



Strategies for Improving Text Reading Ability Based on Human-Computer Interaction in Artificial Intelligence

Guorong Shen*

School of Foreign Languages, Henan University of Technology, Zhengzhou, China

OPEN ACCESS

Edited by:

Deepak Kumar Jain,
Chongqing University of Posts
and Telecommunications, China

Reviewed by:

Haiyang Yu,
Dalian Ocean University, China
Li Ming,
Jiangxi University of Finance
and Economics, China
Liu Yang,
Wuhan Institute of Technology, China

*Correspondence:

Guorong Shen
shengurong2006@haut.edu.cn

Specialty section:

This article was submitted to
Educational Psychology,
a section of the journal
Frontiers in Psychology

Received: 12 January 2022

Accepted: 09 February 2022

Published: 11 March 2022

Citation:

Shen GR (2022) Strategies
for Improving Text Reading Ability
Based on Human-Computer
Interaction in Artificial Intelligence.
Front. Psychol. 13:853066.
doi: 10.3389/fpsyg.2022.853066

In order to improve text reading ability, a human-computer interaction method based on artificial intelligence (AI) human-computer interaction is proposed. Firstly, the design of the AI human-computer interaction model is constructed, which includes the Stanford Question Answering Dataset (SQuAD) and the designed baseline model. There are three components: the coding layer is based on a cyclic neural network (recurrent neural network [RNN] encoder layer), which aims to encode the problem and text into a hidden state; the interaction layer is used to integrate problems and text representation; the output layer connects two independent soft Max layers after a fully connected layer, one is used to obtain the starting position of the answer in the text and the other is used to obtain the ending position. In the interaction layer of the model, this manuscript uses hierarchical attention and aggregation mechanism to improve text coding. The traditional model interaction layer has a simple structure, which leads to weak relevance between text and problems, and poor understanding ability of the model. Finally, the self-attention model is used to further enhance the feature representation of text. The experimental results show that the improved model in this manuscript is compared with the public AI human-computer interaction reading comprehension model. According to the data in the table, the accuracy of the model in this manuscript is better than that of the baseline model, in which the exact match (EM) value is increased by 1.4% and the F1 value is increased by 2.7%. However, compared with improvement point 2, the EM and F1 values of the model have decreased by 0.7%. It shows that the output layer has a certain impact on the effect of the model, and the improvement and optimization of the output layer can also improve the performance of the model. It is proved that AI human-computer interaction can effectively improve text reading ability.

Keywords: AI human-computer interaction, reading comprehension, neural network, attention mechanism, SQuAD dataset

INTRODUCTION

Artificial intelligence (AI) has originated around 1950 and has gone through more than half a century. During this period, it experienced two declines and rose for the third time in recent years. As a new engine of the wave of global scientific change in the next stage, it is expected to lead the fourth industrial revolution and apply AI scientific research to various industries, which will bring

broad development prospects. In recent years, China's graded reading software has also begun to develop Chinese graded reading evaluation standards. Developers link evaluation standards with massive learning resources through AI technology to form a knowledge map. Therefore, students can be evaluated, intelligently graded, and predicted. The multi-terminal interactive design of excellent graded reading software provides convenience for the selection of excellent talents (Li et al., 2020). Students complete learning tasks through the student side, such as homework assigned by teachers, independent variant exercises, and a systematic review of wrong problem books. Through the parent side, parents can synchronously understand the learning track, data statistics, children's wishes, strategy settings, etc. Teachers release learning tasks through the teacher side, track students' learning progress, see each student's quantifiable reading ability, and obtain visual data of class reading. The school can obtain the reading data report of the whole school through the headmaster, track the dynamic development of the school or regional reading level, and master the reading status of different grades and classes in real time, to facilitate macro-control and data analysis (Comunian, 2015). For example, check the poetry recitation ranking and student word breakthrough game ranking of all classes in the school, and select the list with the highest accuracy and the fastest speed. The school can also use this to select excellent talents. AI leads the future, which is also the general trend of future education development. Chinese teaching in primary schools is not only facing the challenge of conforming to the trend but also ushering in historical opportunities and a broader platform. Yu, an ancient educator, positioned teachers as "preaching, teaching and dispelling doubts." The application of graded reading software in primary school Chinese teaching can replace some functions of teachers' "teaching and dispelling doubts," help to share some heavy and trivial work and reduce teachers' teaching pressure. So that teachers can concentrate on more valuable work, that is, preaching and cultivate people with ideas, knowledge, wisdom, temperature, and soul (Martínez-Álvarez et al., 2015).

LITERATURE REVIEW

Ungureanu first proposed the match LSTM and answer pointer model. Its core idea is to build a text representation based on question awareness through match LSTM, and then, predict the starting position of answer boundary through pointer network (Ungureanu et al., 2020). Kraska-Szlenk proposed a bidirectional attention flow (BiDAF) model. The model proposes a two-way attention mechanism, which hierarchically obtains text representation information with different granularity in multiple stages, and obtains document representation based on problem perception through a two-way attention flow mechanism (Kraska-Szlenk, 2018). Ulaiwy and Alarnoosy proposed the document reader model. The model does different processing for the word embedding layer of passage and question. In the word embedding layer of passage, traditional grammatical features, such as part of speech, word frequency, and named entity type, are added, while the question is simply processed, and good

results are achieved with a relatively simple model architecture. At the same time, the author also constructs an open domain-based question answering system DrQA based on the document reader model (Ulaiwy and Alarnoosy, 2018). Xiong et al. (2016) proposed a cyclic neural network model. The greatest innovation of the recurrent neural network model is to apply the self-attention model to the reading comprehension task. Firstly, the model obtains the intermediate vector representation of the article with problem perception through the gated attention-based recurrent networks based on the attention mechanism and filters the impact of the problem on the article. Then, the self-attention mechanism is proposed to re-model the article. The self-attention mechanism can solve the problem that the recursive neural network is difficult to model the long-distance dependence in the text. The idea of the whole model can be formalized into the process of human reading comprehension: first, read the article with questions, preliminarily locate the article information related to the questions, then, browse an article again, screen out the real information that can answer the questions, and finally output the answers (Xiong et al., 2016). Renzulli proposed the FusionNet model. The model extends the existing attention methods from three aspects. Firstly, it proposes a new concept of "word history" to represent the attention information embedded from shallow words to the deep semantic level. Secondly, it identifies an attention scoring function that makes better use of the concept of "lexical history." Finally, the model uses full awareness (full attention mechanism) at different levels, which not only uses the feature vectors of the current text and problems but also integrates the low-level feature vectors and high-level vectors through the aggregation mechanism as the input of the model output layer (Renzulli, 2015). Wu and Zhang proposed the SLQA model. The model integrates attention mechanism and fusion methods in each layer of the network. Firstly, the pre-trained language model Embeddings from Language Models (ELMo) is added to the word vector representation layer. Then, the combination of co-attention and fusion is used to represent articles and problems in the shallow layer, and the combination of self-attention and fusion is used in the deep layer. The main idea of the model is to refine layer by layer, decompose the complex task process of reading and understanding, and extract more useful feature information (Wu and Zhang, 2021). Santos and Canello proposed the Question Answering Network (QANet) model. The current end-to-end AI human-computer interaction reading comprehension models are based on recurrent neural networks (RNNs) with an attention mechanism. Although they have advantages in representing sequence information, the training of recurrent neural networks cannot be parallel, resulting in low training efficiency. As we all know, a convolutional neural network can extract local interactive information, and self-attention mechanism can obtain global interactive information (Santos and Canello, 2015). Rasinski et al. proposed a new question answering architecture. All coding layers in the model use convolutional neural network and self-attention mechanism to replace recursive neural network. Finally, compared with the model based on recurrent neural network, the effect is similar, and the training speed and reasoning speed are improved

by about 10 times (Rasinski et al., 2016). Ma researchers proposed the pre-training language model Bidirectional Encoder Representation from Transformers (BERT). The pre-training model proposed by Google refreshed the best results of more than a dozen tasks in the field of natural language processing and improved the effect of the Stanford Question Answering Dataset (SQuAD) to exceed the human level. The release of pre-training model is a milestone in the field of natural language processing. Traditional word vectors, such as Word2Vec, are context independent. The word vector of each word will change with the task or training corpus. The word vector of the pre-training language model is context dependent, which can alleviate the dependence of specific tasks on the model structure (Ma, 2021).

Based on this, this manuscript proposes a human-computer interaction method based on AI human-computer interaction. Firstly, the design of the AI human-computer interaction model is constructed. As shown in **Figure 1**, it includes the SQuAD and the designed baseline model. The overall model has three components: RNN encoder layer based on cyclic neural network, which aims to encode problems and text into a hidden state. The interaction layer is used to integrate problems and text representations. The output layer connects two independent soft Max layers after a fully connected layer, one is used to obtain the starting position of the answer in the text and the other is used to obtain the ending position. In the interaction layer of the model, this manuscript uses hierarchical attention and aggregation mechanism to improve text coding. The traditional model interaction layer has a simple structure, which leads to weak relevance between text and problems, and poor understanding ability of the model. In order to extract more fine-grained text features, this manuscript first fuses the encoded results of the two two-way attention models and then, aggregates the low-level feature vectors into the current vector representation. Finally, the self-attention model is used to further enhance the feature representation of text.

ARTIFICIAL INTELLIGENCE HUMAN-COMPUTER INTERACTION MODEL DESIGN

Data Sets

The experiment uses the data set SQuAD, which contains about 100,000 triples in the form of text, questions, and answers, and SQuAD consists of 536 articles from Wikipedia. The articles are segmented manually. The generated natural segments are the text in the triple. An important feature of each data sample is that the answer always comes directly from the text, which means that the model does not need to generate the answer text, but only needs to intercept a text from the text corresponding to the question as the answer.

Evaluation Index

The data set contains public training set and verification set. The verification set provides three answers to the questions of each

sample, and each answer comes from different taggers. Therefore, the answers to the same question are not always completely consistent (Scammacca et al., 2015).

The performance of the model is measured by two indicators: F1 and exact match (EM) score.

EM (exact match) means exact match. It is used to detect whether the model output completely matches the real answer. There are only true and false results. For a question in a sample, if the model output answer is enough happiness and the real answer is have enough happiness, then the EM score of the model is 0.

Compared with EM, F1 is a relatively loose evaluation index, which is the harmonic average of precision and recall. In data science, precision measures how many of the answers given are real answers, and recall measures how many of the real answers are predicted. For the same example, if the output answer of the model is enough happiness and the real answer is have enough happiness, the model has 100% accuracy, because the real answer contains the prediction of the model, and the recall rate is 2/3, because only 2 of the 3 single word models in the real answer are given (Khalifian et al., 2015). As described above, according to the calculation formula of F1:

$$F1 = \frac{2 \times P \times R}{P + R} \quad (1)$$

At this time, F1 score is 80%.

As mentioned earlier, since the validation set of the SQuAD has three human-provided answers to each question, the largest EM and F1 scores are selected from the three human-provided answers when evaluating the model performance. Using the above example, if one of the answers has indeed enough happiness, the model can obtain 100% EM and 100% F1 scores, which makes the evaluation more relaxed. Finally, average the whole data set to be evaluated to obtain the final EM and F1 scores (Arch and Stanhope, 2015).

Baseline Model Design

For each sample of SQuAD, the pre-trained GloVe word vector is used to represent the text and problem words. The text is represented by a d-dimensional word vector sequence $x_1, \dots, x_N \in R^d$ with length N, and the problem is represented by a d-dimensional word vector sequence $y_1, \dots, y_M \in R^d$ with length M.

The designed baseline model has three components: RNN encoder layer, which aims to encode problems and text into a hidden state. The interaction layer is used to integrate problems and text representations. The output layer connects two independent softmax layers after a full connection layer, one is used to obtain the starting position of the answer in the text and the other is used to obtain the ending position. **Figure 2** shows the schematic diagram of the baseline model architecture.

Recurrent Neural Network-Based Coding Layer

The word vectors of the text and the question are sent to a single-layer bidirectional cyclic neural network Gated Recurrent Unit (GRU), as shown in the following formula:

$$\{\vec{c}_1, \vec{c}_1, \dots, \vec{c}_N, \vec{c}_N\} = biGRU(\{x_1, \dots, x_N\}) \quad (2)$$

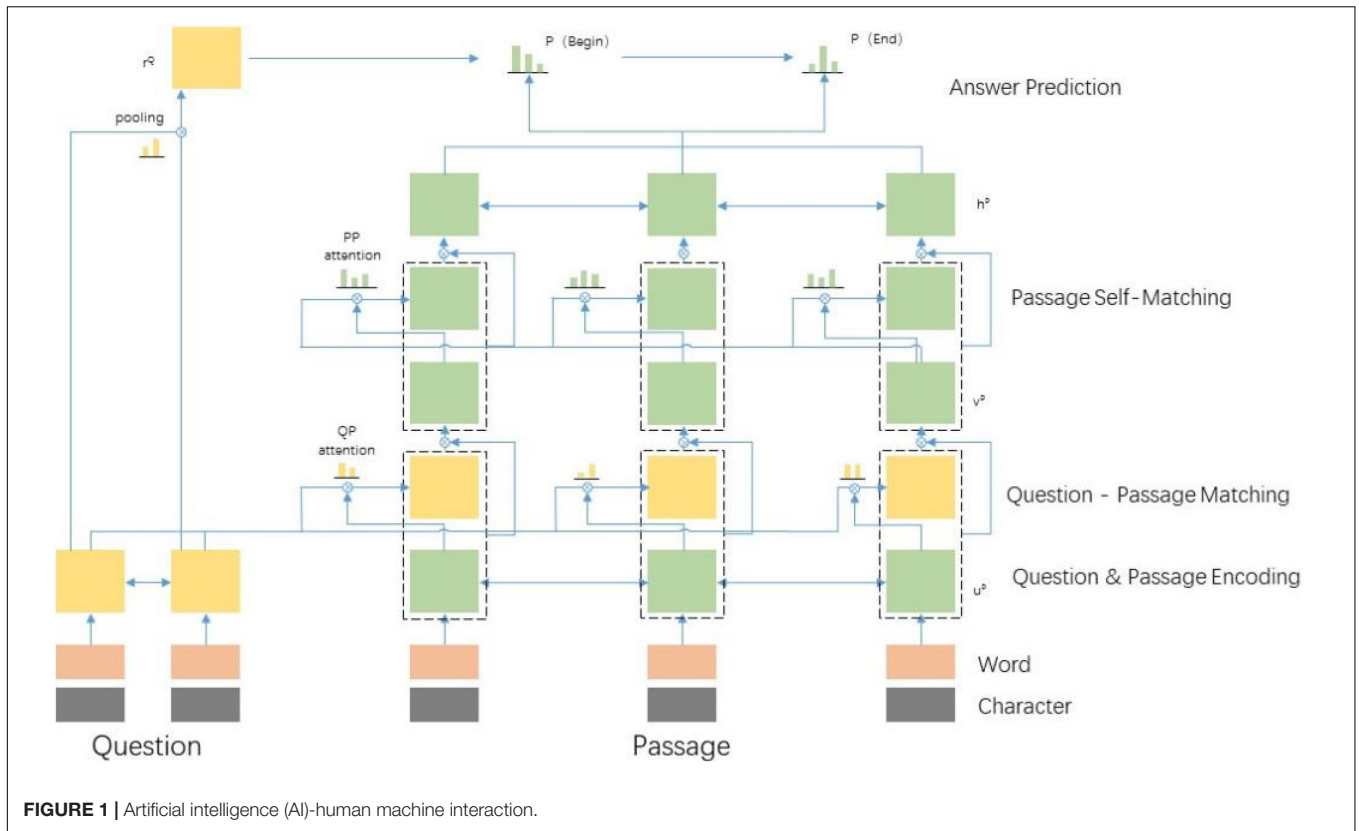


FIGURE 1 | Artificial intelligence (AI)-human machine interaction.

$$\{\vec{q}_1, \vec{q}_1, \dots, \vec{q}_N, \vec{q}_N\} = biGRU(\{y_1, \dots, y_M\}) \quad (3)$$

$$q_j = [\vec{q}_j; \tilde{q}_j] \in R^{2h} \quad \forall i \in \{1, \dots, M\} \quad (5)$$

Gated Recurrent Unit represents the dimension of hidden state. The bidirectional cyclic neural network GRU generates a series of text hidden states ($\vec{c}_i \in R^h$ belongs to forward and $\tilde{c}_i \in R^h$ belongs to backward) and a series of the problem hidden states ($\vec{q}_j \in R^h$ belongs to forward and $\tilde{q}_j \in R^h$ belongs to backward). The forward and backward hidden states are spliced to obtain the text hidden state c_i and problem hidden state q_j respectively:

$$c_i = [\vec{c}_i; \tilde{c}_i] \in R^{2h} \quad \forall i \in \{1, \dots, N\} \quad (4)$$

Interaction Layer

The interaction layer is used to fuse text and problem information. It includes Context2Query from the text perspective and Query2Context from the text perspective. The former is used to obtain which words in which problems the text words focus on, and the latter is used to obtain which words in the text are more important for this problem (Choi et al., 2015).

Specifically, given the text hiding state $c_1, c_2, \dots, c_N \in R^{2h}$ and problem hiding State $q_1, q_2, \dots, q_M \in R^{2h}$ obtained from the coding layer, a similarity matrix $S \in R^{N \times M}$ is constructed, and each element $S_{ij} \in R$ is obtained by the following formula:

$$S_{ij} = w^T [c_i; q_j; c_i \times q_j] \quad (6)$$

Where $w \in R^{6h}$ is a weight vector, \times represents the element level product of two vectors. This similarity matrix s is used to create text to question interactions and problem to text interactions. Text to question interaction focuses on the question word most related to each text word (Jaquette and Curs, 2015). The calculation is as follows:

$$a^i = \text{soft max}(S_i) \in R^M \quad (7)$$

$$a_i = \sum_{j=1}^M a_j^i q_j \in R^{2h} \quad \forall i \in \{1, \dots, N\} \quad (8)$$

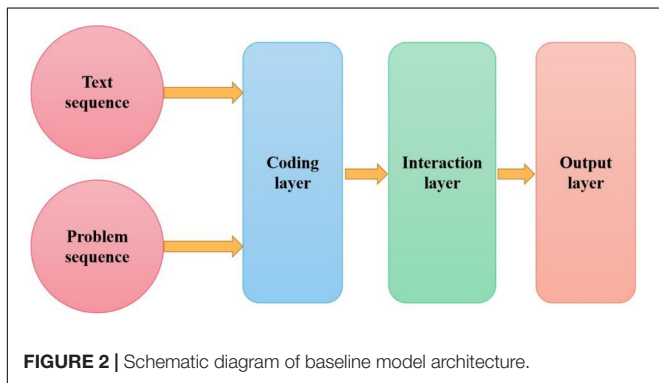


FIGURE 2 | Schematic diagram of baseline model architecture.

The interactive calculation from question to text is as follows:

$$\beta = \text{soft max} \left(\max_j S_{ij} \right) \in R^N \tag{9}$$

$$c' = \sum_{i=1}^N \beta_i c_i \in R^{2h} \tag{10}$$

Representation of the last interaction:

$$b_i = [c_i; a_i; c_i \times a_i; c_i \times c'] \in R^{8h} \quad \forall i \in \{1, \dots, N\} \tag{11}$$

Enter b_i into a two-way GRU to get $b'_i \in R^{2h}, \forall i \in \{1, \dots, N\}$, which contains the interactive information of the corresponding word of the text about the whole text and the problem.

Output Layer

Each union represents b'_i through a full connection layer and uses *ReLU* to activate the function:

$$b''_i = \text{ReLU} (W_{FC} b'_i + v_{FC}) \in R^h \quad \forall i \in \{1, \dots, N\} \tag{12}$$

$W_{FC} \in R^{h \times 2h}$ is the weight of the full connection layer and $v_{FC} \in R^h$ is the offset of the full connection layer. Pass all b''_i through a softmax layer to obtain the probability distribution of the starting position of an answer:

$$\text{scores}_i^{\text{start}} = w_{\text{start}}^T b''_i + u_{\text{start}} \in R \quad \forall i \in [1, \dots, N] \tag{13}$$

$$p^{\text{start}} = \text{soft max} (\text{scores}^{\text{start}}) \in R^N \tag{14}$$

Where $w_{\text{start}} \in R^h$ is a weight vector and $u_{\text{start}} \in R$ is an offset term. Through the same softmax layer, using different parameters w_{end} and u_{end} , the probability distribution of the end position of an answer is obtained.

Loss Function

For the real answer, the starting position in the text is $i_{\text{start}} \in [1, \dots, N]$ and the ending position is $i_{\text{end}} \in [1, \dots, N]$. In the output layer, the probability distributions of the start position and the end position at n positions of the text can be obtained respectively. For a data sample, the cross-entropy loss is used to calculate the loss:

$$\text{loss} = -\log p^{\text{start}}(i_{\text{start}}) - \log p^{\text{end}}(i_{\text{end}}) \tag{15}$$

In the training phase, the average loss is calculated on a batch of data (minibatch).

Forecast

In the prediction stage, given the text and problem of a sample, the output layer of the model can obtain the probability distributions of the start position and the end position at N positions of the text respectively.

First, calculate the indexes idx^{start} and idx^{end} of the start and end positions of the answer in the text as follows:

$$idx^{\text{start}} = \arg \max_{i=1}^N p_i^{\text{start}} \tag{16}$$

$$idx^{\text{end}} = \arg \max_{i=1}^N p_i^{\text{end}} \tag{17}$$

As shown above, using two independent softmax layers to predict idx^{start} and idx^{end} respectively, the idx^{start} given by the model may be greater than idx^{end} , to output a null answer.

Secondly, as a comparison, the length distribution of answers in the training data set is analyzed, and a sliding window with a unified length of W is set. In the prediction stage, given the text and question of a sample, take the first word in the sliding window as the starting position i . Set the other words in the sliding window as the termination position j and the computer probability $p_i^{\text{start}} p_j^{\text{end}}$ in turn and traverse the whole text in the sliding window so that the maximum i and j of $p_i^{\text{start}} p_j^{\text{end}}$ are idx^{start} and idx^{end} , respectively.

Improved Model Design

The text is represented by a d -dimensional word vector sequence $x_1, \dots, x_N \in R^d$ with length N , and the problem is represented by a d -dimensional word vector sequence $y_1, \dots, y_M \in R^d$ with length M . therefore, the text X and problem y of a given sample can be represented as matrices of size $R^{N \times d}$ and $R^{M \times d}$ (Ribeiro et al., 2016).

The length of the filter in the convolutional neural network (CNN) coding layer is equal to the word vector dimension d , and the width is set to k to cover k adjacent words, as shown in **Figure 3**.

The CNN coding layer is connected in series with K convolution layers, excluding the pooling layer. Each convolution layer has 1 filter. Based on the convolution operation of the previous layer, the width of the filter is continuously increased to obtain a wider range of receptive fields in the word dimension and extract a wider range of semantic information. Finally, the operation results of each convolution layer are spliced in the vector representation dimension of the word. The output dimension of text X through the CNN coding layer is $R^{N \times (l \times K)}$, and the output dimension of problem y through the CNN coding layer is $R^{M \times (l \times K)}$.

RESULTS AND ANALYSIS

At present, most AI human-computer interactive reading comprehension data sets are tested online. Therefore, developers cannot get the test set data. This makes it impossible for developers to frequently adjust model parameters and evaluate them on test sets. On the other hand, we observe that the models submitted by many authors have the same effect on the verification set and the test set, which shows that the samples of the verification set and the test set obey the same distribution. Therefore, it is a reasonable way to compare the performance of the model implemented in this manuscript by comparing the evaluation scores on the verification set (Stürmer et al., 2015).

The language model used in this manuscript is BERT, which is better than ELMo and OpenAI human-computer interaction generative pre-training model (GPT). The BERT

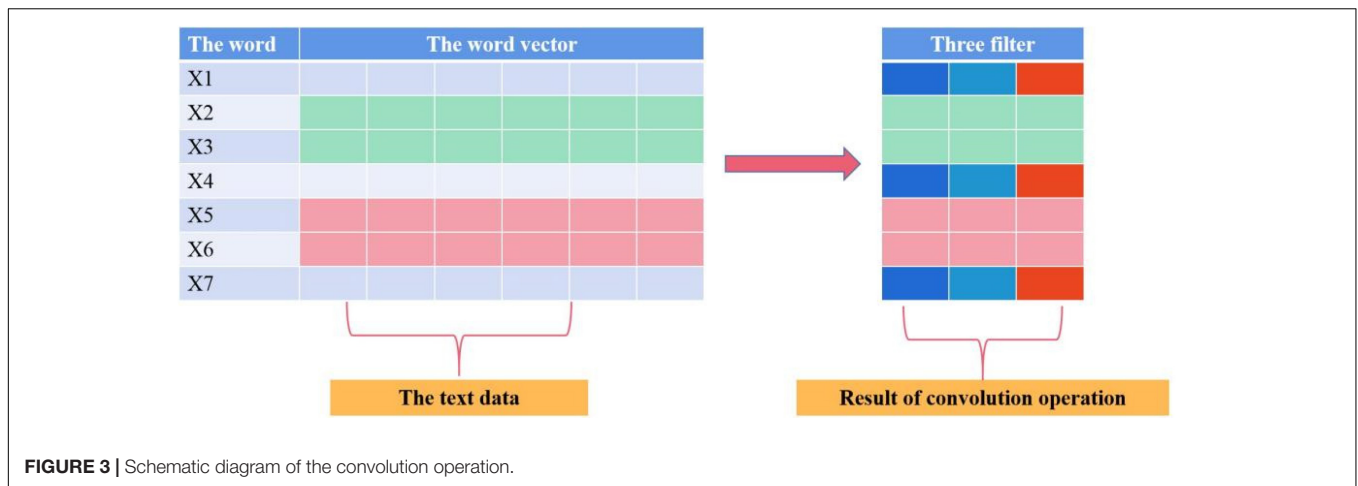


FIGURE 3 | Schematic diagram of the convolution operation.

model architecture uses a two-way multi-layer transformer encoder. The model training is different from the traditional language model to predict the next word as the target task. BERT proposes two new tasks. The first task is to hide 15% of words randomly in the output sequence and predict the hidden words, so that the model can predict the hidden words from any direction. Another task is to enable the model to learn the relationship between sentences and predict the next sentence. Bert can not only be directly applied to downstream tasks in the way of fine-tuning but also can apply the word vector obtained from model pre-training to model word vector initialization in the way of the feature. This manuscript adopts the second method, which uses BERT to extract fixed word vectors, which are obtained by the context representation of fixed length generated by the hidden layer of the pre-trained language model (Yang et al., 2015).

Firstly, the baseline model is reproduced, and the performance of the model is tested on the validation set. The experimental results are shown in Table 1.

From the experimental results, the reproduced results in this manuscript are equivalent to the real results in the manuscript, and the score of F1 value is higher than 0.9% in the manuscript. Because the parameters of the model in the implementation process are not exactly the same as those in the original manuscript, there are also differences in the selection of attention function. Therefore, there are slight changes in the final results, but the overall model effect of the original manuscript is realized (Gaba et al., 2015).

In order to verify the effect of word vectors obtained by BERT on reading comprehension tasks, this manuscript makes a comparative experiment on the effect of BERT. The experimental

results of the improved method of BERT based on the cyclic neural network model are shown in Table 2.

It can be seen from the data in the table that the exact matching value of the cyclic neural network model added with BERT in the verification set has increased by 0.2%, and the F1 value has increased by 2.2%. It can be seen that the word vector generated by the language model trained by unsupervised learning has rich semantics. This context-sensitive word vector is very helpful to improve the effect of natural language understanding task. The word vector obtained by BERT has rich context and contains more lexical and syntactic information than the traditional word vector. In this manuscript, BERT is applied to the cyclic neural network model to verify the role of the original BERT (Sarah et al., 2015).

Then, the attention vector with problem influence based on word level is added to the representation of the text vector, and a three-layer bidirectional gated recurrent unit (BiGRU) network is used to encode the input vector in the coding layer. The network shares parameters between the whole text and the problem. FastText word vector has made three improvements on the basis of word2vec, so that the word vector contains more feature information, which is very helpful to eliminate ambiguity. This manuscript uses a FastText word vector with richer semantics instead of GloVe to calculate the word-level attention. The experimental results show that the FastText word vector is helpful to improve the experimental effect (Srivastava and Wang, 2015). It can be seen that the F1 value of the model is increased by 1.6% compared with the baseline model, which shows that the optimization and improvement of the word vector representation layer of the model can improve the quality of

TABLE 1 | Baseline model evaluation results.

Method	EM	F1
Original paper effect	66.8%	80.6%
This paper realizes the effect	72.8%	81.5%

TABLE 2 | Comparative experimental results based on Bidirectional Encoder Representation from Transformers (BERT).

Method	EM	F1
Original effect	69.3%	80.6%
Reproduction effect	72.4%	83.5%
Improved results based on BERT	72.6%	85.7%

TABLE 3 | Comparison of experimental results based on word vector and context coding.

Method	EM	F1
Original effect	72.6%	80.7%
Reproducing results	71.6%	81.9%
Coding results based on word vector and context	73.6%	83.7%

word vector representation, and then, improve the accuracy of the whole model. The experimental results are shown in **Table 3**.

Statistical Analysis

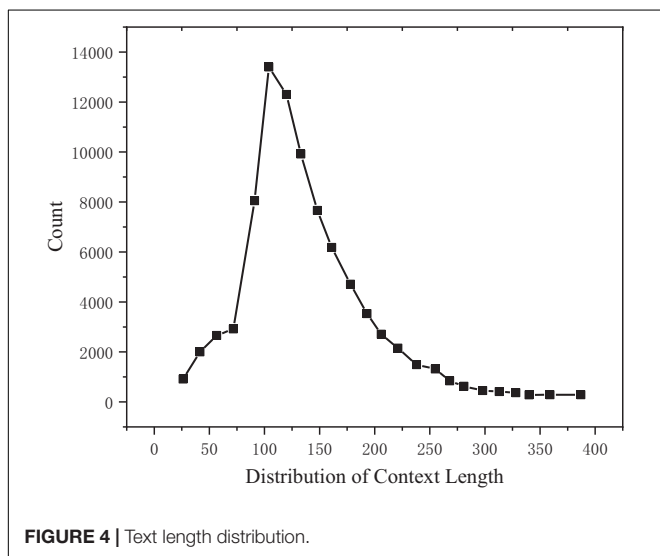
Make statistical analysis on the length of texts, questions, and answers of all samples in the training set. Taking the text length as the abscissa and the text count as the ordinate, it can be observed that the text length presents a right skew distribution, as shown in **Figure 4**. After calculation, the median length is 127, the average is 137, and the maximum value is 766, while 98.35% of the text length is concentrated in the range of 0–300.

Taking the problem length as the abscissa and the problem count as the ordinate, the distribution of the problem length can be observed, as shown in **Figure 5**. After calculation, the median length is 11, the average is 11, and the maximum value is 60, while 99.92% of the problem length is concentrated in the range of 0–30.

Take the answer length as the abscissa and the answer count as the ordinate. The statistical distribution is shown in **Figure 6**. After calculation, the median length is 2, the average is 3, and the maximum value is 46, while 97.5% of the answer lengths are concentrated in the range of 0–15 (Dubreucq et al., 2016).

Model Superparameters

According to the statistical analysis, 300 can cover 98.35% of the text length, the text length n is set to 300, 30 can cover 99.92% of the question length, the question length m is set to 30, 15 can



cover 97.5% of the answer length, and the sliding window length W is set to 15. See **Table 4** for details.

The CNN coding layer of the improved model is stacked with six convolution layers, and a larger filter is used for convolution operation on the convolution output of the previous layer. The super parameters are shown in **Table 5**.

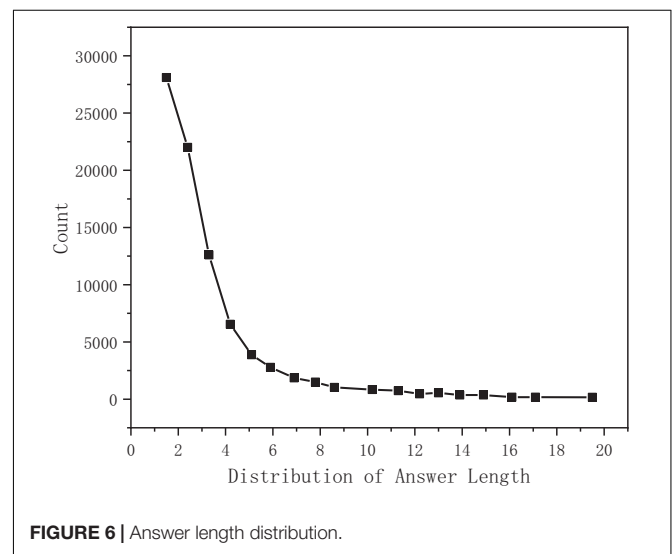
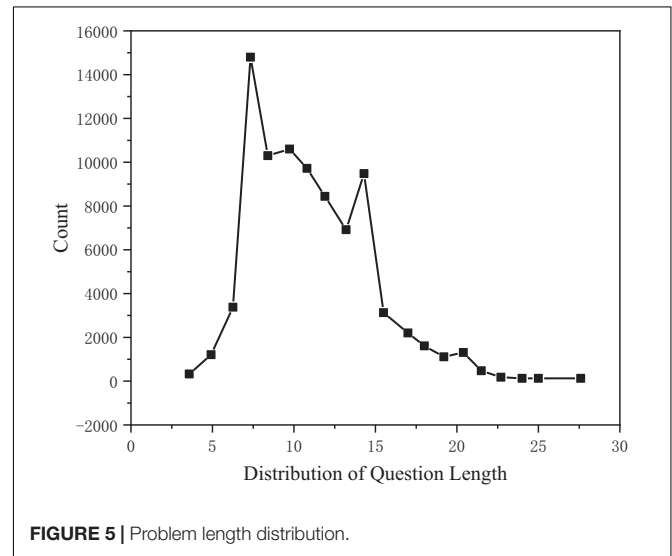


TABLE 4 | Baseline model parameters.

Super parameter	Numerical value
Random inactivation	0.15
Text length N	300
Problem length M	30
Word vector dimension d	100
GRU hidden state dimension H (RNN coding layer)	150
GRU hidden state dimension H (interaction layer)	150
Sliding window length w	15

TABLE 5 | Super parameters of the improved model.

Super parameter	Numerical value
Convolution layer number K	6
Number of convolution layer filters l	50
Convolution layer filter width k	2, 3, 4, 5, 6, 7

Model Training

The training batch size is set to 32, and the maximum gradient norm is set to 5. If it is greater than 5, gradient truncation is performed. Adam algorithm is used for training, and the initial learning rate is 0.001. The total learning rate after the 4,500th iteration of the baseline model is reduced to 0.0005, and the total learning rate after the 6,000th iteration is reduced to 0.0002. Based on the improved model of the CNN coding layer, the total learning rate after 2,500 iterations is reduced to 0.0008, the total learning rate after 11,000 iterations is reduced to 0.0005, the total learning rate after 16,000 iterations is reduced to 0.0002, and the total learning rate after 18,500 iterations is reduced to 0.0001 (Weiss et al., 2015).

Sliding Window

Firstly, verify the effect of the output layer on the baseline model on predicting the answer based on the sliding window and calculate the EM and F1 scores on the verification set respectively. The results are shown in **Table 6**.

Convolutional Neural Network Coding Layer

The baseline model uses RNN coding layer and the improved model uses the CNN coding layer. EM and F1 scores are calculated on the verification set respectively, as shown in **Table 7**.

For different answer lengths, select 1–7 as an example, and the performance of the model will decrease with the increase of the real answer length, as shown in **Figure 7**.

Based on the improved scheme based on word vector representation and context coding, the model is further improved by combining the two-way attention and aggregation mechanism. In order to ensure the comparability of the model, this manuscript does not change the output layer structure of the model and still retains the same output layer structure as experiment 2. Through comparative analysis, the improved effect

TABLE 6 | Effect verification of output layer based on sliding window prediction method.

Prediction mode	F1	EM
Not based on sliding window	68.7	58.0
Based on sliding window	70.7	59.7

TABLE 7 | Comparison of exact match (EM) and F1 scores of baseline model and improved model.

Coding layer	F1	EM
RNN	70.7	59.7
CNN	66.9	54.9

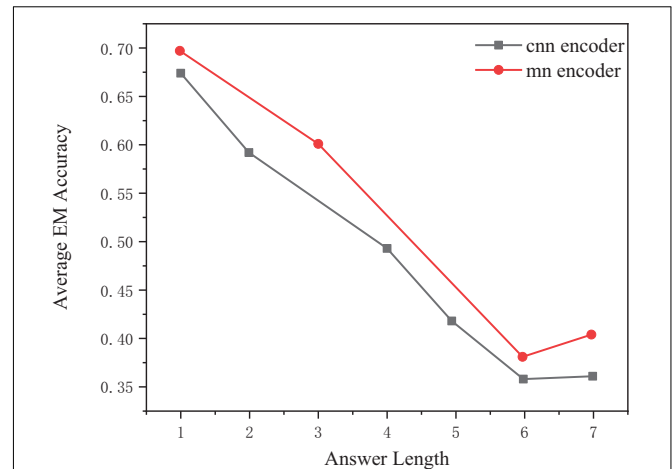


FIGURE 7 | Comparison of exact match (EM) scores of two models on different length answers.

TABLE 8 | Comparison of experimental results based on attention and aggregation mechanism.

Method	EM	F1
Original effect	72%	79.9%
Reproduction effect	70.9%	80.6%
Results based on word vector and context coding	71.9%	83.7%
Results based on attention and aggregation mechanism	76.7%	89.8%

of the algorithm on attention and aggregation mechanism is verified. The experimental data are shown in **Table 8**.

The experimental results in **Table 8** show that after adding the hierarchical attention and aggregation mechanism, the effect of F1 value is improved by 11.8% based on the experiment of improved algorithm based on word vector and context and 7.9% higher than that of the original text. The method based on attention and aggregation mechanism can further improve the model's ability to understand the text. Through attention and aggregation, the model can understand the text at different feature levels. However, the model has not significantly improved the evaluation index of accurate matching, which shows that there is still a certain deviation in the accurate positioning of multiple candidate answers in the text. The advantage of the attention mechanism is that it can locate the text position with the strongest relevance to the question, but the selection of multiple candidate answers depends on the differences of lexical features of the words themselves. The model needs to add more external knowledge to the representation layer to enhance the semantic expression of words. The answer area obtained by the attention mechanism is often accurate. However, to get a more accurate answer, we need more information about the word itself. Finally, combined with the improved methods of word vector representation layer and model interaction layer, this manuscript proposes an improved AI human-computer interactive reading comprehension model architecture. Moreover, further optimizes the reading comprehension model in this manuscript. This

manuscript makes some changes to the output layer of the model based on the baseline model. There are two mainstream methods in the output layer of fragment extraction reading comprehension task. One is to predict the index of the beginning and end position of the answer through the pointer network or combined with the idea of dynamic programming, and the other is to use the two classification methods for prediction. This manuscript balances the advantages and disadvantages of the two methods and uses binary classification to replace the baseline model and uses pointer network to predict the answer. This prediction method may make the model lose certain accuracy, but the efficiency of model training and prediction will be improved a lot (Holmes et al., 2016).

After 45k iteration, the model tends to peak, and the model remains stable in the subsequent 15K iteration. In this manuscript, the experimental parameter values are updated every 1,000 steps, and the whole experimental training process is visualized by tensorboard. It shows that the output layer has a certain impact on the effect of the model, and the improvement and optimization of the output layer can also improve the performance of the model. In addition, this manuscript sacrifices a lot of time in exchange for the accuracy of the model when the experimental hardware allows, which is not desirable in practical engineering. Therefore, how to balance the relationship between model accuracy and training efficiency is a direction of follow-up improvement. In comparison with other public models, it can be seen that the performance of the AI human-computer interactive reading comprehension model implemented in this manuscript is higher than some classical models.

Finally, the comparative analysis shows that the word vectors extracted by language model pre-training can improve the effect of the model, which shows that the word vectors generated by language model pre-training have all rich semantics. At the same time, this manuscript uses the attention mechanism at different levels of the text. This manuscript uses word-level attention in the presentation layer and two-way attention and self-attention models in the model interaction layer. This multi-level attention mechanism can improve the semantic understanding ability of the model. Through experimental analysis, the improved methods of presentation layer and interaction layer can improve the effect of the model to varying degrees. Moreover, the accuracy of the improved AI human-computer interactive reading comprehension model is also better than the traditional model.

CONCLUSION

Artificial intelligence human-computer interactive reading comprehension tasks have various forms. The SQuAD limits the answer to a fragment of the text. The existing models encode the text and questions based on a cyclic neural network to fuse

the semantic information of the context. The main purpose of this manuscript is to try the coding layer based on multi-layer convolution operation. Through experimental analysis, combined with the improved model based on multi-layer convolution operation coding layer, the performance of the given evaluation index is slightly lower than the baseline model, but it is superior in the number of parameters and iteration speed, and the model design is reasonable. For the coding layer based on multi-layer convolution operation, the super parameter setting, such as the number of convolution layers, the number of filters, and the structure design, needs to be further explored. The SQuAD is a classic data set. The reading comprehension form of fragment extraction is relatively simple, while the application of AI human-computer interactive reading comprehension in the search engine, intelligent customer service, and other fields is more generative. Therefore, in the real world, AI human-computer interaction will face more complex and changeable data forms and a larger amount of data. The model needs a more detailed design and takes into account computational efficiency. In addition, human beings solve reading comprehension problems through reasoning, judgment, and induction and can express the thinking process. It is worth further expanding the traditional methods, such as visualization, to improve the interpretability of the model. Finally, with the rise of reinforcement learning, it also has some applications in the field of AI human-computer interactive reading comprehension. How to better combine it with practice and implement it to the application level is worth further exploration.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

GS: editing, writing, and data analysis.

FUNDING

The work was supported by the Innovation Funds Plan of Henan University of Technology (Grant No. 2021-SKCXTD-13), by the Henan Provincial Department of Education -“Henan First-class Post Graduates Course Project – Culture Translation (YJS2021KC13)”, by the Henan Provincial Department of Education-“Henan First-class Undergraduates Course Project (Chinese-English Interpreting)”, and by the Henan University of Technology (Grant No. 2020JKYB05).

REFERENCES

- Arch, E. S., and Stanhope, S. J. (2015). Passive-dynamic ankle-foot orthoses substitute for ankle strength while causing adaptive gait strategies: a feasibility study. *Ann. Biomed. Eng.* 43, 442–450. doi: 10.1007/s10439-014-1067-8
- Choi, J., Jeong, J., Park, Y. I., and Cha, S. W. (2015). Evaluation of regenerative braking effect for e-rev bus according to characteristic of driving cycle. *Int. J. Precision Eng. Manuf. Green Technol.* 2, 149–155.
- Comunian, R. (2015). Rethinking the creative city: the role of complexity, networks and interactions in the urban creative economy. *Urban Stud.* 48, 1157–1179.

- Dubreucq, J., Delorme, C., and Roue, R. (2016). Metacognitive therapy focused on psychosocial function in psychosis. *J. Contemp. Psychother.* 46, 1–10. doi: 10.1111/j.1751-7893.2011.00299.x
- Gaba, S., Lescourret, F., Boudsocq, S., Enjalbert, J., Hinsinger, P., Journet, E. P., et al. (2015). Multiple cropping systems as drivers for providing multiple ecosystem services: from concepts to design. *Agron. Sustain. Dev.* 35, 607–623. doi: 10.1007/s13593-014-0272-z
- Holmes, R., Armanini, D. G., and Yates, A. G. (2016). Effects of best management practice on ecological condition: does location matter? *Environ. Manag.* 57, 1062–1076. doi: 10.1007/s00267-016-0662-x
- Jaquette, O., and Curs, B. R. (2015). Creating the out-of-state university: do public universities increase nonresident freshman enrollment in response to declining state appropriations? *Res. High. Educ.* 56, 535–565. doi: 10.1007/s11162-015-9362-2
- Khalifian, S., Sarhane, K. A., Tammia, M., Ibrahim, Z., Mao, H. Q., Cooney, D. S., et al. (2015). Stem cell-based approaches to improve nerve regeneration: potential implications for reconstructive transplantation? *Arch. Immunol. Ther. Exp.* 63, 15–30. doi: 10.1007/s00005-014-0323-9
- Kraska-Szlenk, I. (2018). Address inversion in swahili: usage patterns, cognitive motivation and cultural factors. *Cogn. Linguist.* 29, 545–583.
- Li, J., Zhang, Y., Qian, C., Ma, S., and Zhang, G. (2020). Research on recommendation and interaction strategies based on resource similarity in the manufacturing ecosystem. *Adv. Eng. Inform.* 46:101183.
- Ma, Y. (2021). The application of schema theory in the teaching of english reading in senior high schools. *Region Educ. Res. Rev.* 3, 17–20. doi: 10.32629/rerr.v3i3.412
- Martínez-Álvarez, I., Mateos, M., Martín, E., and Rijlaarsdam, G. (2015). Learning history by composing synthesis texts: effects of an instructional programme on learning, reading and writing processes, and text quality. *J. Writing Res.* 7, 275–302.
- Rasinski, T. V., Rupley, W. H., Paige, D. D., and Nichols, W. D. (2016). Alternative text types to improve reading fluency for competent to struggling readers. *Int. J. Instr.* 9, 163–178.
- Renzulli, S. J. (2015). Using learning strategies to improve the academic performance of university students on academic probation. *Nacada J.* 35, 29–41.
- Ribeiro, M., De Almeida, A. A. F., Alves, T. F. O., Gramacho, K. P., Pirovani, C. P., and Valle, R. R. (2016). Rootstock × scion interactions on theobroma cacao resistance to witches' broom: photosynthetic, nutritional and antioxidant metabolism responses. *Acta Physiol. Plant.* 38, 1–14.
- Santos, F., and Canello, J. (2015). Brazilian congress, 2014 elections and governability challenges. *Brazil. Polit. Sci. Rev.* 9, 115–134. doi: 10.1590/1981-38212014000200005
- Sarah, G. A., Jaclyn, D., Waylon, B., Shahar, L., Mitch, S., Clarice, L., et al. (2015). Mpneumonia: development of an innovative mhealth application for diagnosing and treating childhood pneumonia and other childhood illnesses in low-resource settings. *PLoS One* 10:e0139625. doi: 10.1371/journal.pone.0139625
- Scammacca, N. K., Roberts, G., Vaughn, S., and Stuebing, K. K. (2015). A meta-analysis of interventions for struggling readers in grades 4–12: 1980–2011. *J. Learn. Disabil.* 48, 369–390. doi: 10.1177/0022219413504995
- Srivastava, M. K., and Wang, T. (2015). When does selling make you wiser? impact of licensing on chinese firms' patenting propensity. *J. Technol. Trans.* 40, 602–628.
- Stürmer, K., Könings, K., and Seidel, T. (2015). Factors within university-based teacher education relating to preservice teachers' professional vision. *Vocat. Learn.* 8, 35–54. doi: 10.1007/s12186-014-9122-z
- Ulaiwy, A. T., and Alarnoosy, D. U. H. (2018). Effect of teaching the reading and texts by the two strategies of (slap & csr) on expressional performance of 2 nd intermediate grade students. *Indian J. Public Health Res. Dev.* 9:1271. doi: 10.5958/0976-5506.2018.02027.2
- Ungureanu, C. R. M., Poiana, M. A., Cocan, I., Lupitu, A. I., and Moigradean, D. (2020). Strategies to improve the thermo-oxidative stability of sunflower oil by exploiting the antioxidant potential of blueberries processing byproducts. *Molecules* 25:5688. doi: 10.3390/molecules25235688
- Weiss, B., Wechsung, E. C., Hsung, I., Kühnel, C., and Mller, S. (2015). Evaluating embodied conversational agents in multimodal interfaces. *Comput. Cogn. Sci.* 1:6. doi: 10.1007/s10916-020-01624-4
- Wu, D., and Zhang, X. (2021). Practical research on computer-based children's literature to improve kindergarten teachers' humanistic quality. *J. Phys.* 1744:e032054.
- Xiong, Y., Li, L., and Gao, S. (2016). Research on the strategies of improving college students' learning initiative in the internet environment. *Psychol. Res.* 6:5.
- Yang, L., Qin, Z., and Tu, L. (2015). Responses of rice yields in different rice-cropping systems to climate variables in the middle and lower reaches of the yangtze river, china. *Food Security* 7, 951–963.

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Shen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.