LSTM over the word representation, generating the forward and backward hidden vector sequences. We input the word vector sequence $V_l = [v_1, v_2, \dots, v_m]$ and the inverse word vector sequence $V_{l\_reverse} = [v_m, v_{m-1}, \dots, v_1]$ into the forward layer and backward layer, respectively, of the Bi-LSTM model. The vector $\mathrm{hw_t}$ at time step t, which combines the vector of forward layer hf_t and backward layer $\mathrm{hb_t}$, can be calculated as follows:

$$hw_t = hf_t + hb_t. \tag{1}$$

The vector sequence $H_w = [hw_1, hw_2, \dots, hw_m]$ refers to the vector concatenation of the Bi-LSTM at all time steps. Finally, we apply the attention mechanism to reweight each word according to its specific role plays when interacting with other words. The sequence representation $\mathrm{v_{ed}}$ is calculated as follows:

[1] https://nlp.stanford.edu/projects/glove/

$$M_s = \tanh(W_{sw}H_w), \tag{2}$$

$$\alpha_d = \mathrm{softmax}(w_s^T M_s), \tag{3}$$

$$v_{ed} = H_w\alpha_d^T, \tag{4}$$

where $W_{sw} \in R^{dl\times dl}$ and $w_s \in R^{dl}$ are projection parameters, $\alpha_d \in R^m$ is the normalized attention, $M_s \in R^{dl\times m}$.

3.3 Type Information Learning

Type Word Embeddings. The type attribute of a relationship or entity is regarded as a sequence of words. Unlike a relation that has a unique type, each entity has various types. Given an entity, we consolidate its type set $\{type_1, type_2, \dots, type_x\}$ into a type-length sequence $s = \{t_{11}, t_{12}, \dots, t_{1j_1}, t_{21}, t_{22}, \dots, t_{2j_2}, \dots, t_{xj_x}\}$, where x is the number of types and j_i ($i \in [1, x]$) is the length of each type.

Given such a type sequence, we look up s in a word embedding matrix $M \in R^{dl\times|V|}$, where $|V|$ denotes the size of the vocabulary. Thus we obtain the sequence of word vectors $V_s = [v_1, v_2, \dots, v_l]$, where l is the fixed length of the word sequence and $v_i \in R^{dl}$, $i \in [1, l]$ is the word vector of dl dimension.

Type Sequence Embedding. Using the same method as shown in Eq(1), we feed the word vector sequence V_s and the inverse word vector sequence $V_{s\_reverse}$ into the forward layer and backward layer, respectively, of the Bi-LSTM network. We then combine the hidden states of the forward layer hf_i and the backward layer hb_i at each time step i using the following element-wise sum:

$$h_i = [hf_i \oplus hb_i]. \tag{5}$$

Word-level Attention. Word-level attention is adopted to capture the informative knowledge of type sequence embedding. The type sequence representation v_{et} of entities is given by

$$M_w = \tanh(H), \tag{6}$$

$$\alpha_t = \mathrm{softmax}(W_w^T M_w), \tag{7}$$

$$v_{et} = H\alpha_t^T, \tag{8}$$

where $H \in R^{l\times m}$ is the type sequence embedding matrix that consists of vectors $\{h_1, h_2, \dots, h_l\}$. W_w^T is the transpose vector, and $W_w \in R^{1\times l}$ is a weight parameter vector. $\alpha_w \in R^{m\times 1}$ is the normalized attention.

In the same way, we obtain the sequence representation v_{rt} of the relation type that is also a m-dimensional vector.

Type level attention. Given a triple, we design a soft entity type constraint in accordance with the relation-specific type:

$$e_i = W_t \tanh(W_e v_{ei} \oplus W_r v_{ri}), \tag{9}$$

where v_{ei} and v_{ri} are the elements of entity representation and relation representation, respectively. W_t, W_r, and $W_e \in R^{l\times l}$ are the weight parameter matrices of type, entity and relation, respectively.

Entity types include informative types and noisy types, where informative types shall be assigned different weights in different scenarios to improve the relation prediction between entities in a given path. Accordingly, we adopt type-level attention to incorporate relation type information into the entity type sequence representation to select the most relevant type of a given entity in a specific triple without any predefined rules or patterns. The type-level attention α_t is given by

$$\alpha_t = \sum_{j=1}^{l} v_{ei} \frac{exp(e_j)}{\sum_{j=1}^{l} exp(e_j)} \tag{10}$$

Afterwards, we multiply each element of v_{et} and its corresponding attention to adjust the entity type encoding v_t with relation constraints:

$$v_t = v_{et} \otimes \alpha_t. \tag{11}$$

3.4 Unifying embeddings

Given a KG, we are to learn the entity and relation representation.

Entity representation. The projection matrices M_s, M_d and M_t are adopted to encode structural embeddings v_{es}, textual embeddings v_{ed} and type embeddings v_t, respectively, into the same representation space:

$$es = M_s v_{es},$$
$$ed = M_d v_{ed}, \tag{12}$$
$$et = M_t v_t.$$

Using the attention mechanism, we adjust the attention weight of structural embeddings, textual embeddings and type embeddings then merge these embeddings to learn the entity representation v_e:

$$M_e = \tanh(W_e[es, ed, et]), \tag{13}$$

$$\alpha_e = \mathrm{softmax}(w_e^T M_e), \tag{14}$$

$$v_e = [es, ed, et]\alpha^T. \tag{15}$$

Relation representation. Entities and relations processed by TransE are in the same vector space. Thus, the structural representation and type representation of relation are given by

$$rs = M_s v_{rs} \qquad rt = M_t v_{rt}. \tag{16}$$

Using the attention mechanism, the relation representation v_r is obtained by

$$M_r = \tanh(W_r[rs, rt]), \tag{17}$$

$$\alpha_r = \mathrm{softmax}(\mathrm{w_r^T M_r}), \tag{18}$$

$$v_r = rs\alpha_r^1 + rt\alpha_r^2. \tag{19}$$

3.5 Path Encoding

We determine the path representation by combining the entity representation $\mathrm{v_e}$ and relation representation $\mathrm{v_r}$ into a multi-hop sequence $p_i = \{v_{e1}, v_{r1}, v_{e2}, v_{r2}, \dots, v_{ei}, v_{ri}, \dots, v_{el}, v_{rl}\}$. We then adopt GRU [51] to encode paths into vector representations.

GRU networks for time step t can be expressed as:

$$z_t = \sigma(W_z x_{t-1} + U_z p_t + b_z), \tag{20}$$

$$r_t = \sigma(W_r x_{t-1} + U_r p_t + b_r), \tag{21}$$

$$h_t = \tanh(W_h(x_{t-1} * r) + U_h p_t + b_h), \tag{22}$$

$$x_t = (1 - z) * h + z * x_{t-1}, \tag{23}$$

where z is the update gate, r is the reset gate, h is the output vector, x_t is plugged into the network unit. σ is the sigmoid function. $U_z, U_r, U_h, W_z, W_r, W_h \in R^{dlxdp}$ are weight matrices, $b_z, b_r, b_h \in R^{dp}$ are offset vectors, * is the hadamard product. The final output vector $\mathrm{h_l}$ is the encoded representation of the path between the given pairwise entities as $p = h_l$.

We form a path matrix P that consists of encoded path $[\mathrm{p_1}, \mathrm{p_2}, \dots \mathrm{p_m}]$ generated by GRU, where $\mathrm{P} \in \mathrm{R^{m\times h}}$ and m is the number of paths between the given entity pair.

Using different types of attention, path encoding can be divided into the following three categories.

One: We randomly choose a path from the path set as a naive baseline of path attention.

Average (Ave): Based on the assumption that each path in the path set has the same impact on the final path matrix representation, we assign the same weight to each path. In other words, the final path matrix representation is equal to the average of each path vector in the path set.

Path Attention (PathAtt): To reduce the influence of noisy paths, path-level attention is designed to assign different weights α_i to paths and obtain the final path matrix representation P_r:

$$M = \tanh(W_p P) \tag{24}$$

$$\alpha_i = \mathrm{softmax}(w^T M) \tag{25}$$

$$P_r = P a_i^T \tag{26}$$

where $\mathrm{M} \in \mathrm{R^{m\times h}}$ is the mapping matrix of path matrix, $\alpha_i \in \mathrm{R^m}$ is the attention model weight, $P_r \in R^m$ is the attentive path representation and $W_p \in R^{m\times m}$, $w \in R^m$ are the projection parameters.

3.6 Output Layer

We use a fully connected layer to convert a path into a k-dimensional vector, where k represents the number of relationship types, including the null relationships. The output of the fully connected layer then goes through a softmax layer to output the probability of different relations. The relation prediction is given by

$$y = \mathrm{argmax}(\mathrm{softmax}(M_o P_r + b)) \tag{27}$$

where $\mathrm{M_o} \in \mathrm{R^{k\times h}}$ is the representation matrix of relation types and $\mathrm{b} \in \mathrm{R^k}$ is a bias vector. TAPR is trained to minimize the negative log-likelihood:

$$\begin{aligned} L(\theta, \Delta_R^+, \Delta_R^-) = -\frac{1}{M} \sum_{e_h, e_t, r \in \Delta_R^+} & logP(r|e_h, e_t) \\ + \sum_{\hat{e}_h, \hat{e}_t, \hat{r} \in \Delta_R^-} & \log(1 - P(\hat{r}|\hat{e}_h, \hat{e}_t)) \end{aligned} \tag{28}$$

where θ presents all parameters of our model, M is the total number of training samples. For all relationships, Δ_R^+ indicates the positive sample, i.e., the existing fact set in the KG, while Δ_R^- denotes the negative sample (unknown fact set).

4 Experiments and Results

In this section, we introduce the dataset and metrics we used in relation paths modeling. Besides, we present the implementation details of our method and a series of state-of-the-art link prediction and entity prediction for comparisons.

4.1 Experiment setup

Dataset and Metrics: In our study, we adopt a subset of Freebase [2] as our dataset. The dataset contains 1,594,253 entities and 27,791 relationships. Among them, 46 types of relationships between more than 2 million entity pairs are regarded as the test set, and 23,599 textual relationship types extracted from Freebase constitutes the intermediate relation nodes in paths. There are 191M paths in total. The average path length is 4.7 and the maximum path length is 7. The mean average precision (MAP), Mean Rank and Hits@10 are adopted as our evaluation metrics.

Implementation detail. Pre-trained GloVE [3] embeddings of 50 dimensions are used as word embeddings. The output vector dimension of TransE is set at 50.

For the Bi-LSTM model, the hidden layer size and the final hidden layer size are both set to 230. The learning rate and the dropout rate are set to 0.001 and 0.5, respectively. We train our models in batches with a size of 128. All other parameters are randomly initialized from [-0.1,0.1]. The model parameters are regularized with a L2 regularization strength of 0.0001. The maximum length of sentence is set to 60. The path length lower limit and upper limit are set to 2 and 5, respectively. The upper limit of the path number is set to 100.

For the GRU model, we adopt the ReLU activation function, use cross-entropy as the loss function, and employ

[2] http://iesl.cs.umass.edu/downloads/akbc16/

[3] http://nlp.stanford.edu/data/glove.6B.zip

AdaGrad as the optimizer. In the implementation, dropout is used on the output layer to prevent overfitting.

4.2 Link Prediction

We adopt several baseline models for comparison.

Graph structure learning methods: (1) PRA [19], a Path Ranking Algorithm for learning proximity measures on labeled graphs. (2) RWR [52], a PRA model to infer different target relations by tuning the weights associated with random walks that follow different paths through the graph. (3) NLFeat [53], a model that captures the compositional structure of textual relations..

Translation-based methods: (1) TransE [25], TransH [26], and TransR [27]. (2) TranSparse [54], a embedding model developed with adaptive sparse matrices, whose sparse degrees are determined by the number of entities (or entity pairs) linked by relations. (3) PTransE [31], a path-based representation learning model which considers relation paths as translations between entities for representation learning.

Neural network models: (1) RNN-PATH [28], an approach that reasons about conjunctions of multi-hop relations, composing the implications of a path using a recurrent neural network (RNN). (2) RNN chains of reasoning [38], a RNN model to compose the distributed semantics of multi-hop paths in knowledge graphs for complex reasoning about entities and relations. (3) Attention-GRU [29], a path-based model using the attention mechanism to highlight the useful paths for the discovery of new entity relation facts. (4) Att-Model+Types [41]: a path-based relation inference model that determines entity pair representations with type information and attentive path combination. (5) DeepPath [55]: a reinforcement learning framework for multi-hop relational path reasoning using a policy-based agent with continuous states based on knowledge graph embeddings to sample the most promising relation to extend its path. (6) Minerva [15]: a neural reinforcement learning approach which learns how to navigate the graph conditioned on the input query to find predictive paths. (7) PRCTA [56]: a path-based reasoning model incorporates constrained type attention but does not explicitly use the textual information.

To better analyze the effectiveness of TAPR, we also report the ablation test in terms of discarding the textual embeddings of entity description (w/o entity description), type embeddings (w/o type) and type attention (w/o type attention). We also compare the effectiveness of different types of path attention, i.e., One, Ave and PathAtt.

The experimental results on Freebase subset are summarized in Table 1 and Table 2. We make the following observations:

- The TAPR model has a comparatively substantial increase relative to other baseline models. The reason may be that a richer semantic knowledge of path information is taken into account compared to models that adopt only structural information or textual information in the path. In addition, unlike the knowledge representation model using single triples, path information takes advantage of all the triple information in paths, which brings significant improvement to the link prediction task.
- Our proposed model outperforms the PRCTA model [55] which also incorporates word-level and type-level attention for knowledge graph completion tasks but does not explicitly use the textual information, indicating the advantage of combining the structural, textural and path information included in the knowledge graph.
- Generally, both entity description and type information contribute, and it provides a larger performance boost to the link prediction. The basic TAPR model (w/o entity description and w/o type) cannot perform as well as the TAPR model, which indicates that description and type information provide supplementary knowledge for embedding, and thus the issue of representation sparseness in large scale knowledge graph can be settled properly.
- It is within our expectation that the adopted path

TABLE I
%MAP PERFORMANCE ON FREEBASE SUBSET

Model	Freebase subset
PRA (Ni et al., 2010)	64.43
RWR (Lao et al., 2012)	64.93
NLFeat (Toutanova et al., 2015)	73.15
TransE (Bordes et al., 2013)	47.10
TransH (Wang et al., 2014)	54.42
TransR (Lin et al., 2015)	58.73
TranSparse (Ji et al., 2016)	71.58
PTransE (Lin et al., 2015)	72.22
RNN-PATH (Neelakantan et al., 2015)	68.43
RNN chains of reasoning (Das et al., 2017)	72.22
Attention-GRU (Wen et al., 2017)	73.70
Att-Model+Type (Jiang et al., 2017)	73.42
DeepPath (Xiang et al., 2017)	75.96
Minerva (Das et al., 2018)	76.27
PRCTA (Lei et al. 2019)	76.44
TAPR+PathAtt	**78.52**

TABLE II
%MAP ABLATION TEST ON FREEBASE SUBSET

Model	Freebase subset
TAPR+PathAtt	**78.52**
TAPR+Ave	76.53 (-1.99)
TAPR+One	73.21 (-5.31)
w/o entity description	74.17 (-4.35)
w/o type	73.99 (-4.53)
w/o type attention	73.57 (-4.95)

attention mechanisms (TAPR+PathAtt) can significantly reduce noise and enhance path representation learning. The MAP of the TAPR+PathAtt model outperforms that of the TAPR+One model, which only contains information about one path involving a small amount of data. In the process of deep learning, few training data may lead to overfitting problems, making its generalization ability weak and leading to low MAP. Compared to the TAPR+Ave model, which assigns an average weight to all paths, the TAPR+PathAtt model increases the weight of reasonable paths and reduces the weight of unreasonable paths via the use of path attention, resulting in the highest precision.

- The results also specify that TAPR can reduce the impact of the noisy types with word-level and type-level attention, and thus it can work as designed to improve performance.

4.3 Entity Prediction

Table 3 and Table 4 report the experimental results of our proposed method on the entity prediction task and the performance of previous models that achieve the state of the art. There are multiple interesting observations, which are as follows:

1. Our proposed method is substantially and consistently superior to the state-of-the-art results in mean rank and Hits@10 on the Freebase subset, which demonstrates the advantage of the simultaneous consideration of KG structural information, textual information and type information.
2. It can be observed that TAPR outperforms the existing type-aware methods. This improvement is attributable to the use of soft type constraints and various semantic-level attention mechanisms, which can effectively reduce the impact of noise types.
3. Using an improper attention mechanism may result in a higher mean rank, since some wrong-predicted instances with an extremely high rank will significantly increase the mean rank.

TABLE III
EXPERIMENTAL RESULTS ON ENTITY PREDICTION

Model	Mean Rank	Hits@10
TransE	265	40.76
TransH	237	43.80
TransR	196	45.28
TranSparse	192	47.18
PTransE	184	47.33
RNN-PATH	177	46.28
RNN chains of reasoning	172	52.76
Attention-GRU	161	55.46
Att-Model+Type	158	55.84
DeepPath	155	56.42
Minerva	151	56.99
PRCTA	144	57.85
TAPR+PathAtt	**136**	**64.80**

TABLE IV
ABLATION TEST ON ENTITY PREDICTION

Model	Mean Rank	Hits@10
TAPR+PathAtt	**136**	**64.80**
TAPR+Ave	147	58.28
TAPR+One	150	57.33

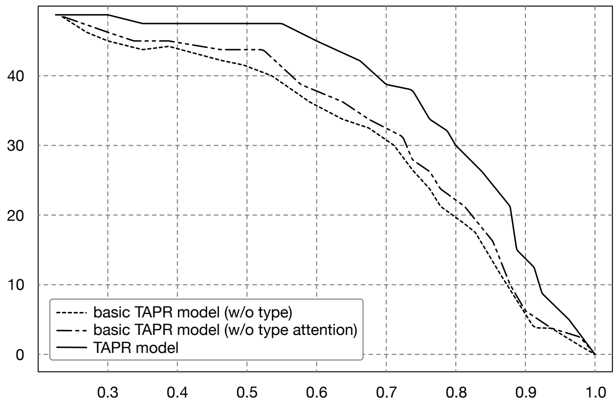


Fig. 3. Effect of type.

4.4 Effect of Type

To further analyze the performance of our model with respect to the type utilization, we report the MAP results of TAPR and its basic model (w/o type and w/o type attention) in Fig. 3. We can observe that, in the interval [0.5, 0.9] of MAP accuracy, the TAPR achieves a significant improvement, indicating that type information can largely alleviate the issue of sparsity when the structural and textual information can provide certain but insufficient support. In addition, the adopted attention mechanism can aid in reducing noises introduced by the redundancy of type information.

Furthermore, we show different performances of various models on paths with a large number of entity types and on paths with a limited number of entity types (see Fig. 4). From our experiments, we observe that our model is superior to other models on both subsets with different magnitude of noise, illustrating the advantage of considering various entity types. Fig. 4 further illustrates that as our model benefits from the proposed attentive denoising mechanism, our model can achieve more significant improvement on the subset with more entity types (2.12% better than PRCTA and 5.01% better than Att-Model+Types).

4.5 Case Study – Path Denoising

We have explored several versions of our model with different path denoising attention mechanisms. To evaluate the path denoising performance of the TAPR+PathAtt model, we randomly select three entities Suffolk County, Fire Island and Jean-Claude Gelin as an example.

Given the textual description of Jean-Claude Gelin "In

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2020.2970044, IEEE Transactions on Knowledge and Data Engineering

TABLE V
SAMPLED RELATIONAL PATH DISCOVERED BY TAPR

Triple	Score	Path
(infectious shock, medical department, emergency department)	max: 0.6169	(infectious shock, disease alias, septic shock), (septic shock, medical department, emergency department)
	min: 0.0320	(infectious shock, complication, disseminated intravascular coagulation), (disseminated intravascular coagulation, complication, abdominal pain), (abdominal pain, complication, electrolyte disturbance), (electrolyte disturbance, medical department, emergency department)
(beryllium poisoning, complication, pulmonary edema)	max: 0.6752	(beryllium poisoning, complication, pneumonia), (pneumonia, disease alias, lower respiratory infections), (lower respiratory infections, syndrome, pulmonary edema)
	min: 0.0047	(beryllium poisoning, disease examination, urinary calcium), (urinary calcium, possible disease with higher score, hypercalcemic nephropathy), (hypercalcemic nephropathy, complication, uremia), (uremia, complication, pulmonary edema)

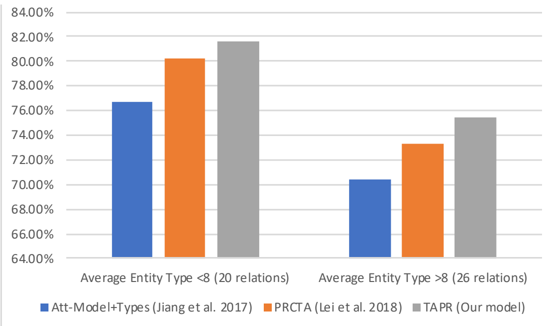


Fig. 4. Performance comparison on paths with a large number of entity types and on paths with a limited number of entity types.

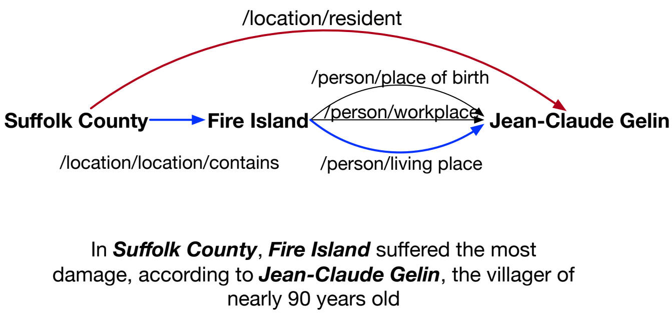


Fig. 5. Example – Path Denoising.

Suffolk County, Fire Island suffered the most damage, according to ***Jean-Claude Gelin***, the villager of nearly 90 years old", the conventional link prediction models may predict wrong relation "place of death" between entity Suffolk County and Jean-Claude Gelin probably because the appearance of entity "damage" and "90 years old".

However, there are various relation-specific types between location entity and person entity, such as /people/deceased_person/place_of_death, /person/place_of_birth, /person/workplace, /person/living_place etc.

In the case of considering type information and paying due attention to different relationship, our proposed model correctly identifies the relationship /location/resident between Suffolk County and Jean-Claude Gelin, indicating the robustness of the model to noisy paths (see Fig. 5).

4.6 Case Study – Link Prediction

Table 5 demonstrates two actual prediction examples of relational path reasoning selected from the test data. Each example has positive and negative instances with the highest and the lowest attention weight, respectively. In the case in which the triple is determined, KG semantic knowledge is adopted to further verify the authenticity of the specific relationship in an instance.

As the first triple indicates, the two entities connected by the relationship "alias" are basically different expressions of the same entity, so the path is logical. However, paths with lower scores have different level "complication" relations, which makes it difficult to distinguish the relationship between the two diseases "infectious shock" and "abdominal pain" from either pathological or semantic.

For the second triple, we observe that TAPR could discover some diseases that are not directly related to a certain symptom through the "alias" or "complications" relation. The path with the lower score speculates from "beryllium poisoning" to "pulmonary edema". The intermediate word, "uremia", is a kind of kidney disease with little connection to "beryllium poisoning". TAPR successfully identifies this instance as not a true case and assigns it a lower weight.

4.7 Computational Cost

To compare the model efficiency, we investigate the computational cost of the proposed TAPR model and several baseline methods. All these methods are run on a single NVIDIA GeForce GTX 1080 Ti. The processing time (minutes) on the Freebase subset is given in Table 6. Due to the limited space, we only report the running time of ten

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2020.2970044, IEEE Transactions on Knowledge and Data Engineering

TABLE VI
COMPUTATIONAL COST (MINUTES) OF EACH MODEL'S PROCESSING OF RELATIONSHIPS

	Att-Model+Types (Jiang et al. 2017)	PRCTA (Lei et al. 2018)	TAPR (Our model)
/cvg/game version/game	27	26	30
/education/educational institution/school type	29	28	33
/music/artist/label	31	29	34
/architecture/structure/address	25	25	28
/people/family/member	32	30	34
/tv/tv program/country of origin	28	27	31
/people/person/place of birth	42	42	44
/soccer/football player/position	36	36	38
/geography/river/cities	27	26	29
/people/person/nationality	26	26	28
All Relationship	491 minutes	455 minutes	570 minutes

different relationships. The running time of other relationships exhibit a similar trend. We also present the overall running time of each model's processing of all relationships.

The RNN-based methods (i.e., Att-Model+Types [41] and TAPR) is computationally expensive, which is caused by the complex operations in each RNN unit along the input sequence. In addition, we can observe that the consideration of various KG information is indeed computationally expensive, thus TAPR model have longer training time than other methods. Specifically, the processing time on Freebase subset is about 44 mins and 29 mins for the /people/person/place of birth relationship and /geography/river/cities relationship, respectively.

5 CONCLUSION

In this paper, we propose a type-aware attentive path reasoning method for knowledge graph completion by simultaneously considering KG structural information, textual information and type information. The type information is regarded as constraints in type selection. Experimental results on a real-world dataset demonstrate that our model is obviously superior to other baseline models on link prediction and entity prediction tasks.

In the future, we will consider improving the network embedding method to further explore KG structural information. In addition, we will utilize more attention mechanisms and information in the knowledge graph to find more relevance between triples.

ACKNOWLEDGMENT

This work was financially supported by the National Natural Science Foundation of China (No.61602013 and No. 61773229), the Shenzhen Fundamental Research Project (Grant No. JCYJ20170818091546869), and Basic Scientific Research Program of Shenzhen City (Grant No. JCYJ20160331184440545). The corresponding author Min Yang was sponsored by the CCF-Tencent Open Research Fund.

Ying Shen realized the development methodology and the creation of models. Ning Ding and Yaliang Li conducted an investigation process, and implemented algorithms and programming. Hai-Tao Zheng analyzed the experimental results. Min Yang was responsible for the management and coordination responsibility for the research activity planning and execution.

REFERENCES

[1] Y. Shen, Y. Li, S. Si, J. Zhang, M. Yang, K. Lei, "Gastroenterology Ontology Construction using Synonym Identification and Relation Extraction," IEEE Access, vol. 6, no. 1, pp. 52095-52104, 2018.

[2] K. Bollacker, C. Evans, P. Paritosh, T. Sturge, J. Taylor. (2008). Freebase: a collaboratively created graph database for structuring human knowledge. Presented at the 2008 ACM SIGMOD international conference on Management of data. pp. 1247-1250.

[3] G. A. Miller, "WordNet: a lexical database for English," Communications of the ACM, vol. 38, no.11, pp. 39-41, 1995.

[4] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, Z. Ives. (2007) Dbpedia: A nucleus for a web of open data. Presented at the semantic web. pp. 722-735.

[5] C. Deaton, B. Shepard, C. Klein, C. Mayans, B. Summers, A. Brusseau, D. Lenat. (2005). The comprehensive terrorism knowledge base in cyc. Presented at the 2005 International Conference on Intelligence Analysis.

[6] D. Vrandečić, M. Krötzsch. "Wikidata: a free collaborative knowledge base," Communications of the ACM. pp. 1-7, 2014.

[7] F. M. Suchanek, G. Kasneci, G. Weikum. (2007) Yago: a core of semantic knowledge. Presented at the 16th international conference on World Wide Web. pp. 697-706.

[8] X. Yao, B. Van Durme. (2014). Information extraction over structured data: Question answering with freebase. Presented at the 52nd Annual Meeting of the Association for Computational Linguistics.

[9] P. Nguyen, P. Tomeo, T. Di Noia, E. Di Sciascio. (2015, October). Content-based recommendations via DBpedia and Freebase: a case study in the music domain. Presented at the International Semantic Web Conference. pp. 605-621.

[10] X. Dong, E. Gabrilovich, G. Heitz, W. Horn, N. Lao, K. Murphy, W.

Zhang. (2014, August). Knowledge vault: A web-scale approach to probabilistic knowledge fusion. Presented at the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 601-610.

[11] P. Castells, M. Fernandez, D. Vallet, "An adaptation of the vector-space model for ontology-based information retrieval," IEEE transactions on knowledge and data engineering, vol. 19, no.2, pp. 261-272, 2007.

[12] Y. Shen, L. Zhang, J. Zhang, M. Yang, B. Tang, Y. Li, K. Lei, "CBN: Constructing a Clinical Bayesian Network based on Data from the Electronic Medical Record," Journal of Biomedical Informatics, vol. 88, pp. 1–10. 2018.

[13] R. West, E. Gabrilovich, K. Murphy, S. Sun, R. Gupta, D. Lin, (2014). Knowledge base completion via search-based question answering. Presented at the 23rd international conference on World wide web. pp. 515-526).

[14] M. Nickel, K. Murphy, V. Tresp, E. Gabrilovich. (2016). A review of relational machine learning for knowledge graphs. Presented at the IEEE 2016, vol. 104, no. 1, pp. 11-33.

[15] R. Das, S. Dhuliawala, M. Zaheer, L. Vilnis, I. Durugkar, A. Krishnamurthy, A. McCallum. (2018) Go for a walk and arrive at the answer: Reasoning over paths in knowledge bases using reinforcement learning. Presented at the Sixth International Conference on Learning Representations (ICLR 2018).

[16] Q. Wang, Z. Mao, B. Wang, L. Guo, "Knowledge graph embedding: A survey of approaches and applications," IEEE Transactions on Knowledge and Data Engineering, vol. 29, no.12, pp.2724-2743. 2017.

[17] R. Xie, Z. Liu, M. Sun. Representation Learning of Knowledge Graphs with Hierarchical Types. Presented at the 25th International Joint Conference on Artificial Intelligence IJCAI-16. pp. 2965-2971, 2016

[18] T. Trouillon, C. Dance, É. Gaussier, J. Welbl, S. Riedel, G. Bouchard, "Knowledge graph completion via complex tensor factorization," The Journal of Machine Learning Research, vol. 18, no.1, pp.4735-4772. 2017.

[19] N. Lao, W. Cohen, "Relational retrieval using a combination of path-constrained random walks," Machine learning, vol. 81, no.1, pp.53-67. 2010.

[20] Q. Wang, J. Liu, Y. Luo, B. Wang, C. Lin. (2016). Knowledge base completion via coupled path ranking. Presented at the 54th Annual Meeting of the Association for Computational Linguistics. vol. 1. no. 6. pp. 1308-1318.

[21] L. He, B. Liu, G. Li, Y. Sheng, Y. Wang, Z. Xu, "Knowledge base completion by variational bayesian neural tensor decomposition," Cognitive Computation, pp.1-10. 2018.

[22] I. Sutskever, J. Tenenbaum, R. Salakhutdinov. (2009). Modelling relational data using bayesian clustered tensor factorization. Presented at Advances in neural information processing systems. pp. 1821-1828.

[23] A. Bordes, X. Glorot, J. Weston, Y. Bengio, "A semantic matching energy function for learning with multi-relational data." Machine Learning, vol. 94, no. 2, pp. 233-259. 2014.

[24] B. Yang, W. Yih, X. He, J. Gao, L. Deng. (2014) Embedding entities and relations for learning and inference in knowledge bases. Presented at International Conference on Learning Representations 2015 (ICLR 2015). vol. 1412, no. 6575.

[25] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, O. Yakhnenko. (2013) Translating embeddings for modeling multi-relational data. In Processing of Advances in neural information processing systems. pp. 2787-2795.

[26] Z. Wang, J. Zhang, J. Feng, Z. Chen, (2014). Knowledge graph embedding by translating on hyperplanes. Presented at the Twenty-Eighth AAAI conference on artificial intelligence AAAI 2014. pp. 1112-1119.

[27] Y. Lin, Z. Liu, M. Sun, Y. Liu, X. Zhu, (2015). Learning entity and relation embeddings for knowledge graph completion. Presented at the Twenty-ninth AAAI conference on artificial intelligence AAAI 2015. pp. 2181-2187.

[28] A. Neelakantan, B. Roth, A. Mccallum, "Compositional Vector Space Models for Knowledge Base Completion," Computer Science, pp. 1-16. 2015.

[29] D. Wen, Y. Liu, K. Yuan, S. Si, Y. Shen. (2017). Attention-Aware Path-Based Relation Extraction for Medical Knowledge Graph. Presented at International Conference on Smart Computing and Communication. pp. 321-331.

[30] R. Socher, B. Huval, C. Manning, A. Ng. (2012). Semantic compositionality through recursive matrix-vector spaces. In Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning EMNLP 2012. pp. 1201-1211.

[31] Y. Lin, Z. Liu, H. Luan, M. Sun, S. Rao, S. Liu. (2015). Modeling relation paths for representation learning of knowledge bases. Presented at the 2015 Conference on Empirical Methods in Natural Language Processing, pp. 705–714, 2015.

[32] A. Neelakantan. (2017). Knowledge Representation and Reasoning with Deep Neural Networks. Doctoral Dissertations. 1114.

[33] M. Nickel, V. Tresp. (2013) Tensor factorization for multi-relational learning. Presented at the Joint European Conference on Machine Learning and Knowledge Discovery in Databases. pp. 617-621.

[34] W. Zeng, Y. Lin, Z. Liu, M. Sun. (2017) Incorporating Relation Paths in Neural Relation Extraction. Presented at the 2017 Conference on Empirical Methods in Natural Language Processing. pp. 1768–1777.

[35] M. Nickel, X. Jiang, V. Tresp. (2014) Reducing the rank in relational factorization models by including observable patterns. Presented at the Advances in Neural Information Processing Systems. pp. 1179-1187.

[36] Z. Wang, J. Li. (2016). Text-Enhanced Representation Learning for Knowledge Graph. Presented at the Twenty-Fifth International Joint Conference on Artificial Intelligence IJCAI 2016. pp. 1293-1299.

[37] S. Cao, W. Lu, Q. Xu. (2015). Grarep: Learning graph representations with global structural information. Prsented at the 24th ACM International on Conference on Information and Knowledge Management CIKM 2015. pp. 891-900.

[38] R. Das, A. Neelakantan, D. Belanger, A. McCallum. (2016). Chains of reasoning over entities, relations, and text using recurrent neural networks. Presented at the 15th Conference of the European Chapter of the Association for Computational Linguistics EACL 2016. vol 1, Long Papers, pp. 132–141.

[39] W. Chen, W. Xiong, X. Yan, W. Wang. (2018). Variational Knowledge Graph Reasoning. Presented at the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies NAACL HLT 2018. vol. 1, Long Papers. pp. 1823–1832.

[40] R. Xie, Z. Liu, M. Sun. (2016). Representation Learning of Knowledge Graphs with Hierarchical Types. Presented at the Twenty-Fifth International Joint Conference on Artificial Intelligence IJCAI 2016. pp. 2965-2971.

[41] X. Jiang, Q. Wang, B. Qi, Y. Qiu, P. Li, B. Wang. (2017). Attentive Path Combination for Knowledge Graph Completion. Presented at the Asian Conference on Machine Learning. pp. 590-605.

[42] D. B. Lenat and E. A. Feigenbaum, "On the thresholds of knowledge," Artificial intelligence, vol. 47, no. 1, pp. 185–250, 1991.

[43] G. Weikum, M. Theobald. (2010) From information to knowledge: harvesting entities and relationships from Web sources. Presented at the

twenty-ninth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems. pp. 65–76.

[44] B. Suh, G. Convertino, E. H. Chi, P. Pirolli. (2009) The Singularity is Not Near: Slowing Growth of Wikipedia. Presented at the 5th International Symposium on Wikis and Open Collaboration. New York, NY, USA: ACM, 2009, pp. 8:1–8:10.

[45] M. Clark, Y. Kim, U. Kruschwitz, D. Song, D. Albakour, S. Dignum, A. De Roeck, "Automatically structuring domain knowledge from text: An overview of current research," Information Processing & Management, vol. 48, no. 3, pp.552-568. 2012.

[46] T. Berners-Lee, J. Hendler, and O. Lassila, "The Semantic Web," 2001. [Online]. Available: http://www.scientificamerican.com/article/the-semantic-web/

[47] J. Biega, E. Kuzey, F. M. Suchanek. (2013) Inside YAGO2s: A transparent information extraction architecture. Presented at the 22Nd International Conference on World Wide Web. Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee, pp. 325–328.

[48] Annervaz, K. M., Chowdhury, S. B. R., & Dukkipati, A. (2018). Learning beyond datasets: Knowledge graph augmented neural networks for natural language processing. Presented at the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies NAACL-HLT 2018, pp. 313–322

[49] Cannaviccio, M., Barbosa, D., & Merialdo, P. (2016). Accurate fact harvesting from natural language text in wikipedia with Lector. Presented at the 19th International Workshop on Web and Databases. pp. 9.

[50] A. Carlson, J. Betteridge, B. Kisiel, B. Settles, E. R. H. Jr, T. M. Mitchell, Toward an Architecture for Never-Ending Language Learning. Presented at the Twenty-Fourth Conference on Artificial Intelligence (AAAI 2010). pp. 1306–1313.

[51] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078.

[52] N. Lao, T. Mitchell, W. Cohen. (2011). Random walk inference and learning in a large scale knowledge base. Presented at the 2011 Conference on Empirical Methods in Natural Language Processing EMNLP 2011. vol. 2012. pp.529-539.

[53] K. Toutanova, D. Chen, P. Pantel, H. Poon, P. Choudhury, M. Gamon. (2015). Representing text for joint embedding of text and knowledge bases. Presented at the 2015 Conference on Empirical Methods on Natural Language Processing EMNLP 2015. pp.1-8.

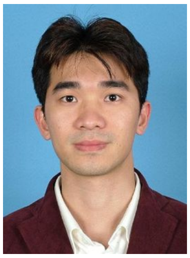
[54] G. Ji, K. Liu, S. He, J. Zhao. (2016). Knowledge Graph Completion with Adaptive Sparse Transfer Matrix. Presented at the Thirtieth AAAI Conference on Artificial IntelligenceAAAI 2016. pp. 985-991.

[55] W. Xiong, T. Hoang, W.Y. Wang. (2017). Deeppath: A reinforcement learning method for knowledge graph reasoning. Presented at the 2017 Conference on Empirical Methods in Natural Language Processing. pp. 1-10.

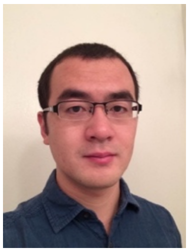
[56] K. Lei, J. Zhang, Y. Xie, D. Wen, D. Chen, M. Yang, Y. Shen, "Path-based reasoning with constrained type attention for knowledge graph completion," Neural Computing and Applications, pp. 1-10. 2019.

Ying Shen is now an Associate Professor in School of Intelligent Systems Engineering, Sun Yat-Sen University. She received her Ph.D. degree from the University of Paris Ouest Nanterre La Défense (France), specialized in Computer Science. She received her Erasmus Mundus Master degree in Natural Language Processing from the University of Franche-Comté (France) and University of Wolverhampton (England). Her research interests include Natural Language Processing and deep learning.

Ning Ding received the B.E degree in Computer Science from Huazhong University of Science and Technology. He is currently a Ph.D student at the Department of Computer Science and Technology, Tsinghua University. His research interest mainly focuses on Deep Learning, Natural Language Processing and Information Extraction.



Haitao Zheng is now an Associate Professor in School of Information Science and Technology at Tsinghua University. He received his Ph.D. degree from the Seoul National University (Korea), specialized in Medical Informatics. He received his Master and bachelor degree in Computer Science from the Sun Yat-Sen University (China). His research fields include Web Science, Semantic Web, Information Retrieval, and Machine Learning.



Yaliang Li received his Ph.D. degree in Computer Science from University at Buffalo, USA, in 2017. He is broadly interested in machine learning, data mining and information analysis. In particular, he is interested in analyzing information from multiple heterogeneous sources, including but not limited to information integration, knowledge graph, anomaly detection, data stream mining.

Min Yang is currently an Associate Professor with the Shenzhen Institutes of Advanced Technology, Chinese Academy of Science. She received her Ph.D. degree from the University of Hong Kong in February 2017. Prior to that, she received her B.S. degree from Sichuan University in 2012. Her current research interests include machine learning, deep learning and natural language processing. She is a member of the IEEE and the IEEE Computer Society.