# Combining Neural Network Models with Rules for Chinese Knowledge Base Question Answering

**Pengju Zhang[1],Kun Wu[1],Zongkui Zhu[12],Yonghui Jia[1],Xiabing Zhou[1],Wenliang Chen[1],Dongcai Lu[2]**

[1] **School of Computer Science and Technology, Soochow University, China**
{pjzhang,kwu,yhjia}@stu.suda.edu.cn
{zhouxiabing,wlchen}@suda.edu.cn
[2] **Huawei cloud & ai**
{zhuzongkui,ludongcai}@huawei.com

**Abstract.** The task of Chinese Knowledge Base Question Answering (CKBQA) in the open domain can be handled as a process of subject entity determination, candidate relationship and path determination, and answer determination based on the knowledge base. In order to find the correct answers, it is important to accurately identify the subject entity and list the candidates of relationship and path. Our system consists of three components: path similarity model, relationship similarity model, and rule similarity model. Finally, we combine the three components as our final system with certain strategic rules. Our final system achieves an average F1 score of 0.73075 on the test set in the CCKS 2019 CKBQA task.

**Keywords: Question Answering · Named Entity Recognition · Entity Linking**

## 1 Introduction

Chinese Knowledge Base Question Answering (CKBQA) refers to a process of determining the subject entity, candidate relationship or path determination based on the knowledge base, and finally determining the answer. The representation of the knowledge map here is a knowledge map, it consists of ' the entity(node)-relation(edge)-the entity value(node)/the attribute value (node)'. The whole process of Q&A is to identify one/more subject entities from the question, and then determine the relationship of the subject entity. The relationship here is divided into a single-hop relationship (only one knowledge triple is associated) and a multi-hop relationship (multi knowledge triples are associated), and finally find answers in the knowledge map based on the identified subject entities and their corresponding relationships.

This paper proposes three model systems: path similarity model, relationship similarity model[9] and rule similarity model. In determining the candidate relationship/path part, unlike the previous method, we propose to use the path to replace the relationship for answer candidates.

This is because the methods used to determine candidate relationships are mostly similarity calculations and relationship classifications. From the perspective of similarity calculation, the simple calculation of the relationship would not take the information of the entity into consideration, in which the context semantics may be lost in the calculation. On the other hand, if the relationship candidates are selected from the perspective of classification, there are too many types of Chinese relations. In the case of less training data, the accuracy of classification will be greatly reduced, so classification is not a good choice. Therefore, we still choice the similarity calculation with adding information such as entity and sentence structure of the path. For example, "《我不是主角》的主角是谁？ 《我不是主角》-主角-PAD", the sentence is computing Similarity by path, in which "PAD" is generalizing the entity value/attribute value. Since the tail entity may not appear in the question,the entity value or attribute value cannot be considered when calculating the similarity. When calculating the similarity by using the relationship,such as"《我不是主角》的主角是谁？ 主角", The structure and path of this sentence has lacked the information about the structure of the sentence，The word "主角" should appear on the R (relationship) in S-R-O when determining the candidate relationship, but if the relationship is used to calculate the similarity, the position information of the word "主角" cannot be reflected.

However, in order to achieve the best results, our path similarity calculation and relationship similarity calculation are realized, and they can complement each other.At the same time, we construct a rule model to correct the results of the above two similarity models, and solving some problems that are difficult to solve by the above similarity models. Finally, we combined the results of these three parts and achieved certain results.

## 2 Related Work

Chinese Knowledge Base Question Answering (CKBQA) is an important task in natural language processing. The current mainstream methods are: semantic analysis based methods and retrieval based methods[1].

The idea of semantic analysis of CKBQA is to transform the question into a semantic representation that allows the knowledge base to "understand" through semantic analysis of natural language, then reasoning the final answer by the knowledge of the knowledge base. In short, what semantic parsing needs to do is transforming the problem of natural language into a semantic representation that allows the knowledge base to "understand". However, this method is difficult to implement.

The retrieval-based method is a process of determining the subject entity, candidate relationship or path determination based on the knowledge base, then finally determining the answer[4][10]. In the the subject entity determination, the most commonly used method is to use sequence labeling model to solve the problem. However, similarity calculation is the most used method in candidate relationship or path determination[2]. This method is used in this paper.

## 3 System Introduction

The flow chart of the whole system is shown in **Fig.**1.

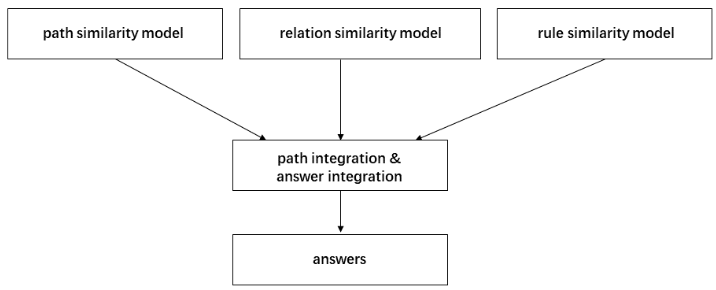


**Fig. 1.** Path similarity model, relationship similarity, rule similarity model are the models of our three subsystems. When they generate their own optimal paths, they perform path fusion operations to select the optimal path as the final result of system 1. Then the answers of the three subsystems are merged as the final answer.

### 3.1 System 1: Path similarity matching

The answer path, which we define as: the order of entities and predicates related to the question, within a certain number of triples from the process of finding the answer from the entity.

For one sentence, the number of candidate answer paths is indeterminate. When a question can find an answer within one hop, the number of candidate answer paths is $n_e \times n_r$ ($n_e$ represents the number of entities, $n_r$ refers to the average number of relationships connected by a single entity). When a question is two hop question, the number of candidate answer paths is $n_e \times n_{r_1} \times n_{r_2}$. Regardless of the question classification, the candidate path size of each sentence is in $n_e \times n_{r_1} \times n_{r_2}$. Question classification can reduce the order of magnitude of candidate paths and improve the efficiency of the system.

The path similarity matching system adopts a method of classifying questions to improve the performance and efficiency of the system. The types of questions in this paper include multi-limit questions, one-hop questions, chain questions, and difficult-to-categorize questions. Question classification classifies each question into one category and regenerates the candidate path of the corresponding template. By this method, the order of the total number of question candidate paths is reduced, which greatly improves the efficiency of the system. The process of the path similarity matching system is shown in **Fig**.2.

Firstly, the BERT sequence labeling[7] model is used to identify the entity of the question, and the entity information obtained by the violent retrieval is

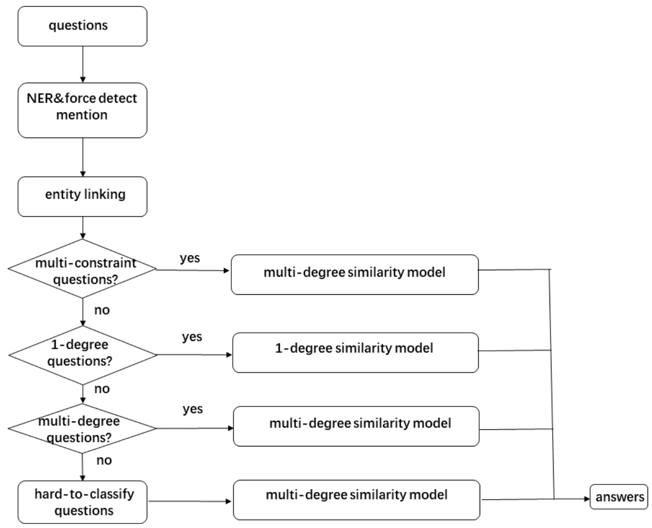


**Fig. 2.** System 1: framework of Path similarity matching

used to perform exact matching and fuzzy matching to obtain the sentence in the question and complete the entity link.

Secondly, the BERT-trained classifier and the set rules are used to divide the questions into multi-limit questions, one-hop questions, chain questions, and difficult-to-categorize questions.

Finally, candidate paths for different templates are generated for different types of questions, and the questions and paths are calculated by the BERT similarity model to obtain the optimal path and the final answer is generated.

### 3.2 System 2: Relationship similarity matching

The process of the relationship similarity matching system is shown in **Fig**.3. The first step is to classify the question (to determine whether it is a one hop question or multi-hop question, to judging whether a question is the form of subject-predicate-object, or not) and to refer to the mention entity (using operations similar to the NER form) after getting the problem.

In the second step, expanding or deleting based on the identified mentions, searching all candidate entities in the knowledge base according to the mentions tables, and the candidate entities are sorted according to a set of features (the entity link model), select the candidate entity with the highest score as the subject entity.

In the third step, according to the S-P-O value, all relationships corresponding to entities are searched in the knowledge base. Semantic similarity (relation extraction model) is calculated between all relationships corresponding to each entity and the question, and the candidate relationship with the highest

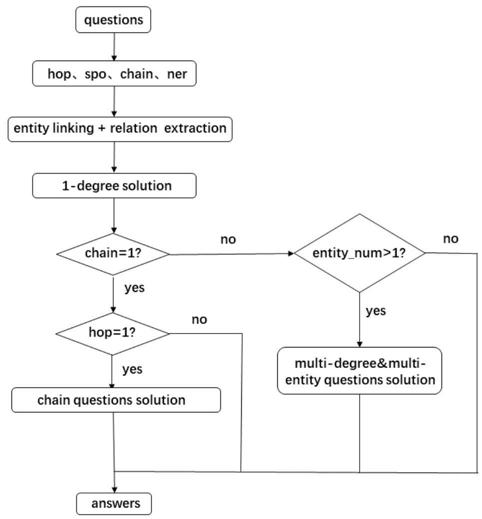


**Fig. 3.** System 2: framework of Relationship similarity matching

score is selected. Then, according to the determined entity-relationship, searching Knowledge Base to solve one-hop question.

Finally, if the question is a chain and multi-hop question, the answer obtained in the third step is used as the entity to perform the third step again, and the solution of the chain question is obtained; If the question is unchained and containing multiple entities, searching for all candidate triples corresponding to database queries for each entity, then find intersection to solve multi-hop and multi-entity questions. In addition, Baidu Encyclopedia Search is added in the search for all candidate entities besides mentioning left and right extensions.Violent search is used if entities are not found after linking. If multiple mentions are identified, the physical links also correspond to multiple, if multiple references are identified and entity links[6] correspond to multiple references, only the highest score is used when searching for answers to chain questions.

### 3.3 System 3: Rule-based method

The rule-based method selects the best answer by calculating the similarity between the path information and the words in the question. Since only the character level similarity is included in the calculation of the similarity, no additional feature information is introduced, so simply defined as the rule method.

**Question segmentation**

Chinese Knowledge Base Question Answering (CKBQA) based on the rule method is the most close to human thinking. It needs to divide the question into the words closest to the representation in the knowledge base before calculating the similarity between the question and answer paths. Such as “获得NBA最有价值球员奖(2008)的球员的球衣号码是多少？”, “NBA最有价值球员奖(2008)” should be treated as a word, which makes it easier to perform similarity calculations. In order to achieve the above word segmentation effect, words or phrases in the knowledge base can be used as a dictionary for word segmentation. In addition, since the knowledge in the knowledge base has errors, especially in dealing with shorter words, the word segmentation results can be adjusted in combination with the existing word segmentation tools. The resulting word segmentation results can be used for path similarity matching.

**Decomposition of question type**

When using the rule method for answer search, in order to perform path search, we divide the problem into one hop and two hop. The single hop problem includes 2 types, and the two hop problem includes 6 types[8]. Table 1 shows the structure of each question type.

**Table 1.** one-hop and two-hop question structure

| type | corresponding structure |
|---|---|
| one-hop type 1 | S-P-? |
| one-hop type 2 | ?-P-O |
| two-hop type 3 | S1-P1-O1/S2-P2-? |
| two-hop type 4 | S1-P1-O1/O2-P2-? |
| two-hop type 5 | O1-P1-?/?-P2-O2 |
| two-hop type 6 | O1-P1-S1/S2-P2-? |
| two-hop type 7 | S1-P1-?/?-P2-O2 |
| two-hop type 8 | S1-P1-?/?-P2-S2 |

**Search snswer**

When performing an answer search, we first calculate the corresponding score corresponding to the answer in the one-hop path. In order to reduce the time and space consumption when searching for two-hop questions, we can set the threshold to filter some unlikely paths. According to the experience of practice, we can generally set the length threshold of the same word in the relational word to be 2, and in the attribute value, the length threshold of the same word can also be set to 2. Then performing a triple search according to the structure of the problem and calculate the answer similarity value that satisfies the above eight problem patterns. In addition, After structural analysis of the question, we can find some rules. Given the combination of two triples of entities or attribute values, the relationship word thresholds may not be set (such as types 5, 7, and

8). However, when only one entity or attribute value is given, the relational word threshold is indispensable (such as types 3, 4, and 6). After completing the score calculation, the answer corresponding to the path with the highest score is the final answer.

## 4 Subsystem Model Introduction

In the previous chapter, this paper introduced the whole system implementation process. In this chapter, the model method adopted by each subsystem will be introduced.

### 4.1 System 1 model

The models of System 1 can be divided into two categories:

The first type is the question classification model, which is used to classify the questions and divide the total questions into different sub-questions to solve.

The second type is the semantic similarity model, which is used to calculate the similarity between the question and the candidate path, so that the optimal path can be selected and the answer can be found through the optimal path.

**Question classification model** (**BERT**)

***One-multi-hop questions classification model data***

The one-multi-hop questions classification model is used to divides the question into two types: one-hop and multi-hop. The one-hop question refers to the question: you only need a triple to get the answer(such as:'姚明-妻子-叶莉'). A multi-hop question refers to a question that requires more than two triples to get an answer(such as:'姚明-妻子-星座-天蝎座'). We label the training set with one-multi-hop questions and train the model.

***Chained question classification model data***

The chain question classification model is used to classify questions into chain and unchain questions. The chain question is marked according to the sparql grammar. For example, the sentence “计算机-发明者?x ?x 民族?y”, which can be expressed as “计算机-发明者-民族”. We label the training set with one-multi-hop questions and train the model.

***One-multi-hop questions classification model structure***

We use the BERT pre-trained model for fine-tune, which converts the classification of one-hop questions, multi-hop questions and chain questions into a two-category problem[3].

**Semantic similarity model（one-multi-hop questions semantic similarity model）**

***Semantic similarity model of one-hop question***

The training corpus of the input model is composed of lable-que-path(such as: "1-姚明的老婆是谁？-姚明-妻子-pad"), "que" is a question, "path" is our candidate path, "pad" here refers to the generalized obj, we generalize the "pad" for all obj, to prevent each different obj from affecting the que-path similarity. Lable is used to determine whether path is the correct path for the que, lable =1 represents that path is the correct path, lable=0 represents the error, We used a 1: 100 positive-negative ratio for training. The model uses the bert pre-training model, which translates the similarity problem of que and path into a two-class problem that determines if the que-path is correct. Finally, in the prediction, the que-path question pair with the highest score is selected, as our predicted que-path, and then the candidate path is obtained.

***Multi-hop question classification model***

The model input corpus of the chained question classification is basically the same as the one-hop question similarity model.(such as: "1-姚明的老婆的星座是什么？-姚明-妻子-星座-pad"). The difference is that the que consists of one-multi-hop questions, and the corresponding path becomes one-multi-paths. However, since the sub, rel, and obj positions of the multi-hop candidate path are different, there are many types of multi-hop paths. But we are only using the chained question path(For example, the path mentioned in the chained question classification model). Because the training set constructed in this way has a large length and limited computing resources, the positive and negative examples of the bert pre-trained model fine-tune change to a positive-negative ratio of 1: 50. Since other types of multi-degree questions are not huge in the training set, the chain questions is replaced by the whole multi-degree questions to train the similarity model.

### 4.2 System 2 model

**Mention entity model （BERT+BiLSTM+CRF）**

We use BERT+BiLSTM+CRF as the sequence labeling model as the model structure[5][3].

The training data is constructed based on the mention to the SPARQL grammar reverse lookup entity of the training corpus.

When we only use one-hop data, the initial recall rate is 0.8728. After the improvement (removing the mentioned book name and double quotes), the result is: 0.8947, but only 0 or 1 entity can be mentioned from the question. Then, adding all the data of one-multi-hop, the model test result is: 0.8824, and multiple references can be mentioned.

**Entity link model（XGBOOST）**

In the entity link model section, this article takes six features: order, mention score, questions and mention char matching, questions and entity semantic similarity, qustions and entity character matching, Maximum similarity between qustions and entity relationship. The model adopts the method of xgboost to do 0-1 classification, and the model test accuracy rate is 0.8915.

**Relation extraction model（BERT）**

We use BERT as a question classification model.

The training data is the same method used in the system 2 to calculate the semantic similarity. However, the training data was changed: the que-rel pair consisting of the problem and candidate relationship words, the candidate relationship words adopt a positive and negative ratio of 1: 5 (random 5 negative examples).

The initial accuracy on single-hop data is: 0.837, the mask-off entity mentions 0.8106, and the final model is to mix all data for NLPCC2016 and CCKS2019. We made improvements: the negative example of NLPCC2016 data is extracted from the ensemble of relationships, and the negative example of CCKS2019 data is extracted from all candidate relationships of the corresponding entities.

## 5 Our Ensemble System

### 5.1 Path integration

None of the three systems can involve paths of all patterns. In order to expand candidate paths, path integration takes best paths of three systems into account. First, get answer paths of the three systems. Then adapt multi-degree similarity model in system 1 to determine the best path. After the procedure, we expand the candidate set of the answer paths. Experiments result shows that this method significantly improves the performance of our system.

### 5.2 Answer integration

Since the three systems have their own advantages in dealing with different types of problems, this paper uses the answers given by the three systems to integrate the answers. First, the type of the answer is determined according to the prompt word in the question, for example, when the words "在哪里,是哪个国家" appear in the question, the answer type is always related to types like "地点" or "国家". Next, after searching types of every answer from types table, we compare the question type, the answer type and screen out the wrong type of answers. Finally, when the types of questions and answers do not match, we use the voting method to complete the final answer search. When we can not distinguish three answers, we select the answer from system 1 with best performance. After answer integration, the performance of entire system is significantly improved.

## 6 Experimental Setup

There are two sources of data used in training our system: one is CCKS 2019 CKBQA dataset，another is NLPCC 2016 KBQA dataset.

### 6.1 Official dataset

CCKS 2019 CKBQA dataset contains 2,298 train, 776 valid and 776 test QA pairs, as **table 2**,**table 3**. Pkutriple contains triples for searching answers, pkutype for checking types of entites and pkumen for matching alias of entities. Details are shown in .

**Table 2.** details of 2019 ccks ckbqa dataset

| filename | number of data |
|---|---|
| pkutriple | 41,007,385 |
| pkutype | 25,182,627 |
| pkumen | 13,930,029 |
| task6ckbqa-train-2019 | 2,298 |
| task6ckbqa-dev-2019.txt | 776 |

### 6.2 Extenal dataset

Our experiments use one external dataset which released in NLPCC-ICCPOL 2016 open domain CKBQA task. The dataset enriches our data and helps training our model, as **table** 3.

**Table 3.** details of extenal dataset

| filename | number of data |
|---|---|
| nlpcc-iccpol-2016.kbqa.training-data | 14,609 |
| nlpcc-iccpol-2016.kbqa.testing-data | 9,780 |

### 6.3 Main results

In this paper, three subsystems are submitted separately for comparison and fusion. In order to avoid random errors in deep learning model, path similarity system and relationship similarity system are both adopt for fusion. Experiments prove semantic similarity calculation plays the most important role.

At the same time, rule similarity system can trim the results of the first two systems from the character level similarity calculation. For example, “NBA最

有价值球员奖(2008)” in “获得NBA最有价值球员奖(2008)的球员的球衣号码是多少？” should be the subject entity, but neither path similarity system nor relationship similarity system can accurately identify the subject entity, however rule similarity system can recognize it from the character level. In this way, path similarity system and relationship similarity system can be amended, and that the effect of integration of the three will be significantly improved, as **Table** 4.

**Table 4.** system score on the final test set

| Subsystem | F1 |
|---|---|
| Path similarity system | 0.6941 |
| Relationship similarity system | 0.6468 |
| Rule similarity system | 0.5212 |
| Final system | 0.7308 |

## 7 Summary and Outlook

This paper proposes a joint system for solving the complex questions of CKBQA, which combines path similarity model, relationship similarity model and rule similarity model. The system achieves the F1-score of 0.73075 on CCKS 2019 CKBQA task. However, the system fails to solve much more complex questions, such as the ones with another hop after multiple constraints. In the future, we plan to work in solutions for those complex questions.

## 8 Acknowledgement

The research work is supported by the National Natural Science Foundation of China (Grant No. 61525205, 61876115). In addition, thanks to Yeqian Wu and Haoqiang Liang, interns of the huawei cloud & ai department , for their help in this work.

## References

1. Bao, J.W., Duan, N., Zhou, M., Zhao, T.: An information retrieval-based approach to table-based question answering. In: NLPCC (2017)
2. Chen, Z.Y., Chang, C.H., Chen, Y.P., Nayak, J., Ku, L.W.: Uhop: An unrestricted-hop relation extraction framework for knowledge-based question answering. In: NAACL-HLT (2019)
3. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. ArXiv **abs/1810.04805** (2018)
4. Lai, Y., Lin, Y., Chen, J., Feng, Y., Zhao, D.: Open domain question answering system based on knowledge base. In: NLPCC/ICCPOL (2016)

5. Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., Dyer, C.: Neural architectures for named entity recognition. In: HLT-NAACL (2016)
6. Li, Y.D., Miao, Q., Yin, C., Huo, C., Mao, W., Hu, C., Xu, F.: A joint model of entity linking and predicate recognition for knowledge base question answering. In: CCKS Tasks (2018)
7. Ma, X., Hovy, E.H.: End-to-end sequence labeling via bi-directional lstm-cnns-crf. ArXiv **abs/1603.01354** (2016)
8. Sun, Z., Song, L., Yu, J.: A qa search algorithm based on the fusion integration of text similarity and graph computation. In: CCKS Tasks (2018)
9. Zhang, M., Zhang, Y., Fu, G.: End-to-end neural relation extraction with global optimization. In: EMNLP (2017)
10. Zhao, W., Chung, T., Goyal, A.K., Metallinou, A.: Simple question answering with subgraph ranking and joint-scoring. ArXiv **abs/1904.04049** (2019)