Intelligent Question Answering Model Based on CN-BiLSTM

Chengfei Li, Li Liu, Fan Jiang and Yongzhao Yu

May 15, 2018

Intelligent Question Answering Model Based on CN-BiLSTM

Chengfei Li[1], Li Liu[1,1], Fan Jiang[1],Yongzhao Yu[1]

[1]School of Information Science and Engineering, Lanzhou University, Lanzhou, China
lichf16@lzu.edu.cn

Abstract: An intelligent question-answering system can understand the user's question in the form of natural language and return a concise, accurate answer by searching the knowledge base or corpus. Compared with search engines, intelligent question-answering systems have a better understanding of users' intentions, thereby can meet users' demand for accuracy information. This paper proposes a novel hybrid model for the contextual query intelligent question-answering task. The model employs convolutional neural network and bidirectional LSTM network to improve the text information encoding capability and capture long-term dependencies of the context. The experiments on bAbi data show that the model is effective and efficient.

Keywords: intelligent question-answering, natural language processing, convolutional neural network, bidirectional LSTM network

1 Introduction

The demand for acquiring information from huge amounts of data effectively and the rise of AI have facilitated the development of intelligent question-answering(Q&A) systems. The intelligent Q&A systems are expected has the ability of understanding the user's question in the form of natural language and returning a concise, accurate answer by searching the knowledge base or corpus. In recent years, deep learning has produced an excellent performance in many areas, such as computer vision and speech recognition[1-2]. The combination of natural language processing and deep learning

[1] Corresponding author: Liu Li is with School of Information Science and Engineering, Lanzhou University, 730000, China (e-mail:liulnd@lzu.edu.cn)

has gradually become a new research trend[3]. However, the complexity and diversity of natural language and the abstraction of semantics have caused great difficulties. Natural language is essentially different from natural signals such as images and speech. Image and speech data can be directly processed by machine learning and deep learning algorithms, while natural language requires some transformation. Some researchers have tried the bag-based model to represent text[4], but these methods may produce problems such as sparse data, dimensional catastrophes and difficult-to-capture semantics. Others use the word embedding text representation algorithm[5] which has advantages over traditional text representation algorithms in grouping in the vector space according to word vector similarity, and the word vector is a low-dimensional, dense, continuous representation. The word embedding text representation algorithm and its various variants greatly promote the development of natural language processing tasks and the development of intelligent Q&A also benefits from these algorithms. The semantic representation ability of the deep learning method represented by CNN and RNN has attracted the attention of researchers[6-9]. In the field of intelligent Q&A, various deep neural network models have been proposed[10] and obtained encouraging results.

The intelligent Q&A systems obtain the semantic representation of the context, the user's question and the answer firstly, and then achieve the match between the question and the answer. This requires that the intelligent Q&A systems must make full use of the encoded semantic representation information to capture semantic matching tasks between texts. However, the existing intelligent Q&A systems still have some problems with the text information encoding and the context long-term dependency capturing. This paper proposes a new intelligent Q&A model which exploits hybrid coding based on convolutional neural network and bidirectional LSTM network(CN-BiLSTM) for high-level information coding. The one-dimensional convolutional neural network captures the text structure information, while the bidirectional LSTM(Bi-LSTM) captures the dependence relationship among structure information, and therefore the model improves the text information encoding ability and captures the long-term context dependence. We conducted experiments on the intelligent Q&A field public data set bAbi [11] and compared our model with different deep learning network models in the paper. The experimental results show that the proposed model based on CN-BiLSTM improves the accuracy of the system.

2 Semantic Representation Based on CN-BiLSTM

2.1 Introduction to the Model

The model is shown in Figure 1, where Story is the context information, Q is the user's question and A is the answer to return. This model takes Word2vec operations on the original context, user questions and answers, transforming the original natural language into a numerical form that deep neural networks can handle. The word-embedded data is sent to Merge Layer which performs dot product operations on context and user-problem data and the results are merged with the answer data. The output of Merge Layer is fed into CN-BiLSTM layer for high-level information encoding and the encoded text information is sent to Output Layer where the model uses the *softmax* function to perform multi-classification calculations to get the output probability of each answer. The model uses cross- entropy as a loss function, *Rmsprop* as an optimizer and the prediction accuracy as a measure indicator.

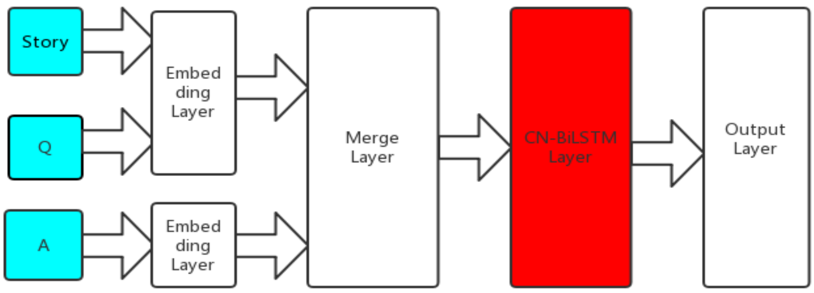

Figure 1. Structural diagram of intelligent question-answering system based on CN-BiLSTM

2.2 Semantic Representation of CN-BiLSTM

In CN-BiLSTM layer, a one-dimensional convolutional neural network is employed firstly to semantically represent sentences for capturing the structural information of it. The convolution kernel checks the text matrix for sliding scanning, performs feature extraction and feature selection through the convolution operation, and finally combines the expression vectors of sentences. In this way, a series of features are obtained at various locations, including information between words, structural information formed between adjacent words, and the like. Repeat the above operations several times and as the operation of different convolution kernels, the structural features of the sentences included in the model are increasing. The final model obtains multiple vector representations, and then these vectors are connected to obtain the high-level semantics of the entire sentence. As indicated, formula(1) is a one-dimensional convolutional formula (examples with 3 convolution kernels).

$$\mathrm{H}_t = F(X_{t-1}, X_t, X_{t+1}) = W[X_{t-1}, X_t, X_{t+1}] + b \tag{1}$$

Here, F is a nonlinear activation function, W is a parameter matrix and b is biases (bias). The semantic representation done by the convolutional neural network is sent to a BatchNormal layer , and its data is normalized.

Next, the coded information passing through the CN layer is sent to the Bi-LSTM network. Essentially, the Bi-LSTM is the integration and screening of information in chronological order. LSTM is a special network structure with "forgotten gate", "input gate" and "output gate". LSTM relies on the structure of some "gates" to allow information to selectively affect the state of each moment in the recurrent neural network. The specific formulas are as follows:

$$\mathrm{C}_t = f_t \times C_{t-1} + \mathrm{i}_t \times \widehat{C}_t \tag{2}$$

$$f_t = \sigma(W_f \bullet [h_{t-1}, x_t] + b_f) \tag{3}$$

$$i_t = \sigma(W_i \bullet [h_{t-1}, x_t] + b_i) \tag{4}$$

$$\widehat{C}_t = \tanh(W_c \bullet [h_{t-1}, x_t] + b_c) \tag{5}$$

Here, f_t, i_t represents forgetting door and input door respectively. At each moment, the forgotten gate controls the degree of forgetting of the previous moment; the input gate controls the extent to which new memories are written into long-term memory. The value of sigmoid function is in [0,1]; tanh function is in [-1,1]. *Ct-1* is the state at time *t-1*, and *Ct* is the state at time *t*.

$$h_t = o_t \times \tanh(C_t) \tag{6}$$

$$o_t = \sigma(W_o \bullet [h_{t-1}, x_t] + b_o) \tag{7}$$

Here, O_t is the output gate, which controls how short-term memory is affected by long-term memory. H_t is the output of t. The Bi-LSTM is based on the structure of the LSTM. Two intermediate values are stored. A participates in the forward calculation and A' participates in the reverse calculation. The final output value h depends on A and A'. In the forward calculation, the Ct of the hidden layer is related to $Ct-1$. In the reverse calculation, the Ct of the hidden layer is related to $Ct+1$:

$$C_t = F(Ux_t + WC_{t-1}) \quad \textbf{(8)}$$

$$C'_t = F(U'x_t + W'C_{t+1}) \quad \textbf{(9)}$$

$$h_t = g(VC_t + V'C'_{t+1}) \quad \textbf{(10)}$$

Here, g is the output activation function U, V, W, and U', V', W' represent the parameter matrix for forward calculation and backward calculation respectively. Bi-LSTM semantic representation modeling is a sequence of word structure vectors completed by each convolution kernel of a one-dimensional convolutional network. Each position has an intermediate representation, which represents the beginning of the clause to this position, the semantics of the end of the sentence to this location. Here we assume that the intermediate representation of each position is determined by the word structure vector of the current position, the intermediate representation of the previous position, and the intermediate representation of the following position. After passing through a Bi-LSTM , the result is input to the Output Layer.

3 Experiments and Results Analysis

bAbi dataset is divided into a training set and a test set and contains a total of 10 categories of tasks. Each sample of it contains contexts, questions and answers. The experimental environment of this paper is a deep learning framework using Keras and TensorFlow under Ubuntu 16.04 system. The CPU model is Intel Xeon E5-2683 v3 and the GPU model is NVIDIA TITAN X (Pascal).

The accuracy of the model on the training set reaches 99.3% and the accuracy on the test set reaches 94% under the iterative 120 rounds. The model effect is shown in Figure 2.

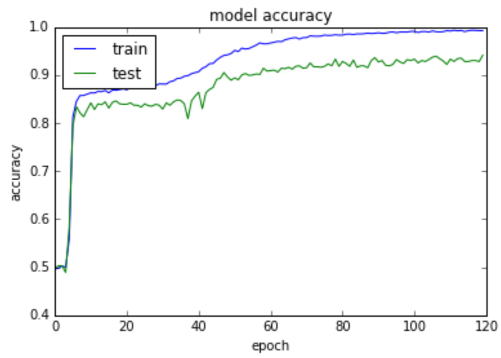

Figure 2. The result based on CN-BiLSTM model

The similar network models with different semantic representation layers are used for comparison experiments. The network models are shown in Figure 3.

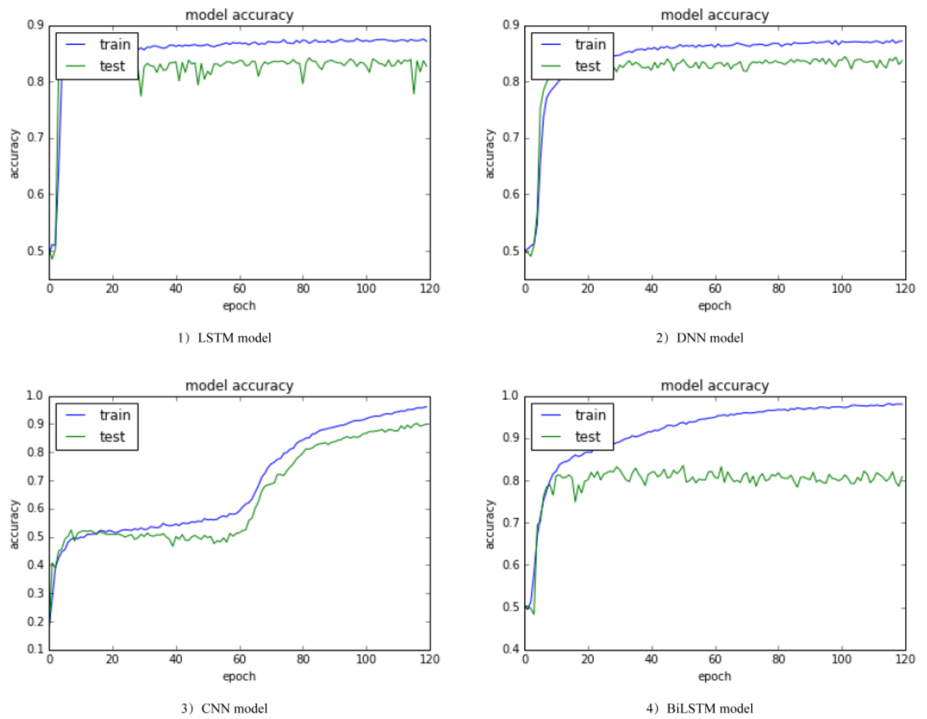

Figure 3. The comparison models

The result of each model is shown in Figure 4.

Figure 4. The effects of the different experimental models.

The final experimental results on the training set and the test set are shown in Table 1.

Table 1. Comparison with the accuracy rate of different models.

Model	Training set accuracy	Test set accuracy
DNN	0.8712	0.8270
CNN	0.8715	0.8370
LSTM	0.9773	0.9
BiLSTM	0.99	0.813
CN-BiLSTM	0.993	0.942

Through the accuracy rate change graphs and the final experimental results on the training set and the test set, it can be known that the final result of DNN model is poor and the accuracy on the test set is not stable. In addition, DNN model requires a large number of parameters, which may cause problems such as time-consuming training and large models. CNN model converges faster and the model training takes less time because the convolution operation shares a large number of parameters. The final accuracy is not high as the volume and operation only involve the internal structural information of the sentence, however, the sequence information convolution operation cannot be extracted. Compared with the former two networks, LSTM network can get a better result, but the convergence speed is slow. BiLSTM network produces good results on the training set, but the test set had poor results, indicating that the model had severe overfitting. CN-BiLSTM network balances model training resources and improves the accuracy. CN-BiLSTM model structure conforms to the human's own way of acquiring textual information. That is, when reading, the structure of textual information is the focus of attention; when carefully understanding textual semantics, sequence information is even more important.

4 Conclusion

CN-BiLSTM model improves the text information encoding capability, captures the long-term dependencies of the context well. The experimental results show that our model improves the accuracy and demonstrate the feasibility of the deep learning model in intelligent Q&A applications.

References

1. He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[J]. 2015:770-778
2. Jing L. Research on speech recognition based on deep learning[D]. Beijing University of Posts and Telecommunications, 2014
3. Socher R, Bengio Y, Manning C D. Deep learning for NLP (without magic)[C]// Tutorial Abstracts of Acl. 2015:5-5
4. Wallach H M. Topic modeling: beyond bag-of-words[C]// International Conference on Machine Learning. ACM, 2006:977-984
5. Boom C D, Canneyt S V, Demeester T, et al. Representation learning for very short texts using weighted word embedding aggregation[J]. Pattern Recognition Letters, 2016, 80(C):150-156.
6. Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11):2278-2324
7. Razavian A S, Azizpour H, Sullivan J, et al. CNN Features Off-the-Shelf: An Astounding Baseline for Recognition[C]// IEEE Conference on Computer Vision and Pattern Recognition Workshops. IEEE Computer Society, 2014:512-519
8. Meng J E, Zhang Y, Wang N, et al. Attention pooling-based convolutional neural network for sentence modelling[J]. Information Sciences, 2016, 373(C):388-403.
9. Filippova K, Alfonseca E, Colmenares C A, et al. Sentence Compression by Deletion with LSTMs[C]// Conference on Empirical Methods in Natural Language Processing. 2015:360-368.
10. Guanghui R, Zhenhua H. Question-based matching method based on deep learning[J]. Journal of Computer Applications, 2017, 37(10):2861-2865
11. Weston J, Bordes A, Chopra S, et al. Towards AI-Complete Question Answering: A Set of Prerequisite Toy Tasks[J].2015, arXiv preprint arXiv:1502.05698