

A Divide-And-Conquer Approach for Multi-label Multi-hop Relation Detection in Knowledge Base Question Answering

Deyu Zhou^{1*}, Yanzheng Xiang¹, Linhai Zhang¹, Chenchen Ye¹,
Qian-Wen Zhang² and Yunbo Cao²

¹School of Computer Science and Engineering, Key Laboratory of Computer Network and Information Integration, Ministry of Education, Southeast University, China

²Tencent Cloud Xiaowei, Beijing 100080, China

{d.zhou, yz_xiang, lzhang472, chenchenye}@seu.edu.cn,

{cowenzhang, yunbocao}@tencent.com

Abstract

Relation detection in knowledge base question answering, aims to identify the path(s) of relations starting from the topic entity node that is linked to the answer node in knowledge graph. Such path might consist of multiple relations, which we call multi-hop. Moreover, for a single question, there may exist multiple relation paths to the correct answer, which we call multi-label. However, most of existing approaches only detect one single path to obtain the answer without considering other correct paths, which might affect the final performance. Therefore, in this paper, we propose a novel divide-and-conquer approach for multi-label multi-hop relation detection (DC-MLMH) by decomposing it into head relation detection and conditional relation path generation. In specific, a novel path sampling mechanism is proposed to generate diverse relation paths for the inference stage. A majority-vote policy is employed to detect final KB answer. Comprehensive experiments were conducted on the FreebaseQA benchmark dataset. Experimental results show that the proposed approach not only outperforms other competitive multi-label baselines, but also has superiority over some state-of-art KBQA methods.

1 Introduction

Knowledge Graph (KG), an important form of structured human knowledge by organizing atomic facts in the triple format, i.e., (head, relation, tail), attracts more and more researchers. However, the ever-growing knowledge graphs make it extremely hard for users to access the information efficiently. To address this issue, Question Answering over Knowledge Base (KBQA) was proposed (Berant et al., 2013; Bordes et al., 2015; Bast and Haussmann, 2015; Yin et al., 2016; Hao et al., 2018). A KBQA system often consists of two core components: (1) entity linking, which identifies the topic

*Corresponding author.

Who is the granddaughter of Queen Elizabeth and George VI?

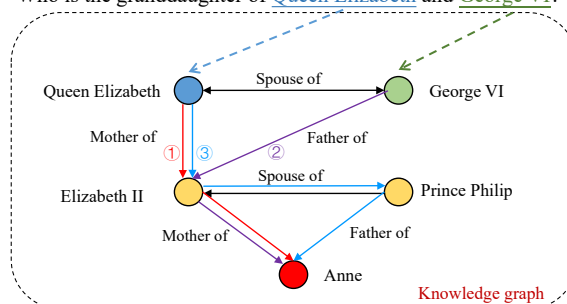


Figure 1: Example of relation paths corresponding to different questions.

entities mentioned in the question and links them to the entity nodes in the KG; (2) relation detection, which extracts the relation paths starting from the topic entity to the answer node based on the question. In this paper, we mainly focus on the second task.

Based on the numbers (hops) of KG triples required to obtain the answer, relation detection methods can be divided into two categories (Qiu et al., 2020), single-hop relation detection and multi-hop relation detection. For single-hop relation detection, only one triple is needed to answer the question. For instance, answering the question “Where is New York?” relies on the KB triple <New York, located in, America>. For multi-hop relation detection, more than one triples are needed to answer the question correctly. For example, to answer the question “When is the birthday of the author of Harry Potter”, two KB triples <Harry Potter, written by, J.K. Rowling> and <J.K. Rowling, birth date, 1965.7.31> are needed.

Most of relation detection approaches mentioned above aim to find one single relation path to answer the question. However, for some complex questions, there might exist multiple paths in the KG leading to the correct answer. As shown in Figure 1, given the question “Who is the granddaughter of Queen Elizabeth and George VI?”, there

are exactly eight relation paths to the final answer “Anne”. Two main reasons might contribute to the multiple relation paths: (1) Multiple topic entities, both ‘Queen Elizabeth’ and ‘George VI’ are topic entities which leads to different relation paths. (2) Diverse paths between nodes in the KG, due to the complex structure of the KG.

Obviously, performing multiple relation paths detection will make the answer prediction more robust. Therefore, in this paper, relation detection in KBQA is cast into multi-label (paths) learning. However, it is not straightforward to detect multiple relation paths with multiple hops since (1) the number of paths (label number) and the length of each path (hop number) are unfixed. (2) the number of valid relation paths (label space) is extremely large due to the exponential growth of relation combination.

To deal with the above challenges, a novel divide-and-conquer approach for multi-label multi-hop relation detection (DC-MLMH) is proposed by decomposing the problem into two sub-problems: (1) head relation detection, which is formulated as multi-label classification since multiple topic entities lead to diverse paths by choosing different head relations. A hybrid attention mechanism is proposed to capture semantic information in both the relations and question texts. (2) conditional relation path generation, long short-term memory network (LSTM) with attention is adopted to generate the entire path under the condition of a given head relation, and capture the order between the relations in the path. Moreover, in order to generate relations paths diversely, a path sampling mechanism is proposed to incorporate uncertainty into the neural network in the inference stage. Therefore, the number of paths and the length of the paths are dynamically determined and relation paths can be generated independently and diversely. Furthermore, the answer detection policy based on majority-vote is proposed to obtain the final answer from the multiple generated relation paths.

In summary, the main contributions of this paper are listed as follows:

- A novel divide-and-conquer approach for multi-label multi-hop relation detection (DC-MLMH) is proposed by decomposing it into head relation detection and conditional relation path generation. As far as we know, we are the first to tackle the detection of multiple relation paths with different lengths.

- A novel path sampling mechanism is proposed to generate diverse relation paths in the inference stage where the number of paths and the length of each path is determined dynamically. Moreover, a majority-vote policy is employed to detect final KB answer.
- Experimental results on the benchmark dataset FreebaseQA show that the proposed approach outperforms other competitive multi-label baselines. It also has superiority over some state-of-art KBQA methods.

2 Related Work

Our work is related to two lines of research, relation detection in KBQA and multi-label learning.

2.1 Relation Detection in KBQA

There are two mainstream branches of relation detection in KBQA (Fu et al., 2020): information retrieval-based and neural semantic parsing-based, which can be further categorized into single-relation (one-hop) detection and multi-relation (multi-hop) detection (Yin et al., 2016).

For single-relation detection, most existing approaches (Yin et al., 2016; Dai et al., 2016; Yu et al., 2017; Lukovnikov et al., 2017; Hao et al., 2018; Yu et al., 2018) embedded questions and candidate relations into the same space and calculated the semantic similarity between the questions and the candidate relations in KBs. For example, (Golub and He, 2016) propose an encoder-decoder based generative framework for relation detection, (Zhang et al., 2021) propose a end-to-end KBQA model based on Bayesian Neural Network (BNN) to estimate uncertainties arose from both model and data.

For multi-relation detection, some approaches (Yih et al., 2015; Yu et al., 2017, 2018) incorporated a constraint detection mechanism to deal with two or three-relation detection tasks. The drawback of such methods is that the number of hops is generally strictly restricted. (Chen et al., 2019) proposed UHop, an unrestricted-hop framework which doesn’t need predefined maximum hop number. (Bordes et al., 2014a,b) constructed the topic-entity-centric subgraph for each question and retrieved answers by ranking the semantic similarity between the question embedding and entity embedding in the subgraph. Such methods suffer from high time complexity as the whole subgraph for

each question is considered. However, all the aforementioned approaches ignore that for some complex questions, there might exist multiple relation paths leading to the correct answer.

2.2 Multi-label Learning in NLP

Multi-label Learning (MLL) approaches can be categorized into the following types: (1) problem transformation methods (Boutell et al., 2004; Read et al., 2011) which transforming MLL into other well-established tasks; (2) algorithm adaptation approaches (Zhang and Zhou, 2007; Ghamrawi and McCallum, 2005) that adapt the existing algorithms to tackle MLL data; (3) deep learning based methods (Nam et al., 2017; Yang et al., 2018; Peng et al., 2019) which exploit deep neural networks to extract high-level features and capture label correlations.

MLL has been applied to multiple crucial tasks in the area of Natural Language Processing (NLP) including question answering (Usbeck et al., 2017), information retrieval (Gopal and Yang, 2010; Jiang et al., 2016), emotion classification (Yang et al., 2019) and so on.

Unlike traditional MLL methods where the label set is fixed, the proposed approach needs to predict the relation path (label) which might not exist in the set of relation paths (label set) in the training data.

3 Methodology

3.1 Problem Setting

In this paper, a KG is represented as $\mathcal{G} = \{\mathcal{E}, \mathcal{R}, \mathcal{F}\}$, where \mathcal{E} , \mathcal{R} and \mathcal{F} are set of entities, relations and facts. A fact is denoted as a triple $\{e_s, r, e_o\} \in \mathcal{F}$.

Given a natural language question $q = \{w_1, w_2, \dots, w_n\}$ and the knowledge graph \mathcal{G} , the KBQA task is to detect the relation paths which link the topic entity mentioned in q to the answer node in \mathcal{G} . In this paper, we consider multi-label multi-hop relation detection which aims at finding a path set \mathbb{P} rather than a single path p_i :

$$\begin{aligned} \mathbb{P} &= \{p_1, \dots, p_i, \dots, p_m\} \\ p_i &= \{r_1^i, \dots, r_j^i, \dots, r_{l_i}^i | r_j^i \in \mathcal{R}\} \end{aligned} \quad (1)$$

3.2 The Architecture

In this section, we introduce the proposed divide and conquer approach for multi-label multi-hop relation detection (DC-MLMH) in details.

3.2.1 Overview

Multi-label multi-hop relation detection can be regarded as a sequence generation problem. We adopt a divide-and-conquer strategy and decompose it into two sub-problems: head relation detection and conditional relation path generation. The object of head relation detection is to detect a set of head relations \mathbb{R}_{head} . Then the conditional path generation generates the entire relation path $p_i \in \mathbb{P}$ for each head relation $r_{head}^i \in \mathbb{R}_{head}$:

$$\begin{aligned} \mathbb{R}_{head} &= \{r_1, \dots, r_i, \dots, r_h | r_i \in \mathcal{R}\} \\ p_i | r_{head}^i &= (r_{head}^i, r_2^i, \dots, r_j^i, \dots, r_{l_i}^i | r_j^i \in \mathcal{R} \\ &\quad, r_{head}^i \in \mathbb{R}_{head}) \end{aligned} \quad (2)$$

The architecture of the proposed divide and conquer approach for multi-label multi-hop relation detection (DC-MLMH) is shown in Figure 2. It consists of three components, (1) BERT (Devlin et al., 2018) encoder, which is shared by the following two components; (2) Head relation detection, head-relation distribution over relation space \mathcal{R} is predicted. A Hybrid-attention mechanism is proposed to capture the semantic information in question and relation texts; (3) Conditional relation path generation, LSTM is adopted to generate the relation path sequentially.

Besides, In order to increase the diversity of generated paths, we propose a novel path sampling mechanism to incorporate uncertainty into the neural network in the inference stage. Therefore, relation paths can be generated diversely and independently.

3.2.2 Head Relation Detection

As mentioned above, one of the main reasons for multiple relation paths is that the question might contain different topic entities. Therefore, the relation paths might comprise multiple head relations focusing on different components of question texts and topic entities.

We leverage BERT (Devlin et al., 2018) to encode question text. The final hidden state vector $H \in \mathbb{R}^{n \times d}$ is served as the representation of the question context. A hybrid attention mechanism is proposed to capture relation-specific parts of the question for each relation.

Relation Attention Mechanism Relations are represented as a trainable embedding matrix $C \in \mathbb{R}^{|V| \times k}$ where $|V|$ is the size of relation vocabulary and k is the dimension of the embedding vector.

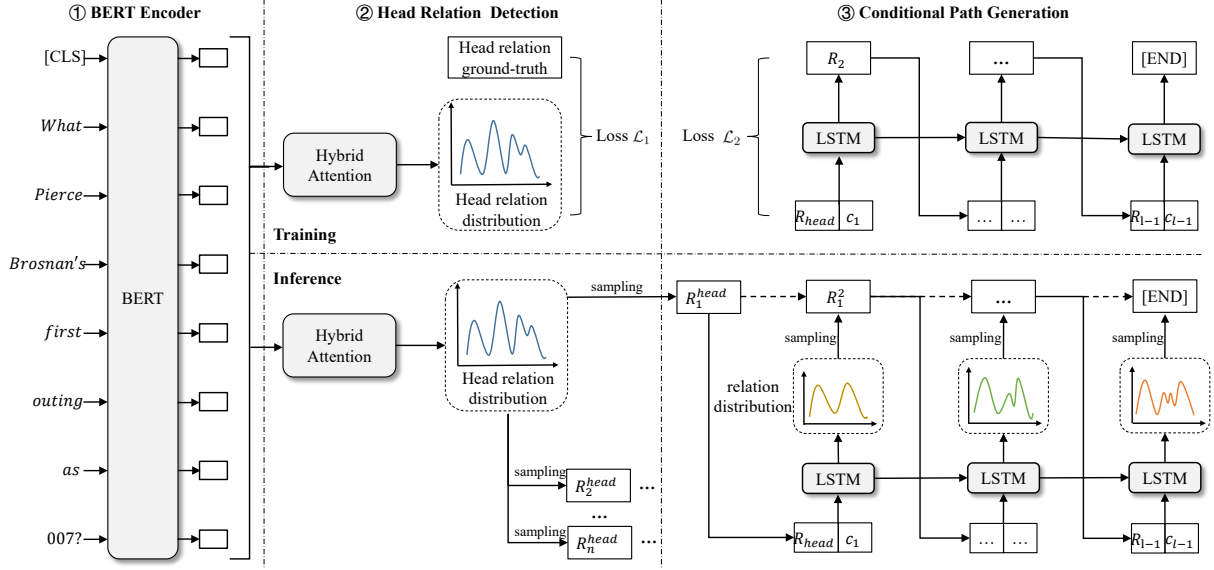


Figure 2: The architecture of the proposed divide-and-conquer approach where head relation distribution is obtained by the head relation detection module and the whole path is generated in the conditional relation path generation module. The training process is in the upper part of the figure while the inference process with the proposed path sampling is in the lower part of the figure.

Each relation embedding is initialized using the bag-of-words representation of the text of the relation:

$$C_t = \sum_{i=1}^{n_t} e_i^t \quad (3)$$

where n_t is the number of the words in the t -th relation text and e_i^t is the i -th word of the t -th relation text retrieved from pre-trained Glove embeddings (Pennington et al., 2014).

We can explicitly determine the semantic relevance between each pair of word and relation by calculating the dot product between word and relation vectors:

$$A_r = \text{softmax}(CW_1H^T) \quad (4)$$

where W_1 is weight parameter to convert the two vectors into the same dimension and $A_r \in \mathbb{R}^{|V| \times n}$ indicates the semantic similarity distribution between words and relations (For simplicity, all bias terms are omitted in this paper). Relation-aware representation $M_r \in \mathbb{R}^{|V| \times d}$ is constructed through linear combination of question tokens for all relations:

$$M_r = A_r H \quad (5)$$

Self-Attention Mechanism Different relations have diverse semantic relevance to the different

part of the question (Xiao et al., 2019). Therefore, self-attention mechanism (Lin et al., 2017) is adopted to determine the different components of question texts for each relation. The self-attention matrix $A_s \in \mathbb{R}^{|V| \times n}$ can be calculated as follows:

$$A_s = \text{softmax}(W_3 \tanh(W_2 H^T)) \quad (6)$$

where W_2, W_3 are self-attention parameters. The i -th row of A_s can be considered as the weight distribution of the question tokens over the i -th relation.

Similar to relation-aware representation, self-attention representation $M_s \in \mathbb{R}^{|V| \times d}$ is constructed as:

$$M_s = A_s H \quad (7)$$

Relation-specific question representation $M_{all} \in \mathbb{R}^{|V| \times 2d}$ is obtained by simply concatenate the two representations:

$$M_{all} = [M_s; M_r] \quad (8)$$

The probability of each relation to be the head relation can be calculated via:

$$\hat{y}^{head} = \text{sigmoid}(M_{all} W_4) \quad (9)$$

where W_4 is weight parameter of the fully connected layer and sigmoid is used as activation function.

Cross-entropy loss \mathcal{L}_1 is adopted as the loss function for multi-label text classification following the same way as (Nam et al., 2014):

$$\mathcal{L}_1 = - \sum_{i=1}^N \sum_{j=1}^{|V|} (y_{ij} \log(\bar{y}_{ij}^{head}) + (1 - y_{ij})(1 - \log(\bar{y}_{ij}^{head}))) \quad (10)$$

where N is the number of questions, $|V|$ is the size of relation vocabulary, $y_{ij} \in \{0, 1\}$ is the ground truth of the i -th question over j -th relation and $\bar{y}_{ij}^{head} \in [0, 1]$ is the predicted probability.

3.2.3 Conditional Relation Path Generation

The remaining sequence of the path under the condition of the head relation is generated through conditional relation path generation module. It consists of a BERT encoder shared with the head relation detection module and a decoder. An attention mechanism is employed to decide the different parts of the question text in the each step of the conditional relation path generation process. The attention weight α_{tj} assigned to the j -th token in the t -th step is computed as follows:

$$e_{tj} = V_a \tanh(W_a [s_{t-1}; H_j]) \quad (11)$$

$$\alpha_{tj} = \frac{e_{tj}}{\sum_{i=1}^n e_{ti}} \quad (12)$$

where W_a, V_a are parameters and s_{t-1} is the hidden state of the decoder at time step $t - 1$. Finally, the context input of the decoder at time step t is computed as the weighted sum of question tokens:

$$c_t = \sum_{i=1}^n \alpha_{ti} H_i \quad (13)$$

LSTM (Hochreiter and Schmidhuber, 1997) is employed to generate relation paths sequentially. It can capture the correlations between relations.

The input format of decoder is:

$$\{R_{head}, R_2, \dots, R_l, [END]\} \quad (14)$$

where l represents the length of the input relation path and the first input R_{head} is the head relation of the path rather than a “[BOS]”. It is obtained in

head relation detection module. The hidden state s_t at time t is calculated by:

$$s_t = LSTM(s_{t-1}, [c_t; r_{t-1}]) \quad (15)$$

where r_{t-1} is the embedding of the relation predicted at time step $t - 1$.

The relation probability distribution \vec{y}_t over relation space at time step t is computed as follows:

$$\vec{y}_t = softmax(\tanh(W_5 s_t)) \quad (16)$$

where W_5 is weight parameter. The training loss of the conditional relation path generation module \mathcal{L}_2 can be calculated the same as the head relation detection module. The total loss function \mathcal{L} is the sum of the loss functions of the two modules:

$$\mathcal{L} = \mathcal{L}_1 + \mathcal{L}_2 \quad (17)$$

3.3 Training and Inference

As shown in the upper part of Figure 2, the ground-truth head relations and relation paths are employed to train the DH-MLMH model. The ground truth paths of a question refer to the relation paths that exist between the topic-entity nodes of the question and the answer-entity node in the KG. In our proposed approach, relation detection is regarded as a sequence generation task. Traditional methods generate sequences by searching over output sequences greedily or with beam search. Even with these mechanisms, the model is still insufficient to generate diverse sequence outputs. In the inference stage, we propose a novel path sampling mechanism to incorporate uncertainty into the neural network that can generate relation paths diversely and independently. The number of paths for each question and the length of the paths are determined dynamically.

Path Sampling Mechanism The process of path sampling mechanism is illustrated in Algorithm 1. The probability of each relation to be the head relation is calculated using (9). We set the threshold to 0.5 to filter the relations and normalize the probability of the remaining ones to form a multinomial distribution $\vec{\theta}_{head}$. A randomly-sampled head relation r_{head} is chosen as the first input of the decoder. Then, the remaining path is generated cyclically through the decoder. At the t -th step of the path generation process, the relation probability distribution can be obtained via (16). Due to the large relation space, we select the top- k highest

Algorithm 1 Path sampling mechanism.

Input: Question q , sampling times T **Output:** Predicted path set \mathbb{P}

```
1:  $\mathbb{P} \leftarrow \emptyset$ 
2: for  $i \leftarrow 1; i \leq T; i++$  do
3:   Relation Sequence  $R \leftarrow ()$ 
4:   Calculate the probability for each relation becoming the head relation (9) and construct the multinomial distribution  $\vec{\theta}_{head}$ .
5:   Sample a head relation  $r_{head}$  from  $\vec{\theta}_{head}$ .
6:    $R \leftarrow R + r_{head}$ 
7:    $r_t \leftarrow r_{head}$ 
8:   while  $r_t \neq [END]$  do
9:     Input  $r_t$  into the decoder and calculate the current relation probability distribution  $\vec{y}_t$  (16). Choosing top-K highest probability relations to form  $\theta_t$ .
10:    Sample a relation  $r_g$  from  $\theta_t$ .
11:     $R \leftarrow R + r_g$ 
12:     $r_t \leftarrow r_g$ 
13:  end while
14:   $\mathbb{P} \leftarrow \mathbb{P} \cup \{R\}$ 
15: end for
16: return  $\mathbb{P}$ 
```

probability relations for normalization to form a new multinomial distribution $\vec{\theta}_t$, thereby avoiding the introduction of noise. A relation r_t is sampled from the relation probability distribution $\vec{\theta}_t$ which is taken as the input for t+1-th step. The process ends until “[END]” is generated and these generated relations make up a path. The length of paths can be determined according to the KG structure information learned by the decoder. The times of sampling T is a hyperparameter which can be set randomly. The final path set \mathbb{P} is generated and the number of paths can be determined through the path sampling process.

It should be noted that the same path might appear multiple times in \mathbb{P} . However, such duplicated paths are not removed as the more occurrences, the more likely it is to be an answer relation path.

3.4 Multi-label Relation Detection-based KBQA

In order to obtain the final answer in the KG, entity-linking is adopted to link the topic entities mentioned in the question with the corresponding entity nodes in the KG. Besides, a majority-vote (MV) policy is designed to integrate the results of entity-linking and relation detection.

Entity Linking Entity linking associates the topic entities with the KG entity nodes. Following the previous approaches (Lukovnikov et al., 2017; Mohammed et al., 2018), the entity linking task is formulated as a fuzzy string matching problem. For each topic entity in the corpus, the matching score between it and each entity in the KG is calculated based on Levenshtein distance. We rank the matching scores and the top k entities form the set of the candidate entities \mathbb{E} .

Majority-Vote Path set \mathbb{P} and candidate entity set \mathbb{E} can be obtained after relation detection and entity linking. We search for candidate answers in the KG using all the entity-relation pairs. The candidate answer appearing the most times will be selected as the final answer. Based on the majority vote strategy, the negative effects of some noise paths and entities are alleviated.

4 Experiment

In order to evaluate the proposed approach, we conduct experiments on the benchmark dataset FreebaseQA (Jiang et al., 2019).

4.1 Dataset

Data set	Total	Train	Dev	Eval
FreebaseQA	28348	20358	3994	3996

Table 1: Statistics of the FreebaseQA dataset.

FreebaseQA contains 28K unique questions in total. The basic statistics are shown in Table 1. It is created by matching the trivia-type question-answer pairs with head-relation-tail triples in Freebase knowledge graph (Bollacker et al., 2008). FreebaseQA provides many linguistically sophisticated questions. Jiang et al. demonstrate that it is a more difficult KBQA task than WebQSP (Berant et al., 2013) and SimpleQuestions (Bordes et al., 2015).

In addition, FreebaseQA dataset provides multiple relation paths with up to two hops for each question as shown in Table 2. As far as we know, it is the only dataset annotated with multiple relation paths.

4.2 Parameter Settings

BERT-base-uncased (12 layers, 768 hidden dimensions, and 12 attention heads) released by Google is adopted as our encoder. We use pre-trained Glove

Num of paths	Num of questions	proportion
$N = 1$	16065	56.7%
$N = 2$	6842	24.1%
$N = 3$	2908	10.3%
$N = 4$	1235	4.4%
$N \geq 5$	1298	4.5%

Table 2: Statistics of the number of paths in the FreebaseQA dataset.

word embeddings of size 300 to initialize relation matrix. For LSTM decoder, the hidden state dimension is set to 512. The whole model is trained by the Adam optimizer (Kingma and Ba, 2014) with a learning rate of $1e-5$, $1e-4$ for BERT encoder and other modules respectively. The number of epochs is 10 and the mini-batch size of the input is set at 16. The parameters are chosen based on the evaluation results from dev dataset.

4.3 Multi Relation Detection Task

Baselines As the proposed approach aims to multiple relation paths detection, the following competitive multi-label classification approaches are chosen baselines:

- **MLKNN (Zhang and Zhou, 2007)**: K -Nearest Neighbors are found for each unseen instance, then maximum a posterior (MAP) principle is adopted to predict the label set based on the statistic information learned from the label sets of the neighbors.
- **CNN (Kim, 2014)**: Convolutional neural network is employed to extract text features, and linear layer with the sigmoid function is to calculate the probability of each label.
- **HAN (Yang et al., 2016)**: a hierarchical structure network with word-level and sentence-level attention mechanism for document classification.
- **SGM (Yang et al., 2018)**: a Seq2Seq model with a decoder structure to capture the correlations between labels. We further replace the RNN encoder with BERT encoder (BERT-SGM) for fair comparison as BERT encoder is employed in our approach.

Evaluation Metrics Following the previous work (Yang et al., 2018; Zhang and Zhou, 2007),

hamming loss and micro- F_1 score are adopted as the evaluation metrics.

- **Hamming Loss (Schapire and Singer, 1999)**: It denotes the fraction of misclassified labels, where the relevant label is missed or the irrelevant label is predicted.
- **Micro- F_1 (Schütze et al., 2008)**: It is an aggregated metric considering both the precision and recall for all classes.

Results Results of different multi-relation detection approaches on FreebaseQA benchmark are listed in Table 3. To further analysis the effectiveness of the hybrid attention mechanism, the proposed DC-MLMH model is modified to **DC-MLMH-SELF** by removing the relation-attention mechanism from DC-MLMH, to **DC-MLMH-RELA** by removing the self-attention mechanism from DC-MLMH. The sampling times for DC-MLMH-SELF-100, DC-MLMH-RELA-100 and DC-MLMH-100 are set to 100 while for DC-MLMH-5, the times of sampling are set to 5.

It can be summarized from the table that: (1) our proposed DC-MLMH and the two modifications outperform all baselines by a large margin on Micro- F_1 and recall score. Compared to BERT-SGM, DC-MLMH-100 achieves an improvement of 16.4% Micro- F_1 score and 36.5% recall which demonstrate that the proposed path sampling mechanism can effectively detect diverse relation paths. (2) Compared with DC-MLMH-100, DC-MLMH-5 achieves an improvement of 6.8% precision and a reduction of 2.5% hamming loss, but its recall and Micro-F1 drop by 19.4% and 7.9%. It means that as the times of sampling increases, our approach can detect more diverse paths. Although there is a slight decrease in precision, the improvement in recall is huge. (3) DC-MLMH-SELF-100 and DC-MLMH-RELA-100 perform worse than DC-MLMH-100 which proves the effectiveness of the two proposed attention mechanisms.

4.4 Knowledge Based Question Answering Task

A complete KBQA system includes Entity-linking and Relation Detection. We perform a simple entity-linking and measure the performance of our proposed method by reasoning in the KG to find final answers with the detected relation paths.

Methods	Precision(\uparrow)	Recall(\uparrow)	Micro F1(\uparrow)	HL($\times 10^{-4}$)(\downarrow)
MLKNN	0.5327	0.3287	0.4066	1.4049
CNN	0.5158	0.3952	0.4475	1.4285
HAN	0.4965	0.4254	0.4582	1.4728
SGM	0.5039	0.3976	0.4445	1.4549
BERT-SGM	0.5992	0.4372	0.5056	1.2437
DC-MLMH-SELF-100	0.5431	0.5579	0.5504	1.3340
DC-MLMH-RELA-100	0.5268	0.5505	0.5384	1.3817
DC-MLMH-5	0.6199	0.4810	0.5417	1.1913
DC-MLMH-100	0.5803	0.5966	0.5883	1.2219

Table 3: Performance comparison of multi-label methods on relation detection. SELF denotes self-attention mechanism, RELA denotes Relation attention mechanism. “HL” represents Hamming loss. (\uparrow) represents “the larger the better” while (\downarrow) is the opposite.

Baselines Two approaches are chosen as the baselines:

- **FOFE-net** (Jiang et al., 2019): A pipeline KBQA system built based on FOFE-net (Zhang et al., 2015) which achieves the outstanding results on both SimpleQuestions and WebQSP datasets.
- **BERT-SGM**: a multi-label method mentioned above.

Evaluation Metrics We evaluate the quality of the KBQA systems based on accuracy of the predicted answers. The quality of entity-linking is evaluated based on recall which refers to the fraction of the correct topic entities that included in the candidate entities.

K	Recall
25	0.7640
50	0.8168
100	0.8694
200	0.8994
500	0.9331

Table 4: Results of entity linking.

Results We create an inverted index for the entities of the FB5M (Petrochuk and Zettlemoyer, 2018), and top K entities sorted by Levenshtein Distance are considered as candidate entities. Table 4 shows the result of entity-linking. As K increases, the recall continues to increase with more noisy entities.

Method	Accuracy
FOFE-net	37.0%
BERT-SGM-GT	38.9%
DC-MLMH-EL	37.7%
DC-MLMH-GT	47.5%
DC-MLMH-GT-SP	35.4%

Table 5: Knowledge based question answering results of different approaches on FreebaseQA test set.

Table 5 shows the performance of different KBQA systems. “EL” means using the results of entity-linking while “GT” represents using ground truth entities to find answers, “SP” means only considering one single path to obtain the final KB answer. It can be observed that DC-MLMH-GT outperforms BERT-SGM-GT with a large margin. Although noises and errors are introduced in the entity-linking stage, DC-MLMH-EL which K is set to 25 in entity-linking still has superiority over FOFE-net.

To further explore whether multi-label relation detection has superiority than single-label relation detection, only one of the generated relations paths and the ground-truth entities (DC-MLMH-GT-SP) is employed to find the final KB answer. It can be observed that the accuracy of DC-MLMH-GT-SP is lower by 25.5% compared to DC-MLMH-GT which uses all the generated relation paths to find the answer. It shows that multi-label multi-hop relation detection indeed improves the performance of KBQA.

5 Conclusion

In this paper, a divide-and-conquer multi-label multi-hop relation detection approach is proposed. In specific, relation detection is decomposed into two steps: head-relation detection and conditional relation-path generation. A novel path sampling mechanism is proposed to incorporate uncertainty into the neural network and generate diverse paths independently in the inference stage. A Majority-Vote policy is employed to integrate the entity-linking and multi-label multi-hop relation detection results to obtain the final answer. Experimental results on the FreebaseQA benchmark dataset show that the proposed method outperforms other competitive multi-label baselines. It also achieves superiority over some state-of-art KBQA methods.

Acknowledgements

We would like to thank anonymous reviewers for their valuable comments and helpful suggestions and we thank Tencent for supporting this project. This work was funded by the National Natural Science Foundation of China (61772132).

References

- Hannah Bast and Elmar Haussmann. 2015. [More accurate question answering on freebase](#). In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, CIKM '15*, page 1431–1440, New York, NY, USA. Association for Computing Machinery.
- Jonathan Berant, Andrew Chou, Roy Frostig, and Percy Liang. 2013. [Semantic parsing on Freebase from question-answer pairs](#). In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1533–1544, Seattle, Washington, USA. Association for Computational Linguistics.
- Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. 2008. Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, pages 1247–1250.
- Antoine Bordes, Sumit Chopra, and Jason Weston. 2014a. Question answering with subgraph embeddings. *arXiv preprint arXiv:1406.3676*.
- Antoine Bordes, Nicolas Usunier, Sumit Chopra, and Jason Weston. 2015. Large-scale simple question answering with memory networks. *arXiv preprint arXiv:1506.02075*.
- Antoine Bordes, Jason Weston, and Nicolas Usunier. 2014b. Open question answering with weakly supervised embedding models. In *Joint European conference on machine learning and knowledge discovery in databases*, pages 165–180. Springer.
- Matthew R Boutell, Jiebo Luo, Xipeng Shen, and Christopher M Brown. 2004. Learning multi-label scene classification. *Pattern recognition*, 37(9):1757–1771.
- Zi-Yuan Chen, Chih-Hung Chang, Yi-Pei Chen, Jijnasa Nayak, and Lun-Wei Ku. 2019. Uhop: An unrestricted-hop relation extraction framework for knowledge-based question answering. *arXiv preprint arXiv:1904.01246*.
- Zihang Dai, Lei Li, and Wei Xu. 2016. Cfo: Conditional focused neural question answering with large-scale knowledge bases. *arXiv preprint arXiv:1606.01994*.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Bin Fu, Yunqi Qiu, Chengguang Tang, Yang Li, Haiyang Yu, and Jian Sun. 2020. A survey on complex question answering over knowledge base: Recent advances and challenges. *arXiv preprint arXiv:2007.13069*.
- Nadia Ghamrawi and Andrew McCallum. 2005. Collective multi-label classification. In *Proceedings of the 14th ACM international conference on Information and knowledge management*, pages 195–200.
- David Golub and Xiaodong He. 2016. Character-level question answering with attention. *arXiv preprint arXiv:1604.00727*.
- Siddharth Gopal and Yiming Yang. 2010. Multilabel classification with meta-level features. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, pages 315–322.
- Yanchao Hao, Hao Liu, Shizhu He, Kang Liu, and Jun Zhao. 2018. Pattern-revising enhanced simple question answering over knowledge bases. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 3272–3282.
- Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation*, 9(8):1735–1780.
- Kelvin Jiang, Dekun Wu, and Hui Jiang. 2019. [FreebaseQA: A new factoid QA data set matching trivia-style question-answer pairs with Freebase](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 318–323, Minneapolis, Minnesota. Association for Computational Linguistics.

- Xiaotian Jiang, Quan Wang, Peng Li, and Bin Wang. 2016. Relation extraction with multi-instance multi-label convolutional neural networks. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 1471–1480.
- Yoon Kim. 2014. [Convolutional neural networks for sentence classification](#). In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1746–1751, Doha, Qatar. Association for Computational Linguistics.
- Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Zhouhan Lin, Minwei Feng, Cicero Nogueira dos Santos, Mo Yu, Bing Xiang, Bowen Zhou, and Yoshua Bengio. 2017. A structured self-attentive sentence embedding. *arXiv preprint arXiv:1703.03130*.
- Denis Lukovnikov, Asja Fischer, Jens Lehmann, and Sören Auer. 2017. Neural network-based question answering over knowledge graphs on word and character level. In *Proceedings of the 26th international conference on World Wide Web*, pages 1211–1220.
- Salman Mohammed, Peng Shi, and Jimmy Lin. 2018. [Strong baselines for simple question answering over knowledge graphs with and without neural networks](#). In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 291–296, New Orleans, Louisiana. Association for Computational Linguistics.
- Jinseok Nam, Jungi Kim, Eneldo Loza Mencía, Iryna Gurevych, and Johannes Fürnkranz. 2014. Large-scale multi-label text classification—revisiting neural networks. In *Joint european conference on machine learning and knowledge discovery in databases*, pages 437–452. Springer.
- Jinseok Nam, Eneldo Loza Mencía, Hyunwoo J Kim, and Johannes Fürnkranz. 2017. Maximizing subset accuracy with recurrent neural networks in multi-label classification. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pages 5419–5429.
- Hao Peng, Jianxin Li, Senzhang Wang, Lihong Wang, Qiran Gong, Renyu Yang, Bo Li, Philip Yu, and Lifang He. 2019. Hierarchical taxonomy-aware and attentional graph capsule rcnns for large-scale multi-label text classification. *IEEE Transactions on Knowledge and Data Engineering*.
- Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.
- Michael Petrochuk and Luke Zettlemoyer. 2018. Simplequestions nearly solved: A new upper-bound and baseline approach. *arXiv preprint arXiv:1804.08798*.
- Yunqi Qiu, Yuanzhuo Wang, Xiaolong Jin, and Kun Zhang. 2020. Stepwise reasoning for multi-relation question answering over knowledge graph with weak supervision. In *Proceedings of the 13th International Conference on Web Search and Data Mining*, pages 474–482.
- Jesse Read, Bernhard Pfahringer, Geoff Holmes, and Eibe Frank. 2011. Classifier chains for multi-label classification. *Machine learning*, 85(3):333.
- Robert E Schapire and Yoram Singer. 1999. Improved boosting algorithms using confidence-rated predictions. *Machine learning*, 37(3):297–336.
- Hinrich Schütze, Christopher D Manning, and Prabhakar Raghavan. 2008. *Introduction to information retrieval*, volume 39. Cambridge University Press Cambridge.
- Ricardo Usbeck, Michael Hoffmann, Michael Röder, Jens Lehmann, and Axel-Cyrille Ngonga Ngomo. 2017. Using multi-label classification for improved question answering. *arXiv preprint arXiv:1710.08634*.
- Lin Xiao, Xin Huang, Boli Chen, and Liping Jing. 2019. Label-specific document representation for multi-label text classification. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 466–475.
- Pengcheng Yang, Xu Sun, Wei Li, Shuming Ma, Wei Wu, and Houfeng Wang. 2018. Sgm: sequence generation model for multi-label classification. *arXiv preprint arXiv:1806.04822*.
- Yang Yang, Deyu Zhou, Yulan He, and Meng Zhang. 2019. Interpretable relevant emotion ranking with event-driven attention. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 177–187.
- Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. 2016. Hierarchical attention networks for document classification. In *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies*, pages 1480–1489.
- Scott Wen-tau Yih, Ming-Wei Chang, Xiaodong He, and Jianfeng Gao. 2015. Semantic parsing via staged query graph generation: Question answering with knowledge base.

- Wenpeng Yin, Mo Yu, Bing Xiang, Bowen Zhou, and Hinrich Schütze. 2016. Simple question answering by attentive convolutional neural network. *arXiv preprint arXiv:1606.03391*.
- Mo Yu, Wenpeng Yin, Kazi Saidul Hasan, Cicero dos Santos, Bing Xiang, and Bowen Zhou. 2017. Improved neural relation detection for knowledge base question answering. *arXiv preprint arXiv:1704.06194*.
- Yang Yu, Kazi Saidul Hasan, Mo Yu, Wei Zhang, and Zhiguo Wang. 2018. Knowledge base relation detection via multi-view matching. In *European Conference on Advances in Databases and Information Systems*, pages 286–294. Springer.
- Linhai Zhang, Chao Lin, Deyu Zhou, Yulan He, and Meng Zhang. 2021. [A bayesian end-to-end model with estimated uncertainties for simple question answering over knowledge bases](#). *Computer Speech Language*, 66:101167.
- Min-Ling Zhang and Zhi-Hua Zhou. 2007. Ml-knn: A lazy learning approach to multi-label learning. *Pattern recognition*, 40(7):2038–2048.
- Shiliang Zhang, Hui Jiang, Mingbin Xu, Junfeng Hou, and Li-Rong Dai. 2015. The fixed-size ordinally-forgetting encoding method for neural network language models. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 495–500.