Answer Selection Using Multi-Layer Semantic Representation Learning

Yuan Yuan and Lei Chen*
Institute of Intelligent Machines, Chinese Academy of Sciences, Hefei, China

*Corresponding author e-mail: chenlei@iim.ac.cn

Abstract. Answer selection is the most important module of question answering system. Mining the semantic relevance between questions and answers is the key point of answer selection. However, the answer texts often contain some colloquial expressions and redundant information in some fields. In these cases, some traditional methods of mining text features are complex and the performance improvement is limited. In this paper we investigate using deep learning to enhance the computability of semantics and eliminate the semantic gap for the answer selection module of non-factual question answering system. By integrating different networks to mine the association semantic between questions and answers, a novel multi-layer answer selection model is proposed. The shallow semantic relevance between questions and answers is mined by adding the attention mechanism to CNN, and then the features are input into Bi-LSTM model for further mining to obtain more efficient semantic representation. The experimental results show that compared with other models, our method can achieve the better accuracy in the task of answer selection.

1. Introduction

As a typical representative of representation learning, deep learning has made breakthroughs in many researches and applications. In the study of answer selection, representation learning has also gained a lot of attention. As the most important module of question answering system (QA), answer selection returns the correct answers according to the semantic relevance between the questions and answers, based on a candidate set of questions and answers. Hence, mining the semantics of questions and answers to find the semantic relevance between them is the key point of answer selection [1]. Answer selection task is regarded as a semantic matching process in representation learning. Specifically, natural language questions and answers are transformed into numerical vectors in low-dimensional semantic space first. And then, semantic features are automatic learned and extracted to enhance the computability of semantics so as to eliminate the semantic gap. Finally, the similarity between semantics is calculated to select the required answers.

Recent years many related works have been proposed. Yu et al. [2] proposed using convolution and pooling operations of CNN to mine text semantics. The position information of words was not taken into account when extracting features of n-gram language model, so text preprocessing was not needed in this method. Zhou et al. [3,4] encoded each question-answer pair and processed the encoded sentence vector using RNN to obtain sequence information, with the output of each moment being the final category of the answers, while the context information can be retained. Wang and Nyberg [5] proposed to overlay two layers of Bi-LSTM to mine the semantics of text, where the output of the former layer was taken as the input of the next layer to extract the high-dimensional semantic features.
Wang et al. [6] analyzed the influence of attention mechanism on semantic representation and proposed three improvements of LSTM that integrates attention mechanism to solve the problem of weight bias in the attention mechanism of single LSTM network. In order to eliminate ambiguity better, Miao et al. [7] proposed a potential random attention mechanism, which used a potential vector to represent the question vectors in the weight multiplication phase of the LSTM that integrates attention mechanism and then the potential vector was used to calculate a weight for the output of each time point of the LSTM in the answer text. Cheng et al. [8] changed the information storage vector in LSTM from single to multiple and stored the output information of all time points, so as to solve the problem of information loss if the information was stored in the fixed length carrier. Ma et al. [9] proposed a hybrid answer selection model for non-factoid QA, where the irrelevant information is removed and better representations are generated for questions and answers.

Meanwhile, the answer selection method based on deep learning also has its shortcomings [10]. The semantic representation model based on neural network often processes the input text uniformly, and cannot recognize the noise information in the answer text. Moreover, a single network structure can only mine the same semantic features of text, which makes the final generated semantic vectors unable to express semantic information efficiently. In addition, state-of-the-art networks often use complex architectures and take a long time to train the model. In order to solve these problems, some novel works have been proposed in the last two years. Instead of focusing on architecture engineering, Nicosia and Moschitti [11] took advantage of small amounts of labelled data that model semantic phenomena in text to encode matching features directly in the word representations. Kratzwald and Feuerriegel [12] proposed an adaptive document retrieval model to deal with the problem that choosing a static number of documents may suffer from a noise-information trade-off. Zhu et al. [13] introduced a pair-to-sequence model for unanswerable question generation, which effectively captured the interactions between the question and the paragraph.

This paper focuses on the answer selection task of non-factual QA. In this kind of QA, question-answer pairs are independent of each other. Meanwhile, answer texts often contain colloquial expressions and irrelevant information. Therefore, the key point of the answer selection task in non-fact QA is how to efficiently represent and learn the semantics of the questions and answers. Some previous methods generally pre-extracted the abstract of the answer text, which may lose some key information while improving the accuracy of answer selection. Besides, the process of mining text features is complex and the improvement of system performance is limited. Inspired by the work of Zhou et al. [14], we propose an answer selection model using multi-layer semantic representation learning to deeply mine the semantic relevance between questions and answers, while retaining the original answer information. Firstly, a CNN with attention mechanism called Attention-CNN is used to extract the semantic features between adjacent words in questions and answers. This Attention-CNN can also find out the semantic features that have great influence on the final results. And then the feature vectors are input into the Bi-LSTM model, which can not only capture the semantic changes caused by different word positions, but also store the interaction between context information. This multi-layer networks can efficiently mine the semantic features of text and find out the semantic relationship between questions and correct answers.

2. Multi-Layer Semantic Representation Learning

2.1. Semantic representation model based on Attention-CNN

In the process of semantic representation learning based on CNN, the generation of semantic vectors for questions and answers is independent of each other. There is no information interaction in the whole process of semantic mining. In this paper, the attention mechanism is added to the answer selection model to mine the semantic relevance between the questions and the answers, and give higher weight to the key information in order to improve the semantic expression ability of the answer selection model. As shown in Fig. 1, we add the attention mechanism after convolution operation in CNN [15]. After performing the convolution operation of the original input layer, the features can be extracted. According to the attention matrix, which is the weight matrix of semantic features, the attention vectors of questions and answers can be obtained by summing the rows and columns of the
attention matrix respectively. And then these two attention vectors are used to guide the pooling operation according to the following formula,

\[ P_{k,i} = \sum_{j=0}^{l} c_{k,j} \alpha_{i,j} \]  

(1)

where \( c \) is the feature set extracted from convolution layer, \( a \) is the attention matrix, and \( P \) is the result of pooling. After this operation, higher weights are given to the features that have greater impact on the results of semantic matching, and smaller weights are given to the extracted irrelevant features, so as to reduce their impact on the results. By introducing the attention mechanism into CNN, we can enhance the interaction of underlying features in convolution and select the features that have great influence on the results of semantic matching.

![Figure 1. Adding attention mechanism to CNN.](image)

2.2. Semantic representation model based on Bi-LSTM

The semantic representation model based on LSTM can preserve sequence information. However, it does not pay enough attention to the interaction within sentences. Unfortunately, this key information often contributes more to understanding the semantics of the whole text. And when the state of the hidden layer changes, this model only considers the impact of the saved information on the current input, while ignoring the impact of the following information in the sequence on the above information [16, 17]. Therefore, we adopt Bi-LSTM to improve the model in this paper. As shown in Fig. 2, Bi-LSTM consists of two LSTM networks in the opposite directions [18].

![Figure 2. The structure of Bi-LSTM.](image)

In the representation learning of semantics, the forward LSTM network takes the normal text sequence information as input, and the state of its hidden layer preserves the above information. Meanwhile, the backward LSTM network takes the sequential text in reverse order as input, and the
state of its hidden layer preserves the following information of the original sequence. The final output of the model is the splicing of the two reverse LSTM outputs.

2.3. Integration of Attention-CNN and Bi-LSTM

Based on the above discussion, we integrate Attention-CNN with Bi-LSTM to construct the multi-layer answer selection model. As shown in Fig. 3, the model first uses Attention-CNN to obtain the shallow semantic features of questions and answers, which are next taken as the input of Bi-LSTM for deeper semantic mining so as to get the vector representation of answer texts. Finally, the correct answer is selected by calculating the similarity between the question vectors and the answer vectors.

In the last step of Fig. 3, we use the cosine similarity of vectors, calculated by the following formula, to measure the semantic matching between questions and answers, where \( v_q \) and \( v_a \) are the vectors of question and answer respectively, and \( S_{q,a} \) is the semantic similarity between \( v_q \) and \( v_a \).

\[
S_{q,a} = \cos(v_q, v_a) = \frac{v_q \cdot v_a}{\|v_q\| \cdot \|v_a\|}
\]  

(2)

In the process of model training, the input consists of three texts, that is, question \( q \), correct answer \( a' \) and wrong answer \( a'' \). This training makes the semantic similarity between the question \( q \) and the correct answer \( a' \) higher than that between the question \( q \) and the wrong answer \( a'' \). The loss function is defined as the following formula:

\[
L = \max \{0, m - S_{q,a'} + S_{q,a''}\}
\]

(3)

where \( S_{q,a'} \) is the semantic similarity between question and correct answer, \( S_{q,a''} \) is the semantic similarity between question and wrong answer, and \( m \) is a threshold used to limit the cosine distance between them. When the obtained semantic similarity does not meet the requirements, the network parameters are updated until the loss function is satisfied.

3. Experiments

The experimental data in this paper is collected from some QA community of agricultural professional websites, including 10,000 high quality agricultural question and answer pairs, which are divided into 8,000 pairs of training set, 1,000 pairs of validation set and 1,000 pairs of test set. The full domain corpus of Chinese Wikipedia is used to train word vectors, with the dimension of 100. The Chinese word segmentation module Jieba [19] is adopted to process the text. In the training of CNN model, the number of input layers is set to 200 to ensure that most of the answer information is retained. The activation function and the gradient descent method are \( tanh \) and Adam respectively. The experiment shows that the best result can be achieved when the batch size is 128 and the learning rate...
is 0.001. Since the effect of Dropout is not obvious in this experiment, Dropout is not used in the training of CNN to prevent over-fitting. In the training of LSTM model, the parameters setting is the same as that of the single-layer CNN. Besides, the number of iterations is set to 1000 and Dropout of LSTM is set to 0.5. This experiment takes CNN-based model and LSTM-based model as the benchmark models. And many other models are compared. The size of the candidate answer pool is set to 100 in the experiments. The accuracy, calculated as follows, is used to evaluate of the experimental results,

\[
\text{Accuracy} = \frac{\text{Number of questions answered correctly}}{\text{Total number of questions in the test set}}
\]

The experimental results are given in Table 1. CNN and LSTM are the baseline models. “Attention-CNN” means adding attention mechanism to CNN and weighting different features after convolution. “LSTM + attention” indicates adding attention mechanism to unidirectional LSTM. “Bi-LSTM + Attention” denotes adding attention mechanism to Bi-LSTM model. The rest are various combination of different models, where “Attention-CNN + Bi-LSTM” is the proposed model.

| Model                  | Accuracy (%) |
|------------------------|--------------|
| CNN (baseline)         | 48.6         |
| LSTM (baseline)        | 51.2         |
| Bi-LSTM                | 53.1         |
| Attention-CNN          | 52.6         |
| LSTM + attention       | 57.4         |
| Bi-LSTM + attention    | 58.3         |
| CNN + LSTM             | 60.5         |
| CNN + Bi-LSTM          | 61.4         |
| Attention-CNN + LSTM   | 63.8         |
| Attention-CNN + Bi-LSTM| **66.2**     |

According to the experimental results in Table 1, the introduction of attention mechanism into a single neural network model can improve some performance, and the integrations of different models are better than any single model. The integration model of Attention-CNN and Bi-LSTM achieves the highest accuracy of 66.2%, which is 17.6 percentage points and 15 percentage points higher than the baseline system of CNN and LSTM, respectively. Concretely, introducing the attention mechanism can enhance the accuracy of answer selection in non-factual QA. The result of each single model with attention mechanism is better than that of the original model, respectively increased by 4, 6.2 and 5.2 percentage points of CNN, LSTM and Bi-LSTM. Meanwhile, the performance of Bi-LSTM is better than unidirectional LSTM under the same conditions. This is because the length of answer is often long in non-fact QA. The unidirectional LSTM can only process the current input with the saved information above, which leads to the problem of information weakening. While Bi-LSTM model can process the input data in two opposite directions, so it can solve the problem of information weakening. Besides, answer selection model based on the integration of models can deeply mine semantics through different network structures. So that it can better mine the related information and generate efficient semantic representation. The Attention-CNN model can mine the association information of adjacent words. The Bi-LSTM model can better represent the interaction between long text sequences. Therefore, integrating these two models can compose a more efficient text semantic representation learning model, which achieves the best results in the answer selection experiments.

4. Conclusion

This paper proposes an answer selection model using multi-layer semantic representation learning. In order to extract the local features of sentences and acquire the keyword information, the attention mechanism is added to CNN model, which can assign corresponding weights to the input vectors
According to their importance in the process of semantic representation learning. Meanwhile, due to the structure limitation of the unidirectional LSTM network itself, we introduce the Bi-LSTM model to solve the problem of information insufficient. Finally, these two models are integrated for semantic representation learning in the process of answer selection.

The work in this paper is still preliminary. And the experimental dataset is also inadequate. In next work, the method of multi-layer semantic representation learning needs to be better exploited and applied. And some comparisons with other open set methods will also be considered.

5. Acknowledgments

This work is supported by the National Natural Science Foundation of China (31771677, 31601268).

References

[1] Z. Liu, M. Sun, Y. Lin, and R. Xie, Knowledge representation learning: A review (in Chinese), Journal of Computer Research and Development, vol. 53, no. 2, 2016, pp.247-261.

[2] L. Yu, K. M. Hermann, P. Blunsom, and S. Pulman, Deep learning for answer sentence selection, CoRR, vol. bs/1412.1632, 2014.

[3] X. Zhou, B. Hu, Q. Chen, B. Tang, and X. Wang, Answer sequence learning with neural networks for answer selection in community question answering, in ACL 2015, Beijing, China, Volume 2: Short Papers, pp. 713-718.

[4] X. Zhou, B. Hu, J. Lin, Y. Xiang, and X. Wang, ICRC-HIT: A deep learning based comment sequence labeling system for answer selection challenge, in NAACL-HLT 2015, Denver, Colorado, USA, pp. 210-214.

[5] D. Wang and E. Nyberg, A long short-term memory model for answer sentence selection in question answering, in ACL 2015, Beijing, China, Volume 2: Short Papers, pp. 707-712.

[6] B. Wang, K. Liu, and J. Zhao, Inner attention based recurrent neural networks for answer selection, in ACL 2016, Berlin, Germany, Volume 1: Long Papers, pp. 1288-1297.

[7] Y. Miao, L. Yu, and P. Blunsom, Neural variational inference for text processing, in ICML 2016, New York City, USA, pp. 1727-1736.

[8] J. Cheng, L. Dong, and M. Lapata, Long short-term memory-networks for machine reading, in EMNLP 2016, Austin, Texas, USA, pp. 551-561.

[9] R. Ma, J. Zhang, M. Li, L. Chen, and J. Gao, Hybrid answer selection model for non-factoid question answering, in IALP 2017, Singapore, pp. 371-373.

[10] T. M. Lai, T. Bui, and S. Li, A review on deep learning techniques applied to answer selection, in COLING 2018, Santa Fe, New Mexico, USA, pp. 2132-2144.

[11] M. Nicosia and A. Moschitti, Semantic linking in convolutional neural networks for answer sentence selection, in EMNLP 2018, Brussels, Belgium, pp. 1070-1076.

[12] B. Kratzwald and S. Feuerriegel, Adaptive document retrieval for deep question answering, in EMNLP 2018, Brussels, Belgium, pp. 576-581.

[13] H. Zhu, L. Dong, F. Wei, W. Wang, B. Qin, and T. Liu, Learning to ask unanswerable questions for machine reading comprehension, in ACL 2019, Florence, Italy, Volume 1: Long Papers, pp. 4238-4248.

[14] P. Zhou, W. Shi, J. Tian, Z. Qi, B. Li, H. Hao, and B. Xu, Attention-based bidirectional long short-term memory networks for relation classification, in ACL 2016, Berlin, Germany, Volume 2: Short Papers, pp. 207-212.

[15] W. Yin, H. Schütze, B. Xiang, and B. Zhou, ABCNN: attention-based convolutional neural network for modeling sentence pairs, TACL, vol. 4, 2016, pp. 259-272.

[16] Y. Wang, M. Huang, X. Zhu, and L. Zhao, Attention-based LSTM for aspect-level sentiment classification, in EMNLP 2016, Austin, Texas, USA, pp. 606-615.

[17] J. Zeng, X. Ma, and K. Zhou, Enhancing attention-based LSTM with position context for aspect-level sentiment classification," IEEE Access, vol. 7, 2019, pp. 20462-20471.

[18] Y. Liu, C. Sun, L. Lin, and X. Wang, Learning natural language inference using bidirectional LSTM model and inner-attention, CoRR, vol. abs/1605.09090, 2016.

[19] Chinese word segmentation component of python: Jieba, https://pypi.org/project/jieba/