Hindawi Wireless Communications and Mobile Computing Volume 2018, Article ID 2678976, 8 pages https://doi.org/10.1155/2018/2678976



## Research Article

# **Using Sentence-Level Neural Network Models for Multiple-Choice Reading Comprehension Tasks**

# Yuanlong Wang D, Ru Li, 1,2 Hu Zhang, Hongyan Tan, and Qinghua Chai 3

<sup>1</sup>School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China

Correspondence should be addressed to Yuanlong Wang; ylwang@sxu.edu.cn

Received 28 March 2018; Accepted 13 June 2018; Published 3 July 2018

Academic Editor: Tianyong Hao

Copyright © 2018 Yuanlong Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Comprehending unstructured text is a challenging task for machines because it involves understanding texts and answering questions. In this paper, we study the multiple-choice task for reading comprehension based on MC Test datasets and Chinese reading comprehension datasets, among which Chinese reading comprehension datasets which are built by ourselves. Observing the above-mentioned training sets, we find that "sentence comprehension" is more important than "word comprehension" in multiple-choice task, and therefore we propose sentence-level neural network models. Our model firstly uses LSTM network and a composition model to learn compositional vector representation for sentences and then trains a sentence-level attention model for obtaining the sentence-level attention between the sentence embedding in documents and the optional sentences embedding by dot product. Finally, a consensus attention is gained by merging individual attention with the merging function. Experimental results show that our model outperforms various state-of-the-art baselines significantly for both the multiple-choice reading comprehension datasets.

#### 1. Introduction

Reading comprehension is the ability of reading texts, understanding their meanings, and answering questions. When machines are required to comprehend texts, they need to understand unstructured text and do reasoning based on the text [1–3]. It is a major task in the field of natural language processing and machine learning.

Recently, machine reading comprehension (MC) is increasingly drawing attention and several large reading comprehension datasets have also been released. For the several released datasets, the task is getting more and more difficult (from CNN/Daily Mail datasets to SQuAD and then to TriviaQA) as system performance has rapidly improved with each new released datasets. The CNN/Daily Mail datasets [4] is a cloze-style reading comprehension task, which aims to comprehend a given document and then to answer questions based on the given document, and the answer to each question is a single word inside of the document. The SQuAD [5]

is a question-answering reading comprehension task, which further constrains answers often including nonentities and being much longer phrases to be a continuous subspan of the document. Clearly, the question-answering task is more difficult than the cloze-style task. The TriviaQA [6] is also a question-answering reading comprehension task, but the task in TriviaQA is more difficult than the task in SQuAD because answers in TriviaQA are independent of the evidence and belong to a diverse set of types.

Different from the above, the task based on the MCTest datasets [3] is a multiple-choice reading comprehension, each example of which consists of one document and four associated questions and each question gives four candidate answers and only one answer is correct among them. In this paper, we focus on such problem of answering multiple-choice questions in documents, and, at the same time, we also release a Chinese reading comprehension dataset for such multiple-choice task. To our knowledge, the dataset is the first Chinese reading comprehension dataset of this

<sup>&</sup>lt;sup>2</sup>Key Laboratory of Computation Intelligence and Chinese Information Processing of Ministry of Education, Shanxi University, Taiyuan 030006, China

<sup>&</sup>lt;sup>3</sup>School of Foreign Languages, Shanxi University, Taiyuan 030006, China

#### Document:

"Ruins" is a derogatory term that it is irrelevant to cultural and aesthetic in many Chinese mind, and interpretation of the word "ruins" is only a "city and village are changed into desolate places by destruction or natural disasters" in the "Modern Chinese Dictionary"; There is no fault for the interpretation, but it is not enough if it is measured by world knowledge. In Europe, the meaning of "ruins" has been enriched and expanded since modern times. It has been endowed with the connotation of culture and aesthetics, and has become an academic concept. The of meaning of the "ruins" is changed from the Renaissance in Europe.

Please choice two incorrect options according to the content of the document:

#### Chaica

- A. One of the purposes of this paper is to correct the misunderstanding of the term "ruins" in the modern Chinese dictionary. B. The Great Wall Ruins have condensed the vicissitudes of time in China and it have a "perception of the intoxicated" as the Acropolis ruins.
- C. Remains of the ruins often reveals the extraordinary wisdom and great efforts of the predecessors, which bring to the future generations with the shock and resonance of the soul.
- D. Awareness of the ruins is related to the aesthetic consciousness of countrymen, but also it is conducive to the popularity of the "repair the old as the old".
- E. This paper not only contains historical interest, but also infiltrated the concern of reality, and express the author's desire to enhance the cultural quality of the nation.

Box 1: Example for the multiple-choice reading comprehension for literature (the original data is in Chinese, we translate this sample in English for clarity).

kind and is even more complex than MCTest datasets. The example of such dataset consists of one document and one associated question which gives five candidate answers. The specific details of this dataset are in Section 2. Frankly, the **multiple-choice reading comprehension** task remains quite challenging. For one thing, answers in the form of an optional sentence usually do not appear in the document; for another, finding the correct answer of the given question requires reasoning across multiple sentences. Hence, sentence comprehension is more important than word comprehension in the task of the multiple-choice reading comprehension.

To carry out the task of sentence comprehension, we propose a sentence-level attention model primarily inspired by the attention model for the Cloze-style reading comprehension [7, 8]. However, unlike the Cloze-style attention model, answers to multiple-choice questions are optional sentences. Karl et al. [9] train an encoder-decoder model to encode a sentence into a fixed length vector and subsequently decode both the following sentences. They also demonstrate that the low-dimensional vector embeddings are useful for other tasks. Pichotta et al. [10] present a sentence-level LSTM language model for script inference. The results show that the model is useful for predicting missing information in text. Similar to the above model, we also present a sentence representation model which uses LSTM network to learn vector representation for sentences. Moreover, we use sentence composition model to represent sentence vector because the model can express hierarchical sentences from words to phrases, and to sentences. In order to retain more information about two kinds of sentences representation model, we employ connection method to compose the final sentence vector. Then, we train a sentence attention model between optional sentences and sentences in the document. The machine is able to learn the relationships between the document and optional sentences by the attention-based neural network.

Experimental results show that our approach can effectively improve the performance of the task of multiple-choice reading comprehension. In the following text, Chinese reading comprehension datasets, related work, details of our model, and experiments will be described, and, afterwards, our experiments will be analyzed.

#### 2. Chinese Reading Comprehension Datasets

In this paper, we focus on the multiple-choice reading comprehension task. Similar to the MCTest datasets, each example consists of one document and one associated questions. And each question gives five candidate answers. However, the dataset is more complex than MCTest datasets, and it is a literary reading comprehension dataset from test materials of final exam in senior high school. Box 1 shows an example of Chinese reading comprehension datasets.

For the dataset, the description of questions is basically fixed, as in the following: "Question". Therefore, the role of question is ignored in the Chinese reading comprehension task. The goal of the task is to understand the individual document and to select the most consistent options with the meaning of the document. Thus the Chinese reading comprehension can be described as a triple:

$$\langle D, C, A \rangle$$
 (1)

where D is the document, C denotes the choice, and A is a set in which each element is marked as 0 or 1 according to the document meaning (if the option is consistent with the document meaning, it is labeled as 1; otherwise it is labeled as 0). The A can be described as the following:

Question: "Please choose two incorrect options according to the content of the document:"

Answer: C E

A(Answer label): (11010)

TABLE 1: Statistics of multiple-choice reading comprehension datasets: train and three tests.

|            | Documents | Sentences | options |
|------------|-----------|-----------|---------|
| Train      | 769       | 28235     | 3845    |
| BCEETest   | 13        | 548       | 65      |
| SBCEETest1 | 12        | 517       | 60      |
| SBCEETest2 | 52        | 2056      | 260     |

In the training stage, we choose a 769-literary-reading-comprehension dataset which is collected from test materials of final exam in senior high school. In the testing stage, the dataset includes three parts: 13 Beijing college entrance examination papers (BCEETest), 12 simulation materials (SBCEETest1) which is provided by iFLYTEK company, and 52 test materials of final exam in Beijing senior high school (SBCEETest2). All of datasets are collected by the Chinese information processing group of Shanxi University. The statistics of training and testing data are shown in Table 1.

#### 3. Related Work

Machine comprehension is currently a hot topic within the machine learning community. In this section we will focus on the best-performing models applied to MCTest and CNN/Daily Mail according to two kinds of reading comprehension tasks.

3.1. Multiple-Choice Reading Comprehension. Existing models are mostly based on manually engineered features for MCTest [11–13]. These engineered feature models are extremely effective. However, this research often requires significant effort on the auxiliary tools to extract the feature and its capacity for generalization is limited.

Yin et al.[14] proposed a hierarchical attention-based convolutional neural network for multiple-choice reading comprehension task. The model considers multiple levels of granularity, from word to sentence level and then from sentence to snippet level. This model performs poorly on MCTest. A possible reason that can explain this is that the dataset is sparse. However, neural model can address the extracted features problem, so it appeals to increasing interest in multiple-choice reading comprehension task. For sequence data, the recurrent neural network often is used. So we propose a recurrent neural network model for the multiple-choice reading comprehension. Our model uses the bidirectional LSTM to get contextual representations of the sentence.

3.2. Cloze-Style Reading Comprehension. Hermann et al. [4] published the CNN/Daily Mail news corpus, where the content is formed by news articles and its summarization. Also, Cui et al. [7] released HFL-RC PD&CFT for Chinese reading comprehension datasets, which includes People Daily news datasets and Children's Fairy tale datasets. On these datasets, many neural network models have been proposed for the Cloze-style reading comprehension tasks. Hermann et al. [4]

proposed the attentive and impatient readers. The attentive reader uses bidirectional document and query encoders to compute an attention and the impatient reader computes attention over the document after reading every word of the query. Chen et al. [1] proposed a new neural network architecture for the Cloze-style reading comprehension. In contrast to the attentive reader, the attention weights of the model are computed with a bilinear term instead of simple dot product. Kadlec et al. [15] proposed the Attention Sum Reader, which uses attention to directly pick the answer from the context. The model uses attention as pointer over discrete tokens in the context document and then directly sums the word attention across all the occurrences. Cui et al. [7] presented the consensus attention-based neural network, namely, Consensus Attention Sum Reader, and released Chinese reading comprehension datasets. The model computes an attention to every time slice of query and makes a consensus attention among different steps. Cui et al. [8] also proposed the attention-over-attention neural network, namely, Attention-over-Attention Reader. The model presents an attention mechanism that places another attention over the primary attention, to indicate the "importance" of each attention. Dhingra et al.[16] proposed the gatedattention readers for text comprehension. The model integrates a multihop architecture with an attention mechanism which is based on multiplicative interactions between the query embedding and the intermediate states of a recurrent neural network document reader.

To summarize, all of them are attention-based RNN models which have been shown to be extremely effective for the word-level task. At each time-step, these models take a word as input, update a hidden state vector, and predict the answer. In this paper, we propose sentence-level attention model for the multiple-choice reading comprehension. Our work is primarily inspired by the attention model for the Cloze-style reading comprehension.

### 4. Sentence-Level Neural Network Reader

In this section, we will introduce our sentence-level neural network models for the multiple-choice reading comprehension task, namely, Sentence-Level Attention Reader. Our model is primarily motivated by that of Cui et al. [7], which aims to directly estimate the answer of optional sentence from the sentence-level attention instead of calculating the answer of entity from the word-level attention. The level structure of our model is shown in Figure 1. Firstly, the document is divided into several sentences  $D = \{s_1, s_2, \dots s_n\}$  and the sentence embedding is computed by embedding layer. Secondly, we use the bidirectional LSTM to get contextual representations of the sentence, in which the representation of each sentence is formed by concatenating the forward and backward hidden states. Thirdly, the sentence-level attention is computed by a dot product between the sentence embedding in the document and the optional embedding. Finally, the individual attention is merged to a consensus attention by the merging function. The following will give a formal description of our proposed model.

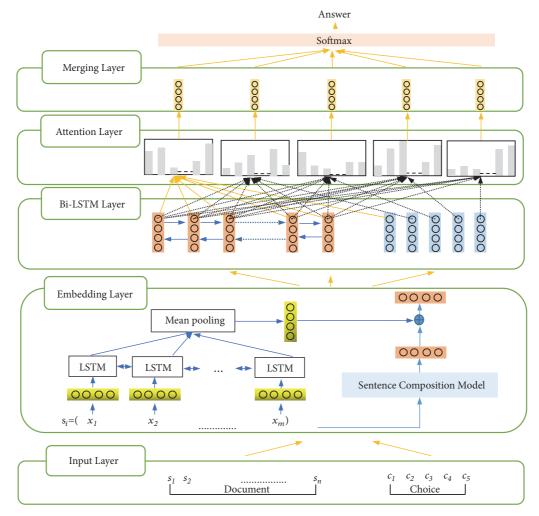


FIGURE 1: Sentence-level attention neural network. The network includes 5 layers: input layer, embedding layer, Bi-LSTM layer, attention layer, and merging layer.

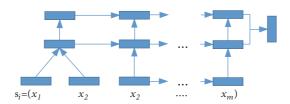


FIGURE 2: Sentence composition model.

4.1. Sentence Representation. The input of our model is the sentences in the document and options, and each sentence consists of word sequence. The sentence is translated into sentence embedding by embedding layer, which is composed of LSTM sentence model and sentence composition model[17] as illustrated in the embedding layer of Figure 1. The LSTM sentence model is a single bi-LSTM layer followed by an average pooling layer. The bi-LSTM layer is used to get the contextual representations of words and the average pooling layer is used to merge word vectors into sentence vectors. On the other hand, we used the sentence composition model to compose sentence vector. The sentence vector is

combined by the trained neural network model, which is trained by the triple consisting of single words and phrases vector (as triple( $w_1, w_2, p$ )). The sentence composition model is illustrated in Figure 2. We denote  $p_i$  as the final sentence vector. In order to retain more information about two kinds of sentences' representation model, we employ a multilayer neural network to compose the final sentence vector,  $p_i(sp_1, sp_2) = sp_1^T M sp_2$ , where  $sp_1$  is the sentence vector for LSTM sentence model,  $sp_2$  is the sentence vector for sentence composition model, and M is a parameter matrix.

In addition to the representation of sentences mentioned above, the context of sentence is also important for inferring

| Method   | MC160 Test | MC500 Test |
|--|------------|------------|
| Richardson et al.(2013)+RTE                    | 0.691      | 0.633      |
| Sachan et al.(2015)                            | -          | 0.678      |
| Wang et al.(2015)                              | 0.753      | 0.699      |
| Trischler et al.(2016)                         | 0.746      | 0.710      |
| Attentive Reader                               | 0.463      | 0.419      |
| Neural Reasoner                                | 0.476      | 0.456      |
| HABCNN-TE                                      | 0.631      | 0.529      |
| Sentence-Level Attention Reader (mode:max+avg) | 0.664      | 0.673      |

TABLE 2: Experimental results for MCTest.

the answer. So the embedding of the sentence in the document is inputted into bi-LSTM layer to get their contextual representations. In our model, the bidirectional LSTM is used as RNN implementation.

$$\overrightarrow{h_i} = LSTM\left(\overrightarrow{h_{i-1}}, p_i\right), \quad i = 1, 2, \dots m$$
 (2)

$$\overleftarrow{h_i} = LSTM\left(\overleftarrow{h_{i+1}}, p_i\right), \quad i = m, m-1, \dots 1$$
(3)

$$h_i = concat\left(\overrightarrow{h_i}, \overleftarrow{h_i}\right)$$
 (4)

Finally, we take  $h_i$  to represent the contextual representations of sentences.  $h_{c.s} \in \mathbb{R}^d$  denote the sentence embedding of the option, where d denotes the number of options.

4.2. Sentence-Level Attention. In attention layer, we directly use a dot product of  $h_i$  and  $h_{c.s}$  to compute the "importance" of each sentence in the document for each option. And we use the softmax function to get a probability distribution. For each sentence in the document, "attention" is computed as follows.

$$\alpha(t) = softmax(h_i(t) \bullet h_{cs})$$
 (5)

where variable  $\alpha(t)$  is the attention weight a tth sentence in document.

In merging layer, the consensus attention is calculated by a merging function as follows.

$$s \propto \begin{cases} \sum_{t=1}^{m} \alpha(t), & \text{if mode = sum;} \\ \frac{1}{m} \sum_{t=1}^{m} \alpha(t), & \text{if mode = avg;} \\ \max_{t=1...m} \alpha(t), & \text{if mode = max;} \\ \frac{1}{n} \sum_{t=1}^{n} \alpha(t), & \text{if mode = max + avg,} \end{cases}$$

$$(6)$$

where n is the top number of the attention weight and n < m.

4.3. Output Layer. Finally, the answer is estimated by the softmax function.

$$a_i = soft \max(W_a * s_i), \quad i = 1...5 \tag{7}$$

where  $W_a$  indicate the weight matrix in the softmax layer and  $a_i$  is a probability distribution of the answer. The prediction of answer labels (such as "1 1 0 1 0") is gotten by the probability. Figure 1 shows the proposed neural network architecture.

#### 5. Experiments

In this section we evaluate our model on the MCTest and our Chinese reading comprehension datasets. We find that although the model is simple, it achieves state-of-the-art performance on these datasets.

5.1. Experimental Details. We use stochastic gradient descent with AdaDelta update rule [18], which only uses the first-order information to adaptively update learning rate over time and has minimal computational overhead. To train model, we minimize the negative log-likelihood as the objective function. The batch size is set to 5 and the number of iterations is set to 25.

For word vectors we use Google's publicly available embedding [19], whose training dataset is 70 thousand literary papers. The dimension of word embedding is set to 200. While we are implementing the sentence-level attention reader, it is easy to overfit the training data. Thus, we adopt dropout method [20] for regularization purpose and handling overfitting problems. The dropout rate is 0.1 on Chinese reading comprehension datasets and 0.01 on MCTest datasets, respectively. Implementation of our model is done with theano [21].

The answer is predicted according to whether the option is consistent with the document meaning for multiple-choice task, so we only evaluate our system performance in terms of precision ( $P = \text{right\_options/sum\_options}$ ).

5.2. Results on MCTest Dataset. To verify the effectiveness of our proposed model, we test firstly our model on public datasets. Table 2 presents the performance of feature engineering and neural methods on the MCTest test set. The first four rows represent feature engineering methods and the last four rows are neural methods. As we can see the feature engineering methods outperform the neural methods significantly. One possible reason is that the neural methods suffered from the relative lack of training data. So we are going to analyze the related feature and add it to our neural network model in future work.

| BCEETest | SBCEETest1                       | SBCEETest2  |
|----------|----------------------------------|---|
| 0.428    | 0.442                            | 0.438   |
| 0.461    | 0.452                            | 0.455   |
| 0.495    | 0.491                            | 0.499   |
| 0.513    | 0.503                            | 0.516   |
| 0.581    | 0.535                            | 0.578   |
|          | 0.428<br>0.461<br>0.495<br>0.513 | 0.428       0.442         0.461       0.452         0.495       0.491         0.513       0.503 |

TABLE 3: Comparison of different reader model on three testing datasets.

For neural methods, the attentive reader [4] is implemented at word representation level and it is a deep model with thousands of parameters, so it performs poorly on MCTest. The neural reasoner [22] has multiple reasoning layers and all temporary reasoning affects the final answer representation. The HABCNN-TE [14] is convolutional architecture network. It can cut down on the parameter count, but the context representation can not be presented enough. Our method addresses the problems of the above methods. Firstly, the recurrent architecture network also cuts down on the parameter count and it can present the context representation at sentence level. Then, we use the max+avg method to reduce the impact of all snippets. Experimental results also demonstrate that our method performs better than the other three neural methods.

- 5.3. Results on Chinese Reading Comprehension Datasets. We have set four baselines for Chinese reading comprehension datasets. One is the HABCNN-TE method which is the most optimal method on MCTest datasets and the other three are as follows.
- (i) The first baseline is inspired by Cui et al. [7]. We use the consensus attention-based neural network (called CAS Reader) for word of document and option. The model computes the attention of each document word directly, in respect to each option word at time t. The final consensus attention of option is computed by a merging function.
- (ii) The second baseline uses a sliding window and matches a bag of words constructed from the document and the option, respectively (called Match Reader). This baseline is inspired from Zhang et al. [23].
- (iii) The third baseline is the sentence similarity measure model (called SM Reader). The similarity is presented by the cosine similarity between the document sentence and the option sentence. The sentence representation is taken from Tai et al. [24]. The experimental results are given in Table 3.

The results on three test sets show that our sentence-level attention reader gives competitive results among various state-of-the-art baselines. We can observe that the accuracy in BCEETest outperforms the other test set. A possible reason can be that the college entrance examination is more standardized than that of the simulation. Also, we have noticed that the performance of the sentence-level model is better than the word-level model. For example, in BCEETest set, the SM Reader (sentence-level) outperforms the Match Reader (word-level) by 3.4% and The Sentence-Level Attention Reader (sentence-level) outperforms the CAS Reader (word-level) by 4.9% in precision, respectively.

Table 4: Results of different merging function.

|              | BCEETest | SBCEETest1 | SBCEETest2 |
|--------------|----------|------------|------------|
| Mode:avg     | 0.562    | 0.513      | 0.550      |
| Mode:sum     | 0.554    | 0.503      | 0.531      |
| Mode:max     | 0.492    | 0.496      | 0.442      |
| Mode:max+avg | 0.581    | 0.535      | 0.578      |

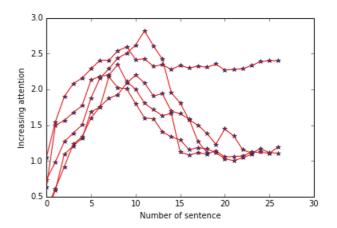


FIGURE 3: Experiment about the top N.

In experimenting we find out that the number of related sentences with the option is very important. So we also evaluate different merging functions as CAS Reader. The results are shown in Table 4. From the results, we can see that the avg and sum methods outperform the max method. A possible reason can be that the max method is equivalent to one sentence of document instead of the original document and a lot of information is lost. However, doing it achieves the best performance in which all sentences in document are used in the model. In order to measure it, we also use the max+avg method as the merging function. The "max" denotes the top N sentences and the "avg" denotes the average of top N sentences. In comparison with the avg method, the accuracy of the max+avg method increased by around 2% on three datasets. And this result is consistent with error analysis in Section 5.5. We suspect that some sentences interfere with the final answer as negative factor. Figure 3 shows the experiment about top N. We select randomly 5 options to do the experiment from the 13 Beijing college entrance examination papers (BCEETest). As we can see, the attention will not continue to increase in around 10. So N is set to 10 in our model. As shown in Box 2. The bold word "Ruins" is a derogatory term that it is irrelevant to cultural and aesthetic in many Chinese mind, and interpretation of the word "ruins" is only a "city and village are changed into desolate places by destruction or natural disasters" in the "Modern Chinese Dictionary"; There is no fault for the interpretation, but it is not enough if it is measured by world knowledge. In Europe, the meaning of "ruins" has been enriched and expanded since modern times. It has been endowed with the connotation of culture and aesthetics, and has become an academic concept.......

 $C_i$ ="One of the purposes of this paper is to correct the misunderstanding of the term "ruins" in the modern Chinese dictionary."

Box 2: Example of related sentences with the choice.

Table 5: Results of two sentence representation models.

|                           | BCEETest | SBCEETest1 | SBCEETest2 |
|---------------------------|----------|------------|------------|
| LSTM sentence model       | 0.518    | 0.495      | 0.489      |
| Sentence composition mode | 1 0.483  | 0.506      | 0.522      |
| Fusion model              | 0.581    | 0.535      | 0.578      |

denotes the most related sentences with the choice  $c_i$ ; the italic word has a little relation with the choice  $c_i$ ; the "....." is not relation.

5.4. Sentence Representation Model Analysis. In this paper, we use two models for the sentence representation, which are LSTM sentence model and sentence composition model [17]. Therefore, we have tested the contribution of the two models to the final model, respectively. The results are shown in Table 5.

The results on three test sets show that the precision of the fusion model is better than that of any single model. Therefore, we use the fusion model in sentence-level attention neural network.

- 5.5. Error Analysis. To better evaluate the proposed approach, we perform a qualitative analysis of its errors. Two major errors are revealed by our analysis, as discussed below.
- (i) The positioning feature word (such as "The second paragraph...") often appears in the options. To further analyze the locating property of our model, we also examine the dependence of accuracy on the positioning feature word. And all sentences are replaced by related sentences of the positioning feature word in document. The accuracy has increased by about 3% on these three datasets. The positioning feature word we use is shown as follows.

[The end of paper; The second paragraph; The end paragraph; The end of paper; The first paragraph]

According to the above description, we will consider adding more features, such as location features, into our model in future work.

(ii) Our model may make mistakes when the option is expressed with emotion (such as "This paper not only contains historical interest, but also infiltrated the concern of reality and express the author's desire to enhance the cultural quality of the nation."). It is very difficult to calculate the attention between the option emotion and the document

emotion. To handle such case correctly, our model will consider the emotion feature in future work. We have about more than 500 emotion feature words, like "thought provoking", "directly express one's mind", and so forth.

#### 6. Conclusion

In this paper, we introduce a sentence-level neural network model to handle the multiple-choice Chinese reading comprehension problems. The experimental results show that our model gives a state-of-the-art accuracy on all the evaluated datasets. We also use the max+avg method as the merging function and the accuracy of the max+avg method increased by about 2%. Furthermore, we analyze the positioning feature word and find that the accuracy increased by about 3%.

The future work will be carried out in the following aspects. First, we would like to extend our Chinese reading comprehension datasets and release it. Second, we are going to analyze the emotion feature and add it to our neural network model.

#### **Data Availability**

The Chinese reading comprehension data used to support the findings of this study are available from the corresponding author upon request.

#### **Conflicts of Interest**

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (nos. 61772324, 61673248), the Shanxi Natural Science Foundation of China (no. 201601D102030), and Program for Fostering Talents of Shanxi Province Joint Postgraduate Training Base (nos. 2017JD05, 2018JD01).

#### References

D. Chen, J. Bolton, and C. D. Manning, "A thorough examination of the CNN/daily mail reading comprehension task," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, ACL 2016, pp. 2358–2367, August 2016.

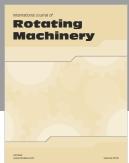
- [2] Y. Z. Liu, S. M. Sun, and L. Y. X. Kand Ruobing, "Knowledge representation learning: a review," *Journal of Computer Research and Development*, vol. 53, no. 2, pp. 247–261, 2016.
- [3] R. Matthew, J. C. Christopher, and R. Erin, "MCTest: a challenge dataset for the open-domain machine comprehension of text," in *Proceedings of the 2013 on Empirical Methods in Natural Language Processing*, pp. 193–203, 2013.
- [4] H. Felix, B. Antoine, C. Sumit, and W. Jason, "The goldilocks principle: reading children s books with explicit memory representations," 2015, https://arxiv.org/abs/1511.02301.
- [5] R. Pranav, Z. Jian, L. Konstantin, and L. Percy, "SQuAD: 100,000+ Questions for machine comprehension of text," in Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pp. 2383–2392, 2016.
- [6] J. Mandar, C. Eunsol, SW. Daniel, and Z. Luke, "TriviaQA: a large scale distantly supervised challenge dataset for reading comprehension," in *Proceeding of the 55th Annual Meeting of the Association for Computational Linguistics*, pp. 1601–1611, 2017.
- [7] Y. M. Cui, T. Liu, and Z. P. Chen, "Consensus attention-based neural networks for Chinese reading comprehension," 2016, https://arxiv.org/help/prep.
- [8] Y. Cui, Z. Chen, S. Wei, S. Wang, T. Liu, and G. Hu, "Attention-over-attention neural networks for reading comprehension," in Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, pp. 593–602, 2017.
- [9] M. H. Karl, K. Tomas, G. Edward et al., "Teaching machines to read and comprehend," *In Advances in Neural Information Processing Systems*, pp. 1684–1692, 2015.
- [10] J. M. Karl PRaymond, "Using sentence-level LSTM language models for script inference," in *Proceeding of the 54th Annual Meeting of the Association for Computational Linguistics*, pp. 279–289, 2016.
- [11] S. Mrinmaya, D. Avinava, P. X. Eric, and R. Matthew, "Learning answer-entailing structures for machine comprehension," in In Proceeding of the 53th Annual Meeting of the Association for Computational Linguistics, pp. 239–249, 2015.
- [12] H. Wang, B. Mohit, G. Kevin, and A. M. David, "Machine comprehension with syntax, frames, and semantics," in *Proceedings of the 53th Annual Meeting of the Association for Computational Linguistics*, pp. 700–706, 2015.
- [13] T. Adam, Z. Ye, and Y. Xingdi, "A parallel-hierarchical model for machine comprehension on sparse," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pp. 432–441, 2016.
- [14] W. Yin, S. Ebert, and H. Schütze, "Attention-Based Convolutional Neural Network for Machine Comprehension," in *Proceedings of the Workshop on Human-Computer Question Answering*, pp. 15–21, San Diego, California, June 2016.
- [15] K. Rudolf, S. Martin, B. Ondrej, and K. Jan, "Text understanding with the attention sum reader network," 2016, https:// arxiv.org/abs/1603.01547.
- [16] D. Bhuwan, H. Liu X, and L. Yang Z, "Gated-attention readers for text comprehension," in *Proceeding of the 55th Annual Meeting of the Association for Computational Linguistics*, pp. 1832–1846, 2017.
- [17] Y. L. Wang, "Sentence composition model for reading comprehension," *Journal of Computer Application*, vol. 37, no. 6, pp. 1741–1746, 2017.
- [18] D. Z. Matthew, "Adadelta: an adaptive learning rate method," 2012, https://arxiv.org/abs/1212.5701.

- [19] M. Tomas, C. Kai, C. Greg, and D. Jeffrey, "Efficient estimation of word representations in vector space," in *Proceedings of the In Proceedings of workshop at ICLR*, pp. 1–12, 2013.
- [20] S. Nitish, E. H. Geoffrey, K. Alex, S. Ilya, and S. Ruslan, "Dropout, a simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.
- [21] Theano Development Team, "Theano: a python framework for fast computation of mathematical expressions," 2016, https:// arxiv.org/abs/1605.02688.
- [22] P. Baolin, L. Zhengdong, L. Hang, and W. Kanfai, "Toward neural network-based reasoning," 2015, https://arxiv.org/abs/ 1508.05508.
- [23] Zhang Z. C., Z. Yu, and T. Liu, "Answer sentence extraction of reading comprehension based on shallow sematic tree kernel," *Journal of Chinese Information Processing*, vol. 22, no. 1, pp. 80– 86, 2008.
- [24] T. Kaisheng, S. Richard, and D. M. Christopher, "Improved semantic representations from tree-structured long short-term memory networks," in *Proceedings of the 53th Annual Meeting* of the Association for Computational Linguistics, pp. 1556–1566, 2015.

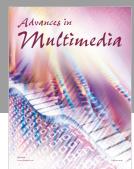


















Submit your manuscripts at www.hindawi.com





