A BRNN Based Question Focus Extraction Model and its Application to a Chinese Airport Question Answering System

Junpeng Bao, Meng Tian, Hui He, Zhonghui Feng and Ke Yang

Department of Computer Science & Technology, Xi’an Jiaotong University, West Xianning Road No.28, Xi’an, P.R. China
Email: Baojp@mail.xjtu.edu.cn

Abstract. Question focus extraction is one of the key of question analysis in a question answering system. This paper presents a simple model (Fk-BRNN) based on bidirectional recurrent neural network (BRNN) for question focus extraction, and applies it to a Chinese Airport Question Answering (CAQA) system. The Fk-BRNN model memorizes different sentence patterns and the focus positions in each sentence pattern, then it extracts focus words at the corresponding positions according to different sentence patterns. For a question, the Fk-BRNN model can extract not only one or more known focus words with correct semantics, but also unknown new focus words. So it can greatly reduce the size of training corpus, and keep excellent generalization ability. As a result, it is more practical and suitable for online learning system. Based on the above ideas, this paper designs and implements a CAQA system, which can accurately answer the questions with a superior performance according to our experimental results.

1. Introduction

Question Answering (QA) system is one of the hotspots in Natural Language Processing research. According to answer generation mechanism, QA systems can be divided into two types, i.e. the retrieval QA systems [1-5] and the generative QA systems [6-8]. According to the question domain, QA systems can be divided into specific domain oriented QA systems [9] and open domain oriented QA systems, such as IBM Watson [1, 5] and Apple Siri [10].

A retrieval QA system uses a question analysis result to retrieve existing answers in a knowledge base and then gives the best answer. The key of this mechanism lies in how to establish the connection between the answer and the question. One way is to design query rules based on the question analysis result and background knowledge, and then query the best answer in the knowledge base according to the query rules. Another way is to sort all candidate answers by matching degree between the answer and the question, and select the answer with the highest matching degree. A generative QA system automatically generates an answer that consists of a sequence of words. The key of this mechanism is to build a natural language generation model using a large amount of interactive data.

A Chinese Airport Question Answering (CAQA) system is a kind of specific domain oriented QA system. It has to consider a lot of background knowledge, such as relevant laws, airport rules, and airport facilities, in order to get the best answer. At present, a generative QA system can hardly make full use of background knowledge to produce a professional answer. It is very often to make a misleading or wrong answer, so it is not suitable for the CAQA system.

This paper presents a bidirectional recurrent neural network based model Fk-BRNN to precisely extract question focus words from a Chinese airport question, and then a rule based inference method uses the question foci to retrieve the correct answer. The question focus extraction [11-14] is one of the
key issues for a QA system. Most existing focus extraction algorithms are based on syntax analysis or pattern matching [11-13]. Due to the diversity of natural language expressions, especially in Chinese, these algorithms have low accuracy and poor scalability. The Fk-BRNN model can simultaneously extract one or more focus words in a question and distinguish the meaning of them. Moreover, The Fk-BRNN model can identify unknown new focus words so that it greatly reduces the size of training corpus.

2. Related Work

Early QA systems were mostly oriented to specific domain, such as LUNAR [9]. The successful application of Apple Siri System [10] and IBM Watson System [1, 5] bring QA system into a new era. Deep learning methods are widely used in QA systems. Many researchers [2-8] use deep learning models to select answer. Tan et al. [5] retrieve the answers that have the highest match degree with the question by a LSTM (Long Short-Term Memory) network. Now, it is popular to make a generative QA system with an encoder-decoder architecture based on a Recurrent Neural Network, such as [6-8].

The traditional keyword extraction algorithms, such as TextRank [11] and TF-IDF based methods are not suitable for focus extraction, especially for multi-foci extraction. Early focus extraction algorithms are mostly based on syntax analysis or pattern matching. For example, Damljanovic et al. [12] extract focus based on syntax analysis. Bayoudhi et al. [13] extract focus through a semi-automatic construction of the lexicon. Ku et al. [14] extract focus by filtering. Zhang et al. [15] use CRF (Conditional Random Field) model to extract focus.

In recent years, the recurrent neural network has been widely used in many fields, especially in natural language processing field, such as speech recognition [16], picture caption [17], language model [18], and machine translation [19]. Bahdanau et al. [19] put an attention mechanism into encoder-decoder model, which greatly improves the accuracy of machine translation. The basic idea is that the importance of the source language words is different according to the different target words. It also inspires us with the Fk-BRNN model.

3. The Fk-BRNN Model and the CAQA System

3.1 The Fk-BRNN Model

3.1.1 The basic idea. The focus in a question is a word or a few words that are crucial for answer retrieval. For example, the focus of the questions about luggage is the luggage name. The focus word of the question “Can I take some bananas on the plane?” is bananas. Since every word in a question may be the focus words. We have to consider the meaning of the word, the influence of all the other words on that word and the relative position of each word when the probability of focus word is calculated. It is not applicable in practice to extract focus words by syntax rules. Rules need to be designed manually by language experts but natural language texts are flexible, non-normative and ambiguous. The statistical focus extraction methods need a lot of training corpus data to get more accurate word distribution probability. Besides, the statistical method is difficult to process the new and unknown words. To solve these problems, this paper proposes the Fk-BRNN model, which learns and memorizes the different sentence patterns, then gives the most likely positions of the focus words. Finally one or more focus words are extracted. The Fk-BRNN model can adapt to the unknown words and accurately output the unknown focus words.

The sentence patterns in the Fk-BRNN model are different from those determined by part of speech tags or syntactic tags. A query question pattern need not to be identical with the sentence pattern memorized in the Fk-BRNN model. As long as they are similar enough, the focus words can be correctly extracted by the Fk-BRNN model. Indeed, The Fk-BRNN model does not directly memorize any focus word itself, but the possible position of a focus in each sentence pattern. Therefore, the Fk-BRNN model can extract unknown focus word even if it does not occur in the training data. Namely, the Fk-BRNN model has very good generalization ability.
3.1.2 Multi-foci extraction network model. A question may contain \(k (k\geq1)\) different focus words. The meaning of these words and the role they play in matching rules are different. Therefore, the Fk-BRNN model uses a multi-foci extraction bidirectional recurrent neural network not only to extract multiple focus words, but also to distinguish different meanings of them by means of setting multiple output groups in the output layer. Each output group extracts a type of focus. As a result, the model can extract multiple focus words simultaneously and distinguish their meanings.

![Figure 1. Multi-foci extraction network model structure expanded by time.](image)

Figure 1 shows the multi-foci extraction network model structure expanded by time. \(x_t\) denotes the word vector of the \(t\)-th word in the sentence. For the \(t\)-th input word, \(h_t\) denotes the forward hidden layer value, and \(h_t'\) denotes the reverse hidden layer value. \(h_0\) and \(h_{n+1}\) are all zero vectors, which denote the initial hidden layer values of forward and backward propagation respectively. \(n\) is the number of words in the sentence. \(o_{t}^{k}\) is the output value of the output group corresponding to the \(k\)-th focus when the \(t\)-th word is input, and represents the possibility that the \(t\)-th word is the \(k\)-th focus. \(W', V'\) and \(U'\) represent the forward propagation weights of the input layer, hidden layer and output layer respectively. \(W, V\) and \(U\) represent the backward propagation weights of the input layer, hidden layer and output layer respectively. Finally, we get all focus words by equation (1).

\[
t^{k*} = \arg \max_{1 \leq t \leq n} (o_{t}^{k})
\]  

(1)

\(t^{k*}\) represents the position of the \(k\)-th focus. The position of the maximum value in the \(k\)-th group is the position of the \(k\)-th focus. \(n\) is the total number of words in a question.

During the training phase, if the input word at moment \(t\) is the focus word of the question. Then, the expected output value at that moment is set to 1, otherwise it is set to 0. Figure 2 shows an example of training question and respective output, which is a question about flight information. There are two foci in this type of question, which are the departure place and the destination. Therefore, in the output group of the departure place, when the input word is Beijing, the expected output value is 1. Similarly, in the output group of the destination, when the input word is Xi'an, the expected output value is 1. The error between the actual output value and the expected output value is calculated and the weights are adjusted by the back propagation algorithm until convergence.
3.1.3 The incomplete focus problem. The incomplete focus problem is that a question contains only a part of focus words, not all of them. For example, the question about flight information may contain only the destination or the departure place. Figure 3 shows an example that only refers to the departure place. In this case, all the expected output values of the group corresponding to the missing focus word are set to 0 during the training phase.

In order to give full consideration to both complete and incomplete focus questions, a threshold $\tau$ is set for each output group during the execution phase. If the maximum output value of the k-th group is greater than $\tau_k$, then the question contains the focus word of the k-th group, and the word at the position of the maximum value is the focus word. Otherwise, the k-th group does not output the focus word. The threshold $\tau_k$ is calculated by equation (2), where $O_{\text{max}}^k$ represents the maximum output value at all non-focus words positions of the k-th group and $O_{\text{min}}^k$ represents the minimum output value at the focus word position of the k-th group.

$$\tau_k = \frac{O_{\text{max}}^k + O_{\text{min}}^k}{2}$$

3.2 The CAQA System Framework

Our CAQA system framework is shown in Figure 4, which includes the knowledge base construction module, question analysis module, answer reasoning module and an airport knowledge base. The question analysis module contains question classification and focus extraction to obtain the question category and focus words respectively. The answer reasoning module retrieves an answer from the knowledge base according to the question analysis results and matching rules in the knowledge base.
The knowledge base is built on rules and respective knowledge content. In order to improve reasoning efficiency, we create a knowledge file for each type of question. Each knowledge file stores all the answers to a type of question, and maps to a category information. As a result, it is easy to maintain the knowledge base. The reasoning machine uses the focus words to find the final answer in the knowledge file through the matching rules. A rule is an if-else statement, in which the condition part is a disjunctive normal form that consists of a list of foci and focus words.

4. Experiments
The experimental data used in this paper comes from Kunming airport question database. The database contains a total of 150,560 answer pairs, including 170 categories. In addition, the average length of the question is 20. The experimental environment is as follows.

Operating System: Ubuntu 16.04 LTS; Memory: 16G;
Processor: Intel Core i7-4790 CPU @ 3.6GHz x 8;

4.1 Focus Extraction Test

4.1.1 Test data and settings. In this section, we select four type questions from the question database, as shown in Table 1. The luggage and airline number questions have only one focus word. The flight information and boarding document questions have two focus words. For each type of question, we select 75% as training data, and the rest as test data. We also ensure that the focus words in test data have not occurred in training data in order to test the generalization ability of the models.

| Question category  | Questions | Focus words | Focus | Focus name                      |
|--------------------|-----------|-------------|-------|---------------------------------|
| luggage            | 62300     | 151         | 1     | luggage name                    |
| airline number     | 23700     | 241         | 1     | airline name                    |
| flight information | 15800     | 115         | 2     | departure place, destination    |
| boarding document  | 36500     | 84          | 2     | available document, unavailable document |

We set the word vector dimension to 150, the number of neurons in the input layer is set to 50. The hidden layer has 100 neurons, including 50 forward-calculated neurons and 50 backward-calculated neurons. The output layer has k (k≥1) neurons.
4.1.2 Results. This paper compares the Fk-BRNN model with the TextRank model[11], QA detector model[12], pattern matching model[13] and CRF model[15] in terms of focus extraction problem. The TextRank model is originally designed to extract key words. In this comparative experiment, the key words extracted by the TextRank model are regarded as the focus words. The window of the TextRank model is set to 10. The QA Detector model uses the Stanford Parse [20] to generate a syntax tree, and select the words with syntax tag NP | NN * as the focus words. The pattern matching model also uses Stanford Parse to generate a syntax tree, and determine the patterns based on syntactic tags. It matches the focus word according to the patterns. The CRF model uses a CRF++ tool to extract the focus words, and its window is set to 5.

Figure 5 compares the above five focus extraction models’ accuracy. Figure 6 shows their average running time that does not include the I/O time and the training time. Obviously, our Fk-BRNN model is superior to and more stable than the others, it achieves the highest near 100% accuracy on all test data. The CRF model is the second best. Since the focus words in the test data do not occur in the training data. It proves that the Fk-BRNN model has very good generalization ability. However, the CRF model is the fastest model, and the Fk-BRNN is the second fastest model. The other three models run much slower than these two.

Since some test question patterns do not appear in the training data. That leads to the main errors of the Fk-BRNN model. But we collect enough question errors, and use these new questions to train an existed Fk-BRNN model. Then the model will get an obvious performance promotion, most of the previous errors are eliminated. This shows that Fk-BRNN model has very good adaptability and migration ability, which make it more practical and suitable for online learning system.

4.2 QA system Test
This experiment uses all the questions in Kunming airport question database. Eighty percent of the questions in each type are randomly selected as training data, the rest is the test data.

The CAQA system with Fk-BRNN model is compared with a CNN-based QA model [3], a LSTM-based QA model [5] and a LSTM-CNN-based QA model [5]. The CNN model and the LSTM model are also the most common deep learning models in the current QA system. They all use an end-to-end answering strategy. Namely, the question and answer pairs are input into the model. Then the model outputs the matching degree of the question-answer pair. The final answer is the answer in the pair that has the highest matching degree. The difference among these three models lies in the sentence vector’s construction process and neural network model, i.e. the representation or encoder of
the question-answer pair.

Figure 7 compares the accuracy of the four models on the airport dataset. The experimental results show that the proposed CAQA system with Fk-BRNN model is the best, it achieves 95% accuracy on the Kunming airport question database. Since all of questions in the Kunming airport question database are short text, there is no problem of gradient disappearance or gradient explosion. So the proposed simple CAQA system is better than the complex end-to-end QA network architectures. Furthermore, our CAQA system gives the final answer by a rule-based reasoning method, which is more precise than the matching degree based method. And the knowledge rules are easy to explain and modify. However, a disadvantage is that the knowledge base needs manual maintenance.

![Figure 7. The accuracy of 4 QA models on the airport dataset](image)

5. Conclusions
This paper proposes a BRNN based multi-foci extraction model, named as Fk-BRNN model. This model can simultaneously extract multiple focus words from a question and distinguish the meaning of these focus words. The Fk-BRNN model can greatly reduce the number of training corpus, and has very good generalization ability. It can adapt to unknown words, accurately extract unknown focus words. The Fk-BRNN model has superior performance compared to the other focus extraction models. Based on the Fk-BRNN model, this paper designs and implements a Chinese Airport Question Answering system, which uses the Fk-BRNN model to extract focus words and infers the final answer from the question category and the focus words. The experimental results show that the proposed CAQA system with the Fk-BRNN model has better performance than the other three end-end neural network QA models.

6. Acknowledgments
This study is supported by the National Key Research and Development Program of China (2016YFB1000604).

7. References
[1] Feng M, Xiang B, Glass M R, Wang L and Zhou B 2016 Applying Deep Learning to Answer Selection: A Study and An Open Task. *IEEE Automatic Speech Recognition and Understanding* pp 813 - 20
[2] Yu L, Hermann K M, Blunsom and Pulman S 2014 Deep Learning for Answer Sentence Selection. *Preprint arXiv/1412.1632*
[3] Severyn A and Moschitti A 2015 Learning to Rank Short Text Pairs with Convolutional Deep Neural Networks. *ACM SIGIR* (Santiago): 373 - 382
[4] Hu B, Lu Z, Li H and Chen Q 2015 Convolutional Neural Network Architectures for Matching Natural Language Sentences. *Int Conf on Neural Information Processing Systems* (Montreal)
[5] Tan M, Santos C D, Xiang B and Zhou B 2015 LSTM-based Deep Learning Models for Non-factoid Answer Selection. *Preprint* arXiv/1511.04108

[6] Shang L, Lu Z and Li H 2015 Neural Responding Machine for Short-Text Conversation. *ACL*: 52 - 58

[7] Yin J, Jiang X, Lu Z, Shang L, Li H and Li X 2016 Neural Generative Question Answering. *IJCAI* (New York) 27: 2972 - 2978

[8] Vinyals O and Le Q 2015 A Neural Conversational Model. *Preprint* arXiv/1506.05869

[9] Woods W A 1977 Lunar rocks in natural English: explorations in natural language question answering. *Linguistic Structures Processing* pp 521 – 569

[10] Siri Team 2017 Deep Learning for Siri’s Voice: On-device Deep Mixture Density Networks for Hybrid Unit Selection Synthesis. *Apple Machine Learning Journal*

[11] Mihalcea R and Tarau P 2004 TextRank: Bringing Order into Texts. *EMNLP, Held in Conjunction with ACL* pp 404 - 11

[12] Damljanovic D, Agatonovic M and Cunningham H 2011 Identification of the Question Focus: Combining Syntactic Analysis and Ontology-based Lookup through the User Interaction. *Int Conf on Language Resources and Evaluation* (Valletta):17 - 23

[13] Bayoudhi A, Belguith L H and Ghorbel H 2015 Question focus extraction and answer passage retrieval. *IEEE Int Conf on Computer Systems and Applications*: 658 - 665

[14] Ku L W, Liang Y T and Chen H H 2007 Question Analysis and Answer Passage Retrieval for Opinion Question Answering Systems. *Computational Linguistics and Chinese Language Processing* 13(3) pp 307-25

[15] Zhang Z, Zhang Y, Liu T and Li S 2010 Automatic Recognition of Focus and Interrogative Word in Chinese Question for Classification. *Computer and Information Science* 3

[16] Graves A and Jaitly N 2014 Towards end – to - end speech recognition with recurrent neural networks. *Int Conf on Machine Learning* (Beijing):1764 -1772

[17] Devlin J, Cheng H, Fang H, Gupta S, Deng L and He X 2015 Language Models for Image Captioning: The Quirks and What Works. *Preprint* arXiv/1505.01809

[18] Mikolov T, Karafiát M, Burget L, Cernocký J and Khudanpur S 2010 Recurrent neural network based language model. *Conf of the Int Speech Communication Association* (Makuhari): 1045 -1048

[19] Bahdanau D, Cho K and Bengio Y 2014 Neural Machine Translation by Jointly Learning to Align and Translate. *Preprint* arXiv/1409.0473

[20] Dan K and Manning C D 2003 Fast Exact Inference with a Factored Model for Natural Language Parsing. *Conf on Advances in Neural Information Processing Systems*: 201 - 208