IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 30 September 2022, accepted 22 October 2022, date of publication 26 October 2022, date of current version 4 November 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3217479

RESEARCH ARTICLE

SSIN: Sentence Semantic Interaction Network for Multi-Choice Reading Comprehension

XIAOBO XU, TURDI TOHTI, AND ASKAR HAMDULLA
School of Information Science and Engineering, Xinjiang University, Ürümqi 830017, China
Xinjiang Key Laboratory of Signal Detection and Processing, Ürümqi 830017, China

Corresponding author: Turdi Tohti (turdy@xju.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 62166042 and Grant U2003207; in part by the Natural Science Foundation of Xinjiang, China, under Grant 2021D01C076; and in part by the Strengthening Plan of National Defense Science and Technology Foundation of China under Grant 2021-JCJQ-JJ-0059.

ABSTRACT Multiple-choice reading comprehension is a challenging task in natural language processing, which aims to select the correct answer from a set of candidate options when given passage and question. Previous approaches usually focus only on word vector interactions and ignore the importance of sentence semantics for reading when modeling the relationship between passage and question. However, reading is a process that includes complex interactions of various knowledge such as vocabulary, syntax and semantics. Interactions based on word vectors alone do not effectively capture the relationship between passage and question. In this work, we propose the Sentence Semantic Interaction Network (SSIN), which models the relationship among passage, question, and answer options based on sentence semantics. The experimental results show that superior results are achieved on both the RACE and MCTest datasets, confirming that the interaction based on sentence semantic vectors can effectively improve the performance of the model reading comprehension.

INDEX TERMS Multiple-choice reading comprehension, natural language processing, sentence semantic interaction network.

I. INTRODUCTION

Machine Reading Comprehension (MRC) is a fundamental and long-term task in natural language understanding tasks, which is a technique that uses algorithms to enable computers to understand the semantics of texts and answer related questions [1], [2], [3]. With the advent of the information age, especially the rapid development of the Internet in recent years, a large amount of text data has been generated in various industries, which is too time-consuming and costly to process manually. MRC technology, which can automatically process text data and accurately extract semantic knowledge of the text, is gradually gaining attention. For example, while traditional search engines can only return documents related to a user's query, reading comprehension models can pinpoint the answer to a question in a document. In customer service, using machine reading comprehension to find the part of a product document relevant to the user's description of the problem and provide a detailed solution can significantly improve customer service efficiency. In the intelligent medical field, reading comprehension models can automatically review many cases and medical papers based on patient symptom descriptions, find possible causes and output treatment plans. In language education, reading comprehension models can be used to review students' essays and give suggestions for improvement. As you can see, machine reading comprehension can help save a lot of manpower and time in any scenario where large amounts of text data need to be processed and analyzed.

Machine reading comprehension can be classified into fill-in-the-blank, extractive, generative, and selective reading comprehensions based on the type of answers. Currently, the accuracy of fill-in-the-blank and extractive machine reading comprehension exceeds that of the human level [4]. However, responses to both the fill-in-the-blank and extractive reading comprehension tasks are based on words or fragments that appear in the original text, and there are no inferential questions in the dataset. Even though they perform well on both

The associate editor coordinating the review of this manuscript and approving it for publication was Okyay Kaynak.

tasks, they rely heavily on pattern recognition and matching at the statistical level and lack human comprehension and inference skills.

This paper will focus on multiple-choice MRC, which aims to select the correct answer from a set of candidate options when given passage and question. The dataset of selective reading comprehension contains many inferential questions, such as article summaries and attitude analysis. Compared to the assessment methods for generative reading comprehension tasks, selective reading comprehension is easy to assess and has uniform assessment metrics, which facilitate objective and fair judgment and comparison of models. In short, selective MRC can, to a certain extent, reflect the machine's comprehension, reasoning, and expression ability. An example of selective MRC is shown in Figure 1.

Passage:
Half a century has passed, but Shao Guohua never forgot the film ticket his teacher gave him that warmed his heart when he was a boy. Shao entered primary school in Zigong, Sichuan Province, in 1960. Back then, sports and watching movies were the students' favorite times. Once, the school organized a trip to the movies for the students, and many of Shao's classmates signed up for it. **But the boy had no money to buy the 0.05-yuan ticket.** When his classmates lined up to enter the movie theater, the boy stood alone outside. Then Shao's Chinese teacher Mo Xunru saw him and bought a ticket for him…
Questions:
1. Why didn't Shao Guohua enter the movie theater at first?
A. Because he didn't like the film.
B. Because his classmates didn't like him.
C. Because he had no money to buy the ticket.
Answer: C

FIGURE 1. Sample reading comprehension problems from RACE.

Most previous work has focused on the interaction between passage, question, and option word vectors for selective machine reading comprehension, ignoring the importance of sentence semantics for reading. Sentences are words and phrases with a specific intonation that expresses a complete meaning. Therefore, understanding sentence semantics is crucial to comprehend the content of a passage.

The word vector-based interaction focuses more on matching the words of question and options with the words of the passage, ignoring the importance of sentence semantics for reading. To solve this problem, this paper proposes a sentence semantic-based interaction network based on word vector-based interaction. In the interaction of passage, question and options, the passage is first converted into a set of sentences, and the question and the candidate options are converted into a sentence, respectively. Then the sentence vectors of question and options are used to interact with a set of sentence vectors of passage, respectively, to obtain sentence vectors of question-aware passage and sentence vectors of option-aware passage. Finally, the sentence vectors of the question-aware passage and the option-aware passage are fused by a gating mechanism to dynamically select the sentence vector of the passage containing the answer to the question.

In recent years, pre-trained models have started to appear in machine reading comprehension, such as the pre-trained models CoVe [5] and BERT [6] which can be used as a contextual encoding layer for machine reading comprehension network architectures, and have achieved very impressive results. Since the pre-training model has proven to have very powerful performance as an encoder, our model uses BERT as a contextual encoding layer, treating the output vector of the BERT model as a word vector of passages, questions, and options.

The contribution of this paper is as follows.

1. In this paper, a framework for reading comprehension is designed. Firstly, we interact with the passage through questions and options, respectively, to extract the passage's content related to the questions and options. Then the content that contains the answer to the question is dynamically selected through a gating mechanism.

2. In this paper, an interaction network based on sentence semantics is proposed. Reading is a process that includes complex interactions of various knowledge such as lexical, syntactic, and semantic. Therefore, we add a sentence semantic-based interaction network to the BERT model of word vector interaction so that the model also focuses on the semantics, syntax, and sentiment of sentences. Compared with other existing approaches, our approach no longer interacts only with word vectors but with an interaction model that combines word and sentence semantics.

This paper is structured as follows. In Section I and Section II, we introduce the research background and related works. In Section III, we introduce the word vector-based interaction model and the sentence vector-based interaction model. In Section IV, we perform an objective evaluation on the selective machine reading comprehension datasets RACE and MCTest. In Section 5, we conduct conclusions and outlook.

II. RELATED WORK

Researchers have been working on machine reading comprehension since the 1970s. For example, Lehnert proposed the QUALM system in 1977, which first showed the importance of text context when answering questions [7]. Hirschman proposed the DEEP READ system in 1999, which mainly uses a rule-based approach, such as adding various types of features to the rules (stemming extraction, semantic class identification, denotational disambiguation, BoW, and other features) to achieve an understanding of text [8].

Between 2013 and 2015, with the development of machine learning, researchers tried to define machine reading comprehension as a form of supervised learning [9], where a mathematical model is trained with manually annotated data that can map passages and questions into one answer. For example, Sachan et al. proposed a hidden (underlying) structure to explain the relationship between the question, the correct answer and the passage and proposed a unified max-margin framework [10] that learns to discover these hidden structures (given a corpus of question-answer pairs) and uses the learned content to answer the questions.

Since 2015, with the rapid development of deep learning, researchers have applied deep learning to machine reading comprehension models. For example, the Bi-Directional Attention Flow (BIDAF) model proposed by Seo et al. [11] is a machine reading comprehension model based on text and interproblem attention construction, which establishes a three-layer architecture of the coding layer-interaction layer-output layer of the reading comprehension model.

Today's machine reading comprehension models use word vectors for interaction when modeling passage, question and option relationships. For example, the model proposed by Sun et al [12], appends information related to questions and answers to the word embedding of passage in the interaction layer through the Highlighting mechanism, which facilitates the model to "remember" this helpful information. However, it still uses word vectors to interact among passages, questions and options. The model proposed by Dai et al [13], in response to the previous model that emphasized global information neglecting local information, so multiple convolutional kernels of different sizes are used in the interaction layer to extract local semantic information of different granularity. However, the word vectors of passage, questions, and options are still used. The model proposed by Zhang et al [14] addresses the previous unidirectional matching strategies that usually calculate question-aware passage representation and ignore passage-aware question representation when modeling the relationship between passage and question, so a Dual Co-Matching matching model was constructed for passage, question and option interaction using an attention mechanism in the interaction layer. However, the word vectors of the passage, question, and option are still used. The model proposed by Zhu et al [15] simplifies the DCMN model by using a Dual multi-headed mutual attention mechanism in the interaction layer in response to the complex structure of the DCMN model. But still, the interaction is performed using word vectors. The model proposed by Hu et al [16] addresses the previous model did not pay attention to the relationship between options, so a four-way bidirectional attention strategy was proposed to formulate the interactions among the passage, questions and candidate options. This enables the model to leverage the option correlation information for inferring the final answer accurately. However, the word vectors of the passage, question, and option are still used.

By looking at the development of today's machine reading comprehension technology, it is easy to see that today's scholars prefer to use the word vectors of passage, question and option in machine reading comprehension for interaction while neglecting to convert the three into sentence vectors for interaction. When we do a reading comprehension passage, we need to locate the words in the passage according to the question and options and understand the meaning of the question and options to find the correct answer. The words in the passage and the options only serve to locate the answer, but the real answer depends on understanding the meaning of passage, questions and options. Therefore, it is not enough to rely on word interaction alone, but sentence-based interaction is also significant.

In summary, this paper proposes an interaction model based on sentence semantics on the basis of word vector interaction. Firstly, the interaction of passage, question and option word vectors is carried out by the BERT model, and the word vectors of passage, question and option are extracted. Then the interaction of sentence vectors is performed by converting the word vectors of the three into sentence vectors through the sentence conversion module. Finally, the prediction of answers is performed.

III. MODEL

Our model is shown in Figure 2. The main components of the model are the sentence transformation module and the interaction module based on sentence semantics. We will discuss each part of the model in detail next.

A. TASK DEFINITION

For a multiple choice reading comprehension task, the machine needs to select the correct answer option from a given passage (P), a question (Q), and a set of candidate options (A) for this question.

B. CONTEXTUALIZED ENCODING

The BERT model is a multi-layer Transformer structure. Its input is each word in a piece of text, and its output is the BERT encoding of each word. The BERT encoding of a word contains the word and its contextual information. The pre-trained model BERT is used as the encoding layer of our model, which encodes passage, question and options as a fixed-length word vector. The encoding is shown below when given a passage, question, and options.

$$H^p = Encode(P). \tag{1}$$

$$H^q = Encode(Q). \tag{2}$$

$$H^a = Encode(A). \tag{3}$$

where Encode($\cdot$) represents the word vector output from the last layer of the BERT model. $H^p \in R^{|p|\times l}, H^q \in R^{|q|\times l}, H^a \in R^{|a|\times l}$ is the word vector of passage, question and options, respectively. |p|, |q|, and |a| are the lengths of passage, question, and options, respectively. ℓ is the dimensionality of the word vector.

C. SENTENCE SEMANTICS-BASED INTERACTION MODEL

Sentences are composed of words and phrases that follow a certain grammatical structure and can express a complete meaning. The interaction among passage, question and options based on word vectors ignores the grammatical structure of the sentences, etc. However, the interaction among passage, question and options in the existing MRC model is based on word vectors in the P, Q, A triad, and sentence-level interaction is missing. Therefore, we propose a sentence semantics-based interaction module that focuses on the grammatical structure along with the words.

Reading is a process that includes complex interactions of various knowledge such as lexical, syntactic and semantic; word-based interaction or sentence-semantic-based interaction alone is not an effective reading. In our model, passage, question, and options interact with word vectors through the pre-training model BERT and then with sentence vector-based interactions. Combining the pre-training model with our proposed sentence interaction model allows our model to include various knowledge complex interactions such as lexical, syntactic and semantic.

TABLE 1. RACE dataset passage, question and option length statistics.

Dataset	RACE-M	RACE-H	RACE
Passage Len	231.1	353.1	321.9
Question Len	9	10.4	10
Option Len	3.9	5.8	5.3

Table 1 shows the statistics of passage, question and option lengths in the RACE dataset, where the average length of questions is 10 and the average length of options is 5.3. Therefore, we take the first 19 word vectors of questions and options output from the BERT model to construct new word vectors of questions and options and then convert them into sentence vectors. The rationale for this is that the first 19 word vectors are chosen to retain most of the useful information about the questions and options in the passage, and to save computational resources. The RACE dataset's average passage length is 322 words, so passages are not processed.

The process is shown below.

$$\overline{H^q} = NEW\_QA(H^q). \tag{4}$$
$$\overline{H^a} = NEW\_QA(H^a). \tag{5}$$

where NEW_QA($\cdot$) stands for taking the first 19 word vectors. H^q and H^a are the word vectors of question and options provided by the encoding layer. $\overline{H^q}, \overline{H^a}$ are the newly constructed word vector of question and options.

Then, $\overline{H^q}$ and $\overline{H^a}$ are input to the bidirectional GRU respectively, and the state of the final GRU is taken as the sentence vector of question and options.

The process is shown below.

$$\widehat{H^q} = Last\_hidden(BiGRU(\overline{H^q})). \tag{6}$$
$$\widehat{H^a} = Last\_hidden(BiGRU(\overline{H^a})). \tag{7}$$

where BiGRU($\cdot$) represents the bidirectional GRU. Last_hidden($\cdot$) represents the final GRU status. $\widehat{H^q}, \widehat{H^a} \in R^{1\times l}$ are the sentence vectors of question and options.

The word vector of the article is input to the bidirectional GRU, and the output word vector obtained takes one word vector for every i interval as the sentence vector summarizing these words, this is because the state of the ith word contains the previous words' information.

In this paper, we take all word vectors output by GRU because this allows the model to learn which word vector can represent the sentence vector, rather than artificially specifying the sentence vector.

The process is shown below.

$$\overline{H^p} = OUTPUT(BiGRU(H^p)). \tag{8}$$
$$\widehat{H^p} = NEW\_P(\overline{H^p}). \tag{9}$$

where BiGRU($\cdot$) represents the bidirectional GRU. OUTPUT($\cdot$) represents the output of the bidirectional GRU. NEW_P($\cdot$) represents the newly constructed sentence vector of the passage. H^p is the word vector of the passage provided by the encoding layer.

After the above operation, we can get the sentence vector $\{\widehat{H^p}, \widehat{H^q}, \widehat{H^a}\}$ of passage, question and options triples.

The process of interacting the passage and question sentence vectors is shown below.

$$\widehat{G^{pq}} = SoftMax(\widehat{H^p}\, W_1\, \widehat{H^{q^T}}). \tag{10}$$
$$\widehat{E^{pq}} = \widehat{G^{pq}}\widehat{H^q}. \tag{11}$$
$$\widehat{H^{pq}} = ReLU(\widehat{E^{pq}}\, W_2). \tag{12}$$

where $W_1, W_2 \in R^{l\times l}$ are learnable parameters. $\widehat{G^{pq}}$ is the weight matrix between the passage and the question sentence vector. $\widehat{H^{pq}}$ is the vector representation of the question-aware passage. The vector representation $\widehat{H^{pa}}$ of option-aware passage can be obtained in the same way.

Finally, we fuse the two parts of information, the vector representation $\widehat{H^{pq}}$ of question-aware passage and the vector representation $\widehat{H^{pa}}$ of option-aware passage, through the gating mechanism [17], so that the vector $\widehat{H^p}$ of passage sentence vectors containing the answers to the questions is obtained.

The process is shown below.

$$g = \sigma(\widehat{H^{pq}}\, W_3 + \widehat{H^{pa}}\, W_4 + b_1). \tag{13}$$
$$\widehat{H^p} = g * \widehat{H^{pq}} + (1-g)\widehat{H^{pa}}. \tag{14}$$

where $W_3, W_4 \in R^{l\times l}$ are learnable parameters. g is a reset gate. $\widehat{H^p} \in R^{|p|\times l}$ is the vector of article sentences containing the answers to the questions inferred from the vector of questions and option sentences.

D. OBJECTIVE FUNCTION

If A_k is the correct option, then the objective function is calculated as follows.

$$C = MaxPooling(\widehat{H^p}). \tag{15}$$

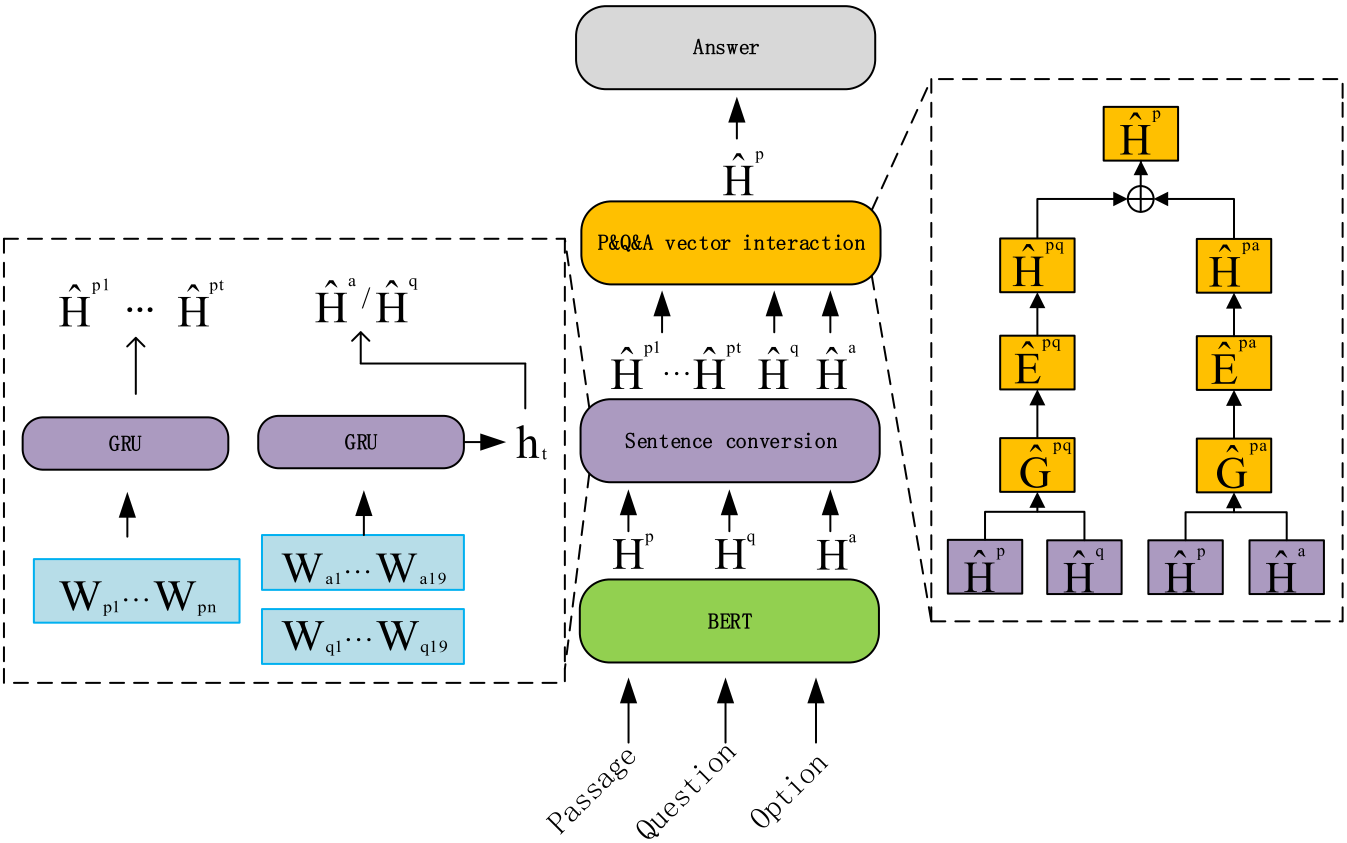


FIGURE 2. Sentence vector-based interaction model.

$$L(A_k\,|P, Q) = -\log \frac{\exp(V^T C_k)}{\sum_{j=1}^{m} V^T C_j}. \tag{16}$$

where $C \in R^{1\times l}$ is the maximum pooling. V^T is a learnable parameter. m is the number of answer choices.

IV. EXPERIMENT

A. DATASET

To validate the effectiveness of the proposed model, our model is evaluated on 2 multinomial choice MRC datasets in this paper. Details about these 2 datasets are shown below.

TABLE 2. Statistics for the multi-choice machine reading comprehension dataset. #O is the average number of candidate options for each question. #P and #Q are the number of passage and questions in the dataset.

Task	Domain	#O	#P	#Q
RACE[18]	General	4	27,933	97,687
MCTest[19]	Stories	4	660	2640

1) RACE

The dataset is drawn from the reading comprehension of middle and high school English exams in China between the ages of 12 and 18, and contains 28,000 short passages and nearly 100,000 questions. It contains a wide variety of questions used to assess students' comprehension skills. RACE consists of two subsets: RACE-M and RACE-H, which correspond to the middle school and high school difficulty levels, respectively, and is recognized as one of the largest and most difficult datasets for multiple-choice reading comprehension.

2) MCTest

Richardson, a researcher at Microsoft Research, published a dataset at EMNLP 2013. In this dataset, all documents are narrative stories. It examines reasoning skills that are limited to what is acceptable for a 7-year-old child and contains many common-sense inferences that are both factual and non-factual. The dataset contains two parts, MC160 and MC500, containing 160 and 500 passages, respectively.

B. EVALUATION INDICATORS

For multiple-choice tasks, since the answers are derived from a given set of options, the model answers can be compared directly with the correct answers during the assessment, so the indicator of accuracy is used as an indicator of assessment.

$$Acc = \frac{N^+}{N}. \tag{17}$$

where N is the total number of questions. N^+ is the number of correct answers given by the model.

C. IMPLEMENTATION DETAILS

Our model is evaluated based on a pre-trained language model. There are 12 layers of Transformer with 768 input and output dimensions, attention contains 12 Heads and 110 million parameters in total.

In our experiments, on the RACE dataset, the maximum sentence input length is set to 380, the dropout rate of each BERT layer is set to 0.1, the optimizer uses BertAdam, the learning rate is set to 5e-5, and we use an NVIDIA GeForce RTX 3090 graphics card, the batch size is set to 4, and the training period On the MCTest dataset, the maximum sentence input length is set to 290, the dropout rate of each BERT layer is set to 0.1, the optimizer uses BertAdam, the learning rate is set to 2e-5, we use an NVIDIA GeForce RTX 3090 graphics card, the batch size is set to 10, and the training period is 10.

The specific parameter settings are shown in Table 3.

TABLE 3. Experimental parameters settings.

	RACE	MCTest
Learning rate	5e-5	2e-5
Epoch	10	10
Batch size	16	9
gradient accumulation steps	8	1
GPU	ONE 3090	ONE 3090
Max input sequence length	380	290
Dropout rate	0.1	0.1
Optimizer	BertAdam	BertAdam

D. BASELINE METHODS

The comparison methods used in our experiments include the DCMN, DUMA and the pre-trained language model BERT.

BERT: It is a bidirectional pre-trained language model.

DCMN: It is a dual co-matching network, which matches the question and option to the passage bidirectionally, and BERT is used as the encoder.

DUMA: It is a dual Multi-head Co-Attention model, which simulates a human transposition thinking process to capture relationships of key information from the passage, question, and answer options.

E. EVALUATION ON THE RACE DATASET

Table 4 shows the experimental results on the RACE test dataset. We list multiple choice machine reading comprehension models in recent years, where DCMN* and DUMA* is our own implemented model and BERT_base is our baseline model. We give the test result of 64.8% in the DCMN model. Due to the limitation of computational resources, our batch size is 4, so the accuracy of our implemented DCMN model decreases by 2.2% (64.8% vs. 67.0%, original paper batch size is 16, GPU is 8 NVIDIA GeForce RTX 1080Ti).

TABLE 4. Experimental results on the RACE test set. WVIN: P&Q&A word vectors interaction network. Sc: Sentence conversion. SSIN: Sc + WVIN. The * symbols represent our implementation of the model.

Model	RACE-M/RACE-H	RACE
HAF[20]	45.0/46.4	46.0
MRU[21]	57.7/47.4	50.4
HCM[22]	55.8/48.2	50.4
MMM[23]	61.1/52.2	54.7
GPT[24]	62.9/57.4	59.0
RSM[12]	69.2/61.5	63.8
OCN[25]	71.6/64.8	66.8
$BERT_{base}$ + DCMN*	71.2/62.1	64.8
DCMN* + SSIN	**73.7**/64.5	67.1
$BERT_{base}$ + DUMA*	72.8/62.2	65.3
Our models		
$BERT_{base}$*	71.1/62.3	65.0
$BERT_{base}$* + WVIN	71.0/63.0	65.3(+0.3)
$BERT_{base}$* + SSIN	72.5/**65.1**	**67.2(+2.2)**

TABLE 5. Experimental results on the RACE validation set. WVIN: P&Q&A word vectors interaction network. Sc: Sentence conversion. SSIN: Sc + WVIN. The * symbols represent our implementation of the model.

Task	RACE
$BERT_{base}$ + DCMN*	64.9
DCMN+	**67.4**
$BERT_{base}$*	64.6
$BERT_{base}$ + DUMA*	65.3
Our models	
$BERT_{base}$* + WVIN	66.9
$BERT_{base}$* + SSIN	**67.4**

TABLE 6. Experimental results on the MCTest test set. WVIN: P&Q&A word vectors interaction network. Sc: Sentence conversion. SSIN: Sc + WVIN. The * symbols represent our implementation of the model.

Task	MC500	MC160
$BERT_{base}$*	72.9	68.5
$BERT_{base}$ + DCMN*	72.9	66.5
$BERT_{base}$ + DUMA*	72.5	66.8
Our models		
$BERT_{base}$* + WVIN	70.4	69.8
$BERT_{base}$* + SSIN	**73.8**	**71.7**

Table 5 shows the experimental results on the RACE validation dataset, where our sentence interaction model improves 2.8% (67.4% vs. 64.6%) over the baseline model and 2.5% (67.4% vs. 64.9%) over DCMN*, where DCMN* is the model we implemented. Compared to the DCMN+ model, our sentence interaction model achieves the same results with fewer resources (DCMN+: 8 NVIDIA GeForce RTX 1080Ti 1080Ti).

F. EVALUATION ON THE MCTest DATASET

Table 6 shows the experimental results on the MCTest test dataset, on MC500, our sentence interaction model improves 0.9% over the baseline model (73.8% vs. 72.9%), 0.9% over DCMN* (73.8% vs. 72.9%), and 3.2% over WVIN (73.8% vs. 70.4); on MC160, it improves 3.2% over the baseline model by 3.2% (71.7% vs. 68.5%), over DCMN* by 5.2% (71.7% vs. 66.5%), and over WVIN by 0.9% (71.7% vs. 69.8%). The effectiveness of our sentence interaction model on the MCTest test dataset is confirmed.

TABLE 7. Experimental results on the MCTest validation set. WVIN: P&Q&A word vectors interaction network. Sc: Sentence conversion. SSIN: Sc + WVIN. The * symbols represent our implementation of the model.

Task	MC500	MC160
$BERT_{base}$*	71.5	68.3
$BERT_{base}$ + DCMN*	65.0	62.5
$BERT_{base}$ + DUMA*	69.2	66.5
Our models		
$BERT_{base}$* + WVIN	70.8	65.5
$BERT_{base}$* + SSIN	**71.7**	**70.5**

Table 7 shows the experimental results on the MCTest validation dataset. On MC500, our sentence interaction model improved 0.2% over the baseline model (71.7% vs. 71.5%), 6.7% over DCMN* (71.7% vs. 65.0%), and 0.9% over WVIN (71.7% vs. 70.8%); on MC160, it improved 2.2% over the baseline model by 2.2% (70.5% vs. 68.3%), 8% (70.5% vs. 62.5%) over DCMN*, and 5% (70.5% vs. 65.5%) over WVIN. The effectiveness of our sentence interaction model on the MCTest validation dataset is confirmed.

G. EXPERIMENTAL ANALYSIS

Table 4 shows the results of the experiments on the RACE test dataset and its two subtasks. We will analyze them in detail next.

1) WORD VECTOR INTERACTION MODEL

The model WVIN proposed in this paper is an interaction model based on word vectors. As can be seen from the experimental data in Table 4, our model improves 0.5% (65.3% vs. 64.8%) over DCMN* and 0.3% (65.3% vs. 65.0%) over the baseline model. This model confirms the effectiveness of this paper's machine reading comprehension framework.

2) SENTENCE SEMANTIC INTERACTION MODEL

The model SSIN proposed in this paper is an interaction model based on sentence semantics. The model is mainly designed to address the importance of previous models focusing only on word vector interaction and ignoring the importance of sentence semantics for reading. As can be seen from the experimental data in Table 4, the sentence interaction model SSIN improves 1.5% (72.5% vs. 71.0%) on RACE-M, 2.1% (65.1% vs. 63.0%) on RACE-H, and 1.9% (67.2% vs. 65.3%) overall compared to the word vector interaction model WVIN. The sentence semantic interaction model proposed in this paper outperforms the word vector interaction model on RACE-M, a dataset focusing on word matching, and RACE-H, a dataset focusing on inference ability. This confirms that sentence semantic-based interaction can effectively improve the model's performance for reading comprehension.

We summarize the reasons resulting in such contrast as follows: (1) The interaction based on the word vectors can not capture the relationship between two sentences well. (2) Previous methods commonly used element splicing to fuse $\widehat{H^{pq}}$ and $\widehat{H^{pa}}$ in Equation 13 (e.g., $[\widehat{H^{pq}}, \widehat{H^{pa}}]$), and such fusion would allow some useless information to interfere with the selection of the correct option. From the experimental results, it can be seen that the sentence-level interaction is very helpful for the model to select the correct answer.

3) CHANGES TO THE DCMN MODEL BASED ON THE INNOVATIONS IN THIS PAPER

In this paper, the sentence transformation module is added to the DCMN model, namely DCMN* + SSIN. As can be seen from the experimental data in Table 4, compared with the DCMN* model, there is a 2.5% improvement in RACE-M (73.7% vs. 71.2%), a 2.4% improvement on RACE-H (64.5% vs. 62.1%), and an overall 2.3% improvement (67.1% vs. 64.8%). This also shows that the interaction of sentence semantics can effectively improve the performance of the model and confirms the applicability of our proposed improvements to other models as well.

4) OUR MODEL VS DCMN+

Table 5 shows the experimental results on the RACE validation dataset. From the experimental data in Table 5, we can see that our sentence vector-based interaction model and the DCMN+ model differ by only 0.01% (67.39% vs. 67.40%), but our model uses much fewer resources than the DCMN+ model.

The main differences are shown in Table 8.

TABLE 8. Comparison of SSIN and DCMN+ model resource usage.

	SSIN	DCMN+
Batch size	4	16
GPU	ONE 3090	EIGHT 1080Ti

V. CONCLUSION

In this paper, an interaction network based on sentence semantics is proposed to address the difficulty of capturing the relationship between sentences in previous models using word vectors of passage, questions and options for interaction. First, the word vectors of passage, questions and options are converted into sentence vectors. Then, the passage, questions and options are interacted at the sentence level. Finally, useful information is dynamically selected using a gating mechanism. We confirmed that the proposed approach is effective through relevant experiments.

Humans do reading comprehension to answer questions using general life knowledge, etc. as auxiliary knowledge in addition to information from the text. Present-day scholars are also investigating the introduction of external knowledge into machine reading comprehension; for example, Jiang et al [26] propose the use of external knowledge in the form of triads and corpora, Yang et al [27] propose a model for answering questions by searching for external knowledge, Duan et al [28] propose a model that can combine external knowledge and contextual fusion network, and Van et al [29] proposed a model to analyze questions using external knowledge, all of these models using external knowledge have achieved good performance, which also shows the direction for our future research. In our subsequent work, we will focus on how to effectively incorporate external knowledge into machine reading comprehension and, in response to the importance of sentence semantic vectors for reading comprehension, we will continue to investigate how to convert word vectors of articles into sentence vectors.

REFERENCES

[1] K. M. Hermann, T. Kocisky, E. Grefenstette, L. Espeholt, W. Kay, M. Suleyman, and P. Blunsom, "Teaching machines to read and comprehend," in *Proc. Adv. Neural Inf. Process. Syst.* , vol. 28, 2015, pp. 1–9.

[2] P. Rajpurkar, J. Zhang, K. Lopyrev, and P. Liang, "SQuAD: 100,000+ questions for machine comprehension of text," 2016, *arXiv:1606.05250*.

[3] T. Nguyen, M. Rosenberg, X. Song, J. Gao, S. Tiwary, R. Majumder, and L. Deng, "MS MARCO: A human generated machine reading comprehension dataset," in *Proc. CoCo@ NIPS*," 2016, pp. 1–10.

[4] Z. Zhang, J. Yang, and H. Zhao, "Retrospective reader for machine reading comprehension," in *Proc. AAAI Conf. Artif. Intell.*, 2021, vol. 35, no. 16, pp. 14506–14514.

[5] B. McCann, J. Bradbury, C. Xiong, and R. Socher, "Learned in translation: Contextualized word vectors," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–12.

[6] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2018, *arXiv:1810.04805*.

[7] W. G. Lehnert, *The Process of Question Answering*, New Haven, CT, USA: Yale Univ., 1977.

[8] L. Hirschman, M. Light, E. Breck, and J. D. Burger, "Deep read: A reading comprehension system," in *Proc. 37th Annu. Meeting Assoc. Comput. Linguistics*, 1999, pp. 325–332.

[9] R. Caruana and A. Niculescu-Mizil, "An empirical comparison of supervised learning algorithms," in *Proc. 23rd Int. Conf. Mach. Learn. (ICML)*, 2006, pp. 161–168.

[10] M. Sachan, A. Dubey, E. P. Xing, and M. Richardson, "Learning answer-entailing structures for machine comprehension," in *Proc. 53rd Annu. Meeting Assoc. Comput. Linguistics 7th Int. Joint Conf. Natural Lang. Process.*, vol. 1, 2015, pp. 239–249.

[11] M. Seo, A. Kembhavi, A. Farhadi, and H. Hajishirzi, "Bidirectional attention flow for machine comprehension," 2016, *arXiv:1611.01603*.

[12] K. Sun, D. Yu, D. Yu, and C. Cardie, "Improving machine reading comprehension with general reading strategies," 2018, *arXiv:1810.13441*.

[13] Y. Dai, Y. Fu, and L. Yang, "A multiple-choice machine reading comprehension model with multi-granularity semantic reasoning," *Appl. Sci.*, vol. 11, no. 17, p. 7945, Aug. 2021.

[14] S. Zhang, H. Zhao, Y. Wu, Z. Zhang, X. Zhou, and X. Zhou, "Dual co-matching network for multi-choice reading comprehension," 2019, *arXiv:1901.09381*.

[15] P. Zhu, Z. Zhang, H. Zhao, and X. Li, "DUMA: Reading comprehension with transposition thinking," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 30, pp. 269–279, 2021.

[16] L. Hu, D. Zou, X. Guo, L. Qi, Y. Tang, H. Song, and J. Yuan, "Four-way bidirectional attention for multiple-choice reading comprehension," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2021, pp. 238–243.

[17] R. Kumar Srivastava, K. Greff, and J. Schmidhuber, "Highway networks," 2015, *arXiv:1505.00387*.

[18] G. Lai, Q. Xie, H. Liu, Y. Yang, and E. Hovy, "RACE: Large-scale reading comprehension dataset from examinations," 2017, *arXiv:1704.04683*.

[19] M. Richardson, C. J. C. Burges, and E. Renshaw, "MCTest: A challenge dataset for the open-domain machine comprehension of text," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2013, pp. 193–203.

[20] H. Zhu, F. Wei, B. Qin, and T. Liu, "Hierarchical attention flow for multiple-choice reading comprehension," in *Proc. AAAI Conf. Artif. Intell.*, 2018, vol. 32, no. 1, pp. 1–8.

[21] Y. Tay, L. Anh Tuan, and S. Cheung Hui, "Multi-range reasoning for machine comprehension," 2018, *arXiv:1803.09074*.

[22] S. Wang, M. Yu, S. Chang, and J. Jiang, "A co-matching model for multi-choice reading comprehension," 2018, *arXiv:1806.04068*.

[23] M. Tang, J. Cai, and H. H. Zhuo, "Multi-matching network for multiple choice reading comprehension," in *Proc. AAAI Conf. Artif. Intell.*, 2019, vol. 33, no. 1, pp. 7088–7095.

[24] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskeve, "Improving language understanding by generative pre-training," Tech. Rep., 2018.

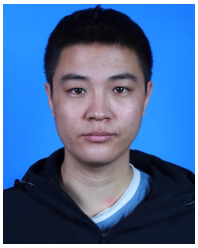
[25] Q. Ran, P. Li, W. Hu, and J. Zhou, "Option comparison network for multiple-choice reading comprehension," 2019, *arXiv:1903.03033*.

[26] Y. Jiang and H. Huang, "Analysis and improvement of external knowledge usage in machine multi-choice reading comprehension tasks," in *Proc. 2nd Int. Conf. Mach. Learn., Big Data Bus. Intell. (MLBDBI)*, Oct. 2020, pp. 85–88.

[27] Y. Yang and S. Kang, "Common sense-based reasoning using external knowledge for question answering," *IEEE Access*, vol. 8, pp. 227185–227192, 2020.

[28] Q. Duan, J. Huang, and H. Wu, "Contextual and semantic fusion network for multiple-choice reading comprehension," *IEEE Access*, vol. 9, pp. 51669–51678, 2021.

[29] K. V. Nguyen, K. V. Tran, S. T. Luu, A. G.-T. Nguyen, and N. L.-T. Nguyen, "Enhancing lexical-based approach with external knowledge for Vietnamese multiple-choice machine reading comprehension," *IEEE Access*, vol. 8, pp. 201404–201417, 2020.



XIAOBO XU is currently pursuing the master's degree with the School of Information Science and Engineering, Xinjiang University. His research interests include natural language processing and machine reading comprehension.

TURDI TOHTI received the Ph.D. degree from Xinjiang University, China, in 2014. He is currently a Professor at the School of Information Science and Engineering, Xinjiang University. His research interests include natural language processing, Q&A and dialogue systems, multimodal interaction and affective computing, multimodal knowledge graph construction, and application.

ASKAR HAMDULLA received the Ph.D. degree from the University of Electronic Science and Technology of China, in 2003. He is currently a Professor at the School of Information Science and Engineering, Xinjiang University. His research interests include speech recognition and synthesis, pattern recognition and image processing, natural language processing, information retrieval, and content security.