Intelligent Question Answering in Restricted Domains Using Deep Learning and Question Pair Matching

LIN-QIN CAI, (Member, IEEE), MIN WEI, SI-TONG ZHOU, AND XUN YAN

Key Laboratory of Industrial Internet of Things and Networked Control, Ministry of Education, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

Corresponding author: Min Wei (waoumin@163.com)

ABSTRACT With the rapid expansion of the Internet, intelligent question answering for information retrieval has once again gained widespread attention. However, current question answering models mainly focus on the general and common-sense questions in open domains and are incapable to effectively solve more complex professional domain questions. This paper proposed an integrated framework for Chinese intelligent question answering in restricted domains. The proposed model fused convolutional neural network and bidirectional long short-term memory network which performs efficient semantic analysis on the question pairs to extract more effective features of the text. Meanwhile, the coattention mechanism and attention mechanism were combined to obtain the semantic interaction and feature representation of the question pair for providing complete information for subsequent calculations. In addition, we introduced the method of question pair matching to implement the Chinese intelligent question answering in a restricted domain. Experiments were tested and evaluated on the open-source CCKS2018 dataset and our private self-built inverted pendulum control question answering (IPC-QA) dataset for automation control virtual learning environment. Experimental results confirm that the proposed models are efficient and achieve a high precision of 0.86042 and 0.8031 on CCKS2018 and IPC-QA respectively.

INDEX TERMS CNN, BiLSTM, attention mechanism, question pair matching, intelligent question answering.

I. INTRODUCTION

As the data information of the Internet grows exponentially, people have an increasing number of ways to obtain information on the Internet, like Baidu, Google and other traditional search engines providing people with numerous information of varying quality [1]. Intelligent question answering is applied to websites, various online communities, and chatbots to offer more accurate and relevant information. Intelligent question answering is generally defined as that the user types the question in natural language form, and the question answering system outputs a streamlined answer or a list of possible answers through the model based on the background knowledge base, rather than returning a bunch of related

The associate editor coordinating the review of this manuscript and approving it for publication was Luigi De Russis.

documents to the user [2]. As the current intelligent question answering mainly focuses on the open domains, for general and common-sense questions, there are main methods based on surface pattern matching, syntactic structure comparison, massive data redundancy, logical reasoning of answers, and multi-feature statistical machine learning. These methods have defects such as incomplete matching and insufficient semantic connection, while deep learning can effectively solve the above problems.

Most of previous intelligent question answering systems used machine learning to analyze and retrieve texts with applying the knowledge base to answer common questions [3]–[5], which required a large amount of annotation data, and could not effectively deal with questions with complex semantic structures. At the same time, machine learning has a poor representation for the semantic information relevance of

questions and answers. With the re-development of artificial intelligence, deep learning has become the major research method in intelligent question answering. Mikolov *et al.* used a neural network model to obtain a new vector representation called word embedding [6], which is a low-dimensional, dense, continuous vector representation and contains both the semantic and grammatical information of the words. Deep neural networks such as recurrent neural network (RNN) have outstanding performance in uncertainty modeling, which can be applied to tasks such as video captioning [7]. On the basis, relevant researchers designed deep neural network models to acquire the vector representation of sentences, such as sentence modeling by RNN and convolutional neural network (CNN) [8], [9]. At the AAAI 2016 conference, in order to extract valuable social information in addition to text content, Fang *et al.* (2016) added heterogeneous social network to RNN to import rich user interaction information, effectively solving the problem of semantic sparsity in intelligent question answering [10]. On a deeper level, answers' authority of the given question also plays an important role in question answering. In [11], Zhao *et al.* (2017) formulated a community question answering system from the viewpoint of asymmetrical multi-faceted ranking network learning, providing a certain idea for professional question answering in restricted domains. Nevertheless, most of these models did not pay much attention to the important information in the texts. At the same time, long short-term memory network (LSTM) performs well in image and video tasks due to its sensitivity to timing information, which provides inspiration for the application of LSTM in natural language processing [12]. In recent years, bidirectional long short-term memory network (BiLSTM) combined with the attention mechanism has been widely concerned by researchers [13]. BiLSTM network can better construct the vector representation of the answers based on the input questions, but the model ignores the influence of the answers on the question vector representations and results in certain result deviations. On the other hand, at present the open domain intelligent question answering has made remarkable progress in the field of deep learning research. However, questions in a restricted domain are difficult to process because the related knowledge base contains specialized and complex specific corpus data of the corresponding domain, which is a bottleneck hindering the development of intelligent question answering. As the knowledge system has strong logic and truth in a particularly professional field, the standard answers corresponding to the questions are mostly limited and unique, and the questions corresponding to a specific answer have multiple representations. The method of modeling and matching between question and question is used to solve the dilemma of one answer corresponding to multiple question representations in a restricted domain. The main advantages are as follows: (1) The stability of corpus. In the restricted domain where the underlying knowledge base or dataset domain is vertical and closed, the question corresponds to the answer one to one, and the data size is delineated in an inductive

range, while the user-input questions have different forms and expressions with a tremendous number [14]. By modeling the question pair, the input question is matched with the existing question, then the corresponding answer output is obtained. (2) Semantic space aspect. There may be a semantic gap when coding between the questions and answers. Whereas when multiple questions correspond to a same answer, their starting point and form in the learning process are not far from each other, thus their semantic space is consistent. (3) On-line running speed. The questions corresponding to similar dense vectors in the model can be indexed with tools such as artificial intelligence markup language (AIML) through the idea of question pair matching, saving computing resources and increasing the running speed.

The main contributions of this paper are summarized as follows:

(1) An integrated framework for Chinese intelligent question answering in restricted domains was presented based on CNN-BiLSTM network, coattention mechanism and attention mechanism. We fused CNN and BiLSTM to extract and represent important textual information and contextual information. The coattention mechanism and attention mechanism were combined to obtain the semantic interaction and feature representation of the question pair for providing complete information for subsequent calculations. Additionally, we introduced the method of question pair matching to implement the Chinese intelligent question answering in a restricted domain. The proposed methods can not only effectively achieve a high accuracy, but also improve resource utilization.

(2) We constructed an inverted pendulum control question answering (IPC-QA) dataset for automation control virtual learning environment to test our proposed methods. The IPC-QA is a typical restricted-domain teaching question answering dataset and contains 9450 question pairs, which involves the main inverted pendulum control experiments in *Automatic Control Theory* course.

The rest chapters of this paper are organized as follows: Section II briefly reviews related work. Section III presents the main frame model of the paper. In Section IV, the datasets and related parameter settings used in the experiments are introduced in detail. Section V is a detailed analysis and summary of the experimental results. We will summarize in Section VI to discuss further work in the future.

II. RELATED WORK

The introduction of the Question Answering Track (QA Track) in the Text Retrieval Conference (TREC) in 1999 has greatly promoted the research and development of natural language processing (NLP) technology in the field of question answering, and numerous scholars have proposed many effective models [15]–[18]. Based on predecessors' work, this section will emphasize the research methods of Chinese intelligent question answering in restricted domains and the advantages of the question pair matching idea,

CNN and BiLSTM, coattention mechanism and attention mechanism in other literatures.

A. CHINESE RESTRICTED DOMAINS AND QUESTION PAIR MATCHING

In 1961, the first intelligent question answering system in restricted domains appeared abroad. The system was built by Green and provided question answering services for the baseball field. In the following decades, the intelligent question answering in restricted domains has received extensive attention. Among them, Pathak and Mishra (2016) followed the principles of information retrieval and natural language processing to propose a question answering system for tourism containing the information about specific tourist places, with precision and accuracy of 70% and 80%, respectively [19]. Although the domestic Chinese intelligent question answering in restricted domains started late, it has achieved continuous development. Similar with [19], Sun *et al.* (2007) [20] focused on intelligent question answering in tourism domain, and used 5 subsystems to form InsunTourQA system, achieving question answering in the specific domain of Chinese tourism. Traditional research methods of intelligent question answering are mainly based on the syntactic matching of question and answer. Moreda *et al.* (2011) [21] presented two proposals for the answer extraction module of question answering system according to semantic information, semantic roles and WordNet. Zhu *et al.* (2011) [22] designed and implemented a Chinese intelligent question answering system based on domain ontology and sentence template, using domain ontology as the knowledge base to supply domain vocabulary to question analysis and answer generation, then conducting knowledge retrieval. Kim *et al.* (2014) [23] improved the baseline system, combining simple rules and unsupervised learning models with deep linguistic features, to choose yes or no answers for questions in the legal domain. Zhou *et al.* (2016) [5] used a vocabulary model based on word relationships to select answer sentences. However, most of these models are derived from the idea of question and answer matching, with huge data computation and excessive dependence on external conditions. For instance, manually marking information and constantly changing external conditions require too much related work to implement. Therefore, Hao and Agichtein (2012) [24] proposed a precise approach of automatically finding an answer to input questions by automatically identifying "equivalent" questions submitted and answered in the past. By means of automatically generating equivalent question patterns, the model achieved over 57% recall and over 54% precision, but still did not completely solve the defects of question and answer matching. After a long period of research, Palmera and Figueroa (2017) [25] matched the intention of the newly published question with the intention of the archived answer presented to the questioner. By manually annotating the number of how-to questions and answers, the accuracy of the best answer retrieval was increased by 4.12%. Inspired by Palmera and Figueroa [25], in this paper, we used the method

of question pair matching to solve the difficulty of the large number and great change of external input questions of Chinese restricted domain in the professional field, and significantly improved corresponding accuracy.

B. CNN AND BILSTM

Since Krizhevsky *et al.* [26] used the extended CNN in the ImageNet Large Scale Visual Recognition Challenge (LSVRC) in 2012 to achieve the best classification effect at the time, the convolutional neural network has once again set off a research boom. As a typical network structure in deep learning, CNN adopts a local perceptual strategy, which can greatly simplify the complexity of traditional neural networks [27]. Jin *et al.* (2017) [28] used CNN to conduct semantic representation and semantic matching of texts. Zhou *et al.* (2018) [29] combined CNN with RNN to capture the semantic matching between questions and answers and the semantic relevance embedded in the answer sequences, achieving the best performance of Macro-F1 at 58.77%. CNN can fully consider the order of sentences in the semantic representation of text and the semantic matching of text, and can take sentence semantics into account from many aspects. Yet, it only combines consecutive phrases and only considers the word order in the sliding window, resulting that it is impossible to capture long-distance dependencies and can't represent complex semantic sentences well.

In recent years, long short-term memory network has been favored to capture contextual semantic relationships [12], [13]. Palangi *et al.* (2016) [30] proposed a text matching model LSTM-RNN based on long short-term memory network. Cheng and Hu (2018) [14] constructed a network model using conditional random field (CRF) and LSTM network to improve information extraction performance and establish a question answering system over knowledge base for the machinery industry. Gao *et al.* (2019) [31] integrated attention mechanism with hierarchical LSTMs for video and image caption tasks. Inspired by the above works, we combined CNN and BiLSTM to effectively capture data and information of local keywords in the sentence. At the same time, the model can also achieve the preliminary analysis and understanding of the relevance of each word in the Chinese sentence.

C. COATTENTION MECHANISM AND ATTENTION MECHANISM

The attention mechanism is used to assign different weights to content vectors of distinct importance according to diverse scenarios, which is beneficial to extract key information in the text, thereby expressing more effectively. Researchers have conducted plentiful researches on the application of attention mechanism in intelligent question answering. Bahdanau *et al.* (2014) [32] added attention mechanism to the model based on bidirectional recurrent neural network to encode and decode sentences in machine translation. Enlightened by the work of Bahdanau *et al.* [32] and Fu *et al.* (2016) [33], Tan *et al.* (2016) [34] and
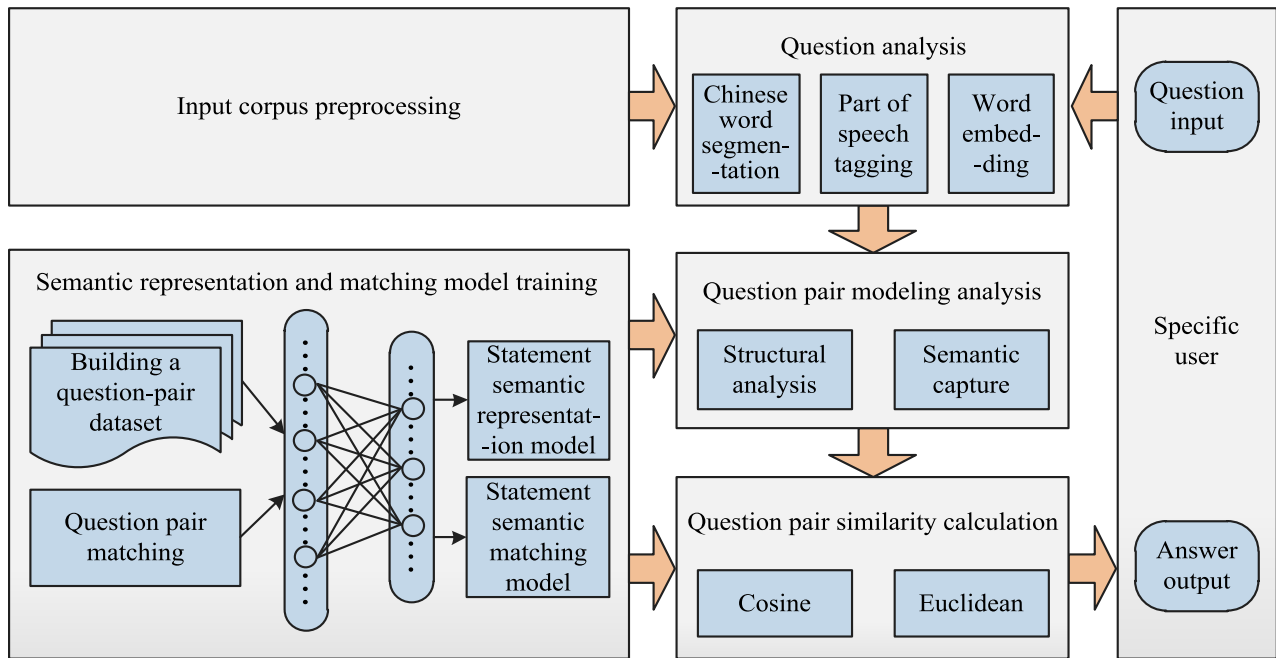
FIGURE 1. The framework of Chinese intelligent question answering in restricted domains based on question pair matching.

Xiang et al. (2017) [35] introduced an attention mechanism fusing bidirectional single layer LSTM for question and answer matching, constructing a better answer representation according to input question. Zhang et al. (2018) [36] studied the internal and external attention mechanisms of discourse representation in implicit discourse relationship recognition. Then, the coattention mechanism has attracted the attention of researchers due to its prominent role in semantic relevance [37]. Yang et al. (2019) [38] developed a coattention mechanism using an end-to-end deep network architecture to jointly learn image and question features, implementing a challenging visual question answering (VQA) task. Motivated by Yang et al. [38], in this paper we combined coattention mechanism and attention mechanism to learn the features of the input question and the existing question together to obtain the interactive vector representation of the question pair. We improved the accuracy of the experiment by making use of the internal semantic connections of the question pairs.

III. PROPOSED MODEL FOR QUESTION ANSWERING
The Chinese intelligent question answering based on question pair matching is a progressive process. Our proposed framework for Chinese intelligent question answering in restricted domains is shown in Fig. 1. After the input corpus is preprocessed by the Chinese word segmentation tool, we encode and extract features of different expressions of the same target. Firstly, it needs a well-organized underlying question answering knowledge base to train the sematic representation and matching model for the question pairs. Secondly, we analyze the semantics and structure of the question pairs to construct the feature vector model. This is a crucial step.

Finally, the matching algorithm is deployed to calculate the similarity between the input question vector and the existing question vector, then outputting the answer.

The model of Chinese sentence semantic feature extraction is constructed in Fig. 2. Firstly, the sentences after word segmentation are sent to the word2vec model trained by the self-built corpus dataset for encoding. With a layer of convolutional neural network, the model can fully consider the word order and semantic context of the input sentence. Then the local key information of the statement is extracted and compressed into a fixed length to prepare for semantic interaction and representation of question pair in subsequent network. Next, semantic information is extracted and feature vectors are further represented by combining stacked BiLSTM network and coattention mechanism. This method not only can solve the dependence problem between the before and after words in the long statements, but also can obtain the related feature representation between the question pair. Finally, the output is predicted by the softmax function.

Chinese sentences are more complex and diverse in expression, and there are also certain difficulties in the specific lexical parsing and coding of restricted domains. Compared with the traditional feature extraction models, the network architecture in Fig. 2 can solve the problem of Chinese semantic understanding. CNN is used to capture crucial feature information, while the BiLSTM network is used to obtain the semantic analysis of the entire sentence.

A. SEMANTIC ANALYSIS WITH THE FUSION STRUCTURE OF CNN AND BILSTM
The relevance of words in Chinese expression is very close, and the analysis of sentence semantics and feature extraction include the sequence coding of each word and the weight
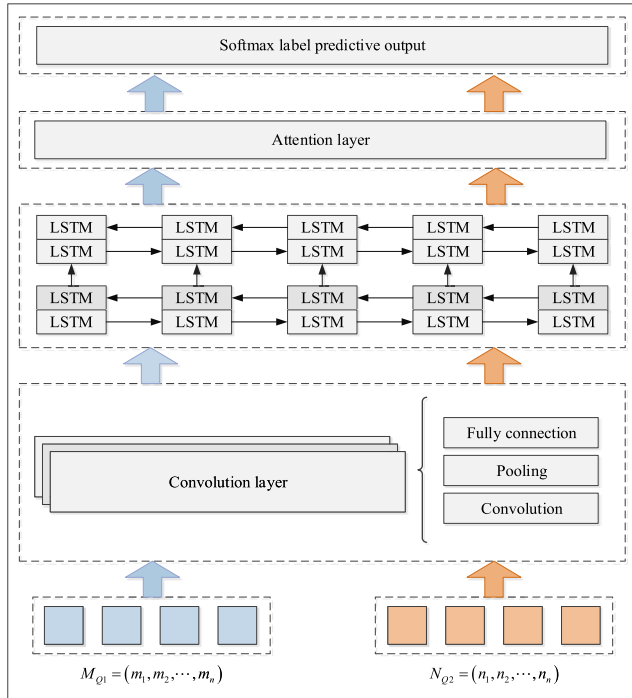
FIGURE 2. Feature extraction model for integrated CNN-BiLSTM network.

assignment among words. Therefore, inspired by related researches [39]–[41], this paper uses the fusion structure of CNN and BiLSTM in the framework of question-sentence feature extraction to calculate the correlation importance of the words and map the local depth of the shallow features in the sentences.

In the CNN network layer, the initial input obtains the preliminary parsing vector of the text mainly through the two processes of convolution and pooling. In the experiment, its internal structure is illustrated in Fig. 3.

The input of the CNN network is the $n \times k$-order word matrix generated by the pre-trained word2vec model, where n is the number of all words in the sentence, i.e., the maximum length of the sentence, and k is the vector dimension corresponding to each word. In this paper, the dimension of the word2vec model is trained as $k = 300$. We set $x_i \in \mathbb{R}^k$ as the k-dimensional word vector of the i-th word of the sentence, where the vector of the unregistered word is filled by the zero-padding method, then the sentence dense vector of length n is expressed as:

$$x_{1:n} = x_1 \oplus x_2 \oplus \cdots \oplus x_n \qquad (1)$$

where $\oplus$ represents the word vector parallel concatenation operator and $x_{a:b}$ represents the word vector matrix composed of $\{x_a, x, \ldots, x_b\}$.

In the process of natural text processed by convolutional neural network, the convolution kernel generally covers the words of the upper and lower several lines of a statement, so the convolution kernel of a text has only length but no width. We set the convolution kernel window $W \in \mathbb{R}^{h \times k}$ with

a fixed length h ($h \times k$, the longitudinal h, and the transverse k represent the convolution window size, the number of words contained in the sliding window, and the dimension of the word vector, respectively) to convolve the words contained in $x_{i:i+h-1}$ in the sentence. The feature operation of the convolution kernel window on the i-th word of the sentence is calculated as follows:

$$c_i = \sigma(W \cdot x_{i:i+h-1} + b) \qquad (2)$$

where c_i, σ, $x_{i:i+h-1}$, and b represent the feature output of the word i after convolution, the activation function, the matrix composed of the word vectors from i to $i + h - 1$, and the bias factor, respectively; W is the operational matrix of the convolutional layer, i.e., the sliding window.

By analogy, when the convolution kernel window with longitudinal length h is acted on the entire sentence with length n, $n - h + 1$ new feature vectors can be obtained for the statement, which form the one-dimensional feature map of the corresponding statement. The specific calculation is as follows:

$$c = [c_1, c_2, \ldots, c_{n-h+1}] \qquad (3)$$

Pooling is a downsampling process for the convolutional layer, which can reduce the size of data to be processed while effectively extracting and retaining important semantic features of statements. In this paper, we extract the maximum eigenvalues of the convolution kernel vectors after convolution by the 1-max pooling strategy, and combine them as input of the BiLSTM network. The calculation is as follows:

$$C = \max[c_1, c_2, \ldots, c_{n-h+1}] \qquad (4)$$

When the 1-max pooling technique is applied to the convolution kernel vector, each feature map only outputs a maximum value. This method can downsample all the dimensions of the feature map output from the convolutional layer to 1. Consequently, the final combined output vector C obtained after pooling is a fixed-size vector matrix whose dimension is the number of feature maps output from convolutional layer.

Through convolution and pooling, the vector matrix of the statement performs initial semantic analysis and feature extraction to obtain local important information features of the statement and effectively reduce the training parameters.

After extracting the structured semantic information with representational ability in the sentences of question pair $Q = (q_1, q_2, \ldots, q_n)$ and $Q' = (q'_1, q'_2, \ldots, q'_m)$ through CNN network, the output is expressed as $C_Q = \max[c_1, c_2, \ldots, c_{n-h+1}]$ and $C_{Q'} = \max[c'_1, c'_2, \ldots, c'_{m-h+1}]$, where n and m are the numbers of words contained in question Q_1 and question Q_2, respectively, namely the sentence length, and h is the convolution window size. The next step is to transfer C_Q and $C_{Q'}$ to BiLSTM network to extract the interdependencies and influences of the distant words in the sentence.

A single LSTM unit can only capture the semantic information of the first half of the sentence in which the word is
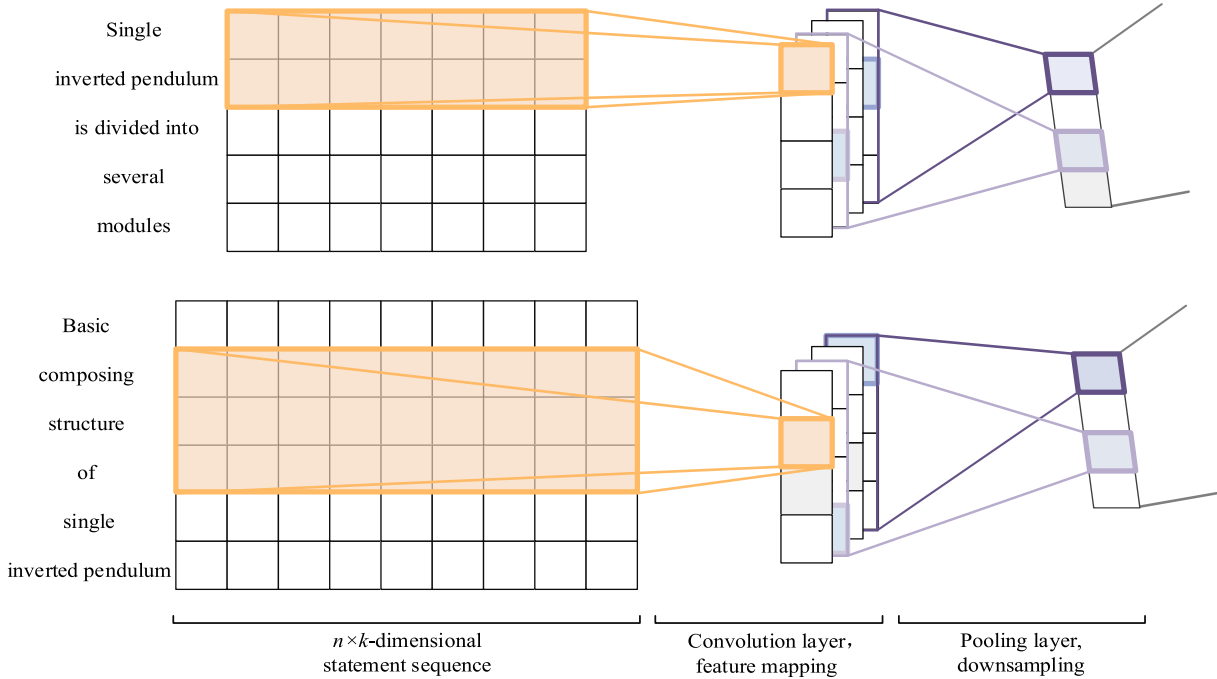
FIGURE 3. Schematic diagram of feature extraction of sentence words based on CNN.

located while the second half of the semantic information cannot be captured. In order to overcome the drawback, Schuster and Paliwal [42] designed a semantic capture model containing the LSTM hidden layer with two opposite directions. The forward and backward LSTM units simultaneously traverse the statement from the beginning to the end so that the forward neuron hidden layer output sequence $\overrightarrow{h}_t = (\overrightarrow{h}_1, \overrightarrow{h}_2, \ldots, \overrightarrow{h}_n)$ and the backward neuron hidden layer output sequence $\overleftarrow{h}_t = (\overleftarrow{h}_1, \overleftarrow{h}_2, \cdots, \overleftarrow{h}_n)$ of the sentence sequence $x = (x_1, x_2, \ldots, x_n)$ can be respectively obtained. The coded output y_t of the hidden layers of bidirectional LSTM network can be obtained by cascading the forward and backward outputs, $y_t = [\overrightarrow{h}_t, \overleftarrow{h}_t]$.

$$\overrightarrow{h}_t = \sigma(W_{\overrightarrow{h}x}x_t + W_{\overrightarrow{h}\overrightarrow{h}}\overrightarrow{h}_{t-1} + b_{\overrightarrow{h}}) \qquad (5)$$

$$\overleftarrow{h}_t = \sigma\left(W_{\overleftarrow{h}x}x_t + W_{\overleftarrow{h}\overleftarrow{h}}\overleftarrow{h}_{t-1} + b_{\overleftarrow{h}}\right) \qquad (6)$$

$$y_t = W_{y\overrightarrow{h}}\overrightarrow{h}_t + W_{y\overleftarrow{h}}\overleftarrow{h}_t + b_y \qquad (7)$$

where W and b represent the weight vectors and bias terms corresponding to the three gates of the LSTM, respectively.

In relevant experiments [43]–[46], reasonable stacking of the networks can effectively improve model's ability of classification and regression. Therefore, this paper construct a stacked BiLSTM network based on the LSTM cell to fully achieve semantic parsing of a single statement. We define the output of upper level BiLSTM network, y_t, as the input of next level BiLSTM network. The stacked BiLSTM network
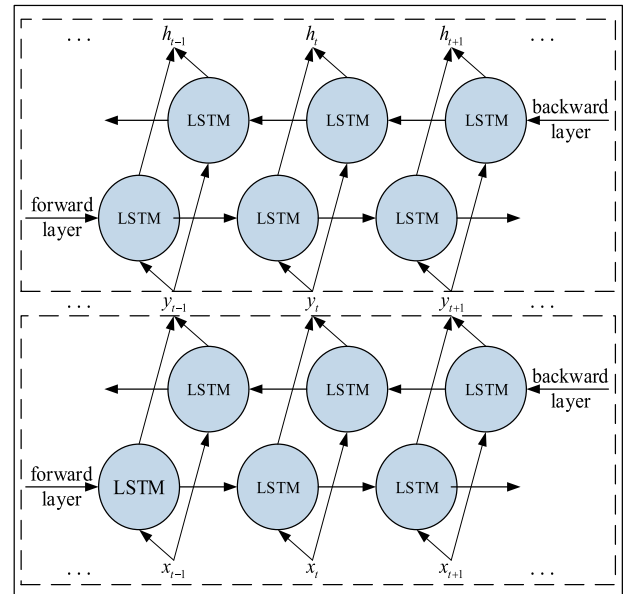

FIGURE 4. Stacked BiLSTM network structure.

structure is shown in Fig. 4. The status output of each word in the sentence is:

$$h_t = W_{h\overrightarrow{h}}\overrightarrow{h}_t + W_{h\overleftarrow{h}}\overleftarrow{h}_t + b_h \qquad (8)$$

We define p_t and p'_t to represent the t-th vectors of the question pair sequences C_Q and $C_{Q'}$, respectively. The question pair sequences after CNN network are sent to the stacked BiLSTM network separately, from which their state matrices

H_Q and $H_{Q'}$ can be obtained as follows:

$$h_0^q = 0, h_t^q = sBiLSTM(h_{t-1}^q, h_{t+1}^q, p_t) \tag{9}$$

$$h_0^{q'} = h_n^q, h_t^{q'} = sBiLSTM(h_{t-1}^{q'}, h_{t+1}^{q'}, p_t') \tag{10}$$

$$H_Q = \left[h_1^q, h_2^q, \ldots, h_n^q\right] \in R^{d*n} \tag{11}$$

$$H_{Q'} = \left[h_1^{q'}, h_2^{q'}, \ldots, h_m^{q'}\right] \in R^{d*m} \tag{12}$$

where d is the dimension of output state matrix of the hidden layer, $d \in R$.

B. SEMANTIC INTERACTION AND FEATURE REPRESENTATION USING COATTENTION AND ATTENTION MECHANISMS

In this paper, we design an incidence matrix in the coattention mechanism to capture the correlation and interaction between vectors, and use the softmax activation function to map the output of multiple neurons to the corresponding interval range. The state matrices H_Q and $H_{Q'}$ of question pair through the stacked BiLSTM network are taken as input, and the internal structure of the coattention mechanism is shown in Fig. 5.
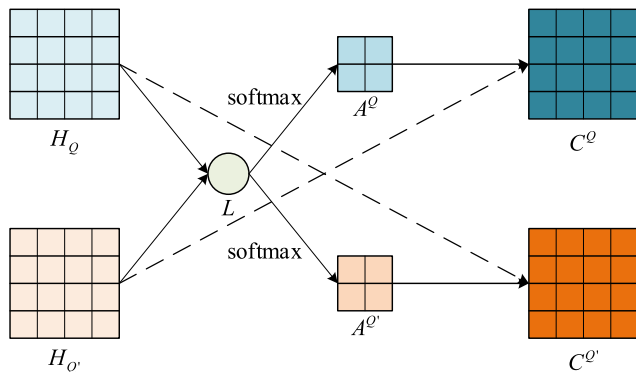

FIGURE 5. Internal structure of coattention mechanism.

We perform matrix multiplication on the state matrices H_Q and $H_{Q'}$ to calculate the correlation matrix L. Each item in incidence matrix is the correlation score between the words of the question pair sentences, that is, the interaction between the question pair statements is reflected as follows:

$$L = H_{Q'}^T H_Q \in R^{m*n} \tag{13}$$

The softmax function normalizes vector elements and performs well on the problems of multi-classification and probability distribution. With the softmax activation function, the attention weights of the hidden layer states of the question pair are:

$$A^Q = soft\max(L) \in R^{m*n}, \quad A^{Q'} = soft\max(L^T) \in R^{n*m} \tag{14}$$

After the interaction, the corresponding feature outputs of the question pair are:

$$C^Q = H_{Q'}A^Q \in R^{d*n} \tag{15}$$

$$C^{Q'} = H_Q A^{Q'} \in R^{d*m} \tag{16}$$

In the process of data calculation and transmission of the stacked BiLSTM network and the coattention mechanism, there may be some information mismatches in the question pair statements. To solve this problem, the attention mechanism layer is added following the coattention mechanism to connect and integrate the statement vector information of the previous few steps.

The input of the last module is the new eigenvector representations of the question pair statements C^Q and $C^{Q'}$, and C_t^Q is set as the t-th attention feature vector of the input question statement. We use max pooling to convert the input into a fixed-length vector O_q, which is the final feature vector output of the input question statement. When the attention mechanism is used to analytically represent the final vector of the existing question statement, the softmax weights of all the feature vectors ($C_1^{Q'}, C_2^{Q'}, \ldots, C_m^{Q'}$) in the existing question statement can be calculated and obtained according to O_q. After allocated weights, the final feature vector output $O_{q'}$ of the existing question statement can be obtained based on the attention mechanism model. Then, attention mechanism can integrate statement information, with which we can acquire the final outputs of the semantic parsing of the question pair as follows:

$$O_q = \max_{0 < t \leq n} C_t^Q \tag{17}$$

$$O_{q'} = \sum_{t=1}^m C_t^{Q'} S_{q'q}(t) \tag{18}$$

where $S_{q'q}$ is standardized by softmax function to represent the attention weight of the text vector of the existing question statement at time t, which is proportional to $C_t^{Q'}$. The higher the value of $S_{q'q}$, the higher the relevance between $C_t^{Q'}$ and the input question statement, that is, the existing question statement will have a greater effect on the eigenvector representation of the input question.

C. QUESTION PAIR MATCHING

After obtaining the final eigenvector representations of the question pair, the output is determined by measuring the similarity between the feature vectors, of which cosine similarity and Euclidean distance are commonly used. The measure standard of cosine similarity is the spatial angle between two vectors, while Euclidean distance is to calculate the absolute spatial distance of two vectors. The spatial diagram of cosine similarity and Euclidean distance is shown in Fig. 6.

Generally, it is hoped that the angle between the feature vectors of the question pair is small enough and the distance is the shortest so as to maximize the similarity capture and calculation of the question pair statement vectors. Therefore, this paper reconciles cosine similarity and Euclidean distance to generate a similarity calculation function that can consider the above both factors simultaneously. The calculation functions of the final vector similarity are as follows:
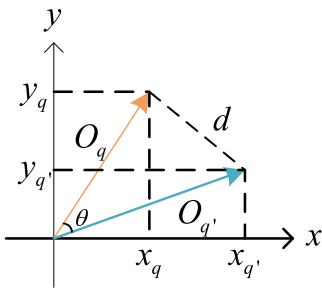
FIGURE 6. Schematic diagram of cosine similarity and Euclidean distance.

Cosine similarity:

$$Score_{\text{cosine}}(O_q, O_{q'}) = \frac{O_q \cdot O_{q'}}{|O_q| \, |O_{q'}|} \qquad (19)$$

Euclidean distance:

$$Score_{Euclidean}(O_q, O_{q'}) = \frac{1}{1 + \left\| O_q - O_{q'} \right\|_2} \qquad (20)$$

The cosine similarity is normalized to the range [0, 1]:

$$Score_{\text{cosine}}(O_q, O_{q'}) = 0.5 Score_{\text{cosine}}(O_q, O_{q'}) + 0.5 \qquad (21)$$

The final vector similarity matching function is as follows:

$$Score(O_q, O_{q'})$$
$$= \frac{2 \cdot Score_{\text{cosine}}(O_q, O_{q'}) \cdot Score_{Euclidean}(O_q, O_{q'})}{Score_{\text{cosine}}(O_q, O_{q'}) + Score_{Euclidean}(O_q, O_{q'})} \qquad (22)$$

where $\cdot$ represents the point multiplication operation, $|O_q|$ and $|O_{q'}|$ respectively represent the modulus length of the corresponding feature vectors, $\left\| O_q - O_{q'} \right\|_2$ is the Euclidean distance between two points. The values of the equations (19) and (20) are in the range of [0, 1].

IV. EXPERIMENTAL DATA AND SETTINGS
A. EXPERIMENTAL DATA
Throughout the current open-source dataset, many corpora are based on traditional question and answer matching for open domains, such as Baidu's WebQA, Stanford's SQuAD, CMU QA dataset, and SemEval 2015 cQA dataset [35]. Based on the experiments on the CCKS2018 dataset, this paper constructed a Chinese question answering dataset for the professional field, i.e., IPC-QA.

1) DATASET CCKS2018
The CCKS2018 dataset is derived from the third task of the WeBank Intelligent Customer Service Question Matching Contest. The core content of the contest is to understand the structure and semantics of two input question statements, and to determine whether they have the same inclination, intention or purpose by matching the correlation similarity of two sentences.

The data in the CCKS2018 task 3 corpus mainly comes from real texts in the financial field, and most of the question sentences are the customers' related business consultations.

TABLE 1. CCKS2018 dataset.

Set	Question pairs	Pos ratio	Average length
Training	100000	50%	≤ 20
Verification	10000	50%	≤ 20
Test	10000	50%	≤ 20
Total	120000	50%	≤ 20

The positive cases of the question pairs with the same intention in training set of the dataset are marked as 1, while the negative cases of the question pairs without the same intention are marked as 0. The original corpus provided by the official website contains 100000 marked question pairs as training set, 10000 unmarked question pairs as verification set, and 10000 unmarked question pairs as test set. The specific representation of the dataset is shown in Table 1.

2) INVERTED PENDULUM DATASET IPC-QA
a: DATA COLLECTION PROCESS AND STANDARDS
As inverted pendulum is a complex and unstable system, many typical questions of *Automatic Control Theory* such as nonlinearity, stabilization and tracking analysis can be simulated on the basic experimental platform. According to the basic knowledge involved in the inverted pendulum experiment, this paper constructed the IPC-QA dataset for restricted-domain teaching question answering by the online searching engines and the professional books.

The collection criteria are set as follows: a total of 60 questions closely related to the inverted pendulum experiment are constructed. Each question has five organized expressions and only one standard answer. The length of each question is set to no more than 35 words, and the length of each answer should not exceed 100 words. Among them, different expression forms corresponding to each question can be conveniently obtained by means of similar question recommendation methods in Baidu or Google search engines. In order to ensure the effectiveness and reliability, the answers corresponding to each question are basically obtained by consulting professional books.

According to the setting criteria, the corpus data is collected and the initial dataset has 60 questions related to the inverted pendulum experiment. Among them, five expressions of each question are selected without deviating from the core semantics but with some certain difference as much as possible. An example of the dataset is shown in Table 2.

The collected question pairs and answers are closely related to the inverted pendulum experiment. The initial dataset covers most of the relevant knowledge content involved in the experiment, which has better coverage and practical application value and is saved in an Excel document.

b: CONSTRUCTION OF EXPERIMENTAL DATASET OF INVERTED PENDULUM
In order to train and learn the model in this paper, the initial dataset needs to be sorted, divided and labeled. According to

TABLE 2. Example of the initial collected question answering dataset of inverted pendulum experiment.

Number	Contents	Question pair and answer statements
	Q_1 expression form 1	Basic units of the inverted pendulum
	Q_1 expression form 2	Which modules are there in the inverted pendulum?
	Q_1 expression form 3	What is the structure of the inverted pendulum?
Question Q_1	Q_1 expression form 4	Basic unit architecture of the inverted pendulum
	Q_1 expression form 5	Key compositions of the inverted pendulum
	A_1 answer	Inverted pendulum is a physical device, mainly including a sensor module and an arithmetic unit...
	Q_2 expression form 1	What are the ways to control the inverted pendulum?
	Q_2 expression form 2	What are the control strategies of the traditional control theory for the inverted pendulum system?
Question Q_2	Q_2 expression form 3	How can the inverted pendulum be controlled?
	Q_2 expression form 4	Common control methods for the inverted pendulum
	Q_2 expression form 5	How to choose the control methods of the inverted pendulum?
	A_2 answer	Variable structure control, predictive function control, iterative control, fuzzy adaptive control...
...	...	...
	Q_{60} expression form 1	Why does the inverted pendulum have the angle ring and the position ring?
	Q_{60} expression form 2	The role of the position ring in the inverted pendulum
	Q_{60} expression form 3	Why should the angle ring and position ring be used together?
Question Q_{60}	Q_{60} expression form 4	Why must the position ring be configured?
	Q_{60} expression form 5	The effect of the angle ring and the position ring on the inverted pendulum
	A_{60} answer	When only the angle ring, the inverted pendulum can make PWM to the pendulum deviation...

the question matching dataset in the financial field provided by CCKS WeBank 2018 question matching evaluation task, the initial dataset is structured as follows: For the five expressions of each question in Table 2, we divide each two into a group but not repeated, then a total of 600 positive cases of question pairs are obtained. The five expressions of each question are combined in turn with one of the expressions of the other questions without repetition, and a total of 8850 negative cases of question pairs can be obtained. According to the general rule of 8:2, the positive and negative cases of question pairs are divided into training verification set and test set. As a result, the training verification set includes 480 positive cases and 7080 negative cases of question pairs, and the test set has

TABLE 3. IPC-QA question answering dataset constructed.

Set	Question pairs	Pos pairs	Neg pairs	Average length
Training	6300	400	5900	≤ 35
Verification	1260	80	1180	≤ 35
Test	1890	120	1770	≤ 35
Total	9450	600	8850	≤ 35

120 positive cases and 1770 negative cases of question pairs. Moreover, the training verification set is further divided into training set and verification set in the proportion of 16.7%. Consequently, the training set has 400 positive cases and 5900 negative cases of question pairs, and the verification set has 80 positive cases and 1180 negative cases of question pairs. The IPC-QA dataset is shown in Table 3.

When labeling the question pairs in the training set, similar with the dataset provided by CCKS2018, the positive cases are marked as 1, and the negative cases are marked as 0.

B. EXPERIMENTAL SETTINGS AND PARAMETER OPTIMIZATION

To complete the training of the model, the specific experimental settings are as follows:

Some data of training sets in CCKS2018 and IPC-QA are sorted out to complete the training of the word2vec model. We set the dimension of the output word vector to 300 and the output length of each question statement to 35. The OOV (Out-Of-Vocabulary) zero-padding method is used to make up the insufficient length.

There are two different convolution windows with h of 2 and 3 respectively in the CNN network, which slide on the statement matrix by the stride of 1. If the number of convolution kernels for both windows is 100, 200 feature maps can be generated. Therefore, each question sentence can obtain corresponding output of a 200-dimensional matrix after max pooling, which is the feature vector input of the stacked BiLSTM network.

We set the initial learning rate of the model, the IPC-QA data batch size, the loss function's fixed margin M, and the regularization parameter λ to 0.001, 30, 0.2, and 1e-5, respectively. During the training process, a dropout layer with a dropout rate of 0.5 is added to the convolutional layer to avoid over-fitting of the model, a gradient clipping method with a clip gradient of 5 is used to avoid the model gradient explosion, and the hinge loss function is set as the objective function of the model training. In this section, Adam algorithm is used to optimize the model with the decay rate of 0.95 to complete the updating and optimization of network parameters. The specific calculation process is shown in Algorithm 1.

V. EXPERIMENTAL RESULTS AND ANALYSIS

According to the evaluation criteria commonly used in intelligent question answering, we use precision, accuracy, recall and F1 score to evaluate the experimental results on

---

Algorithm 1 Feature Extraction and Similarity Calculation of Question Pairs Using Deep Learning

---

Input: The question pair sequence $Q = (q_1, q_2, \ldots, q_n)$ and $Q' = (q'_1, q'_2, \ldots, q'_m)$ after Chinese word segmentation and word2vec vectorization

Output: The similarity degree of question pairs

Process:

1. Input the initial word vector $x_i \in \mathbb{R}^k$ of the statement in turn to perform the convolution operation $c_i = \sigma(W \cdot x_{i:i+h-1} + b)$, obtaining the one-dimensional feature map $c = [c_1, c_2, \ldots, c_{n-h+1}]$ of the statement and the statement vector output $C = \max[c]$ after pooling.

2. The distance dependence of the before and after words in the statement itself is captured to obtain the state matrix outputs H_Q and $H_{Q'}$ of the question pair.

3. The incidence matrix $L = H_{Q'}^T H_Q$ is used to extract the dependency relationship between the question pairs, combining the attention mechanism to obtain the final eigenvector representations O_q and $O_{q'}$ of the question pairs.

4. The harmonic function $Score(O_q, O_{q'})$ is used to calculate the matching degree of vectors, and Adam algorithm is used to optimize and update parameters, then evaluating the model's accuracy.

CCKS2018 and IPC-QA respectively, which are defined by (23)-(26).

$$Precision = \frac{N_{pc}}{N_{pt}} \quad (23)$$

where N_{pc} is the number of positive cases with correctly marked, N_{pt} is the total number of marked positive cases.

$$Accuracy = \frac{N_{qc}}{N_{qt}} \quad (24)$$

where N_{qc} is the number of correctly marked question pairs, N_{qt} is the total number of question pairs.

$$Recall = \frac{N_{pc}}{N_p} \quad (25)$$

where N_{pc} is the number of positive cases with correctly marked, N_p is the total number of positive cases.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (26)$$

A. EXPERIMENTAL RESULTS AND ANALYSIS ON CCKS2018 DATASET

When tested on CCKS2018 dataset, the initial feature representation process was completed using the jieba word segmentation tool and the word2vec model of 300-dimension output. Due to the large dataset, the batchsize was set to

50 and the iterations was 25. According to the evaluation criteria, the task on the dataset can be regarded as a binary classification task, whose label set is $\{1, 0\}$, where 1 indicates that two given question statements have the same or similar intent, and 0 indicates that two given question statements have no same or similar intent. Label 1 is given if the similarity between the two question statements exceeds 0.65, otherwise label 0 is given.

TABLE 4. Comparison of experimental results for each model on CCKS2018 dataset.[1]

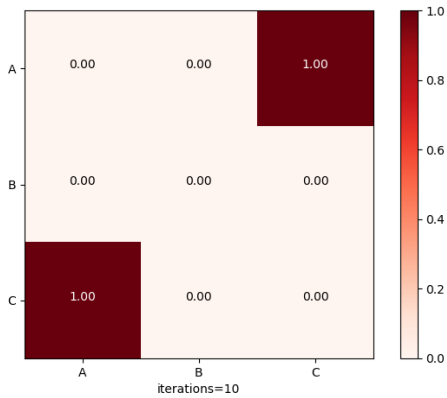| No.	Teams/Models	Precision	Accuracy	Recall	F1
1	Northeastern University NLP Laboratory [47]	**0.86485**	--	--	--
2	dw_xm	0.86396	--	--	--
3	Pingan Medical Text Understanding Department	0.86380	--	--	--
4	SBiLSTM-coA	0.85814	0.86279	0.86951	0.86379
5	CNN + SBiLSTM-coA	**0.86042**	0.86382	**0.87185**	0.86610

In Table 4, the sequence 1~3 is the top three teams in model precision in the competition. The model precision in this paper is 0.86042, which is lower precision of 0.443% than the best performance of the team and rank the eighth in the list of participating teams. However, analyzing lines 4 and 5, it can be found that the precision of the prediction results of question pair is higher of 0.228%. We can obviously see that adding the convolution layer can appropriately improve the analytical ability and prediction level of the model. At the same time, recall and F1 score in our model have achieved favorable label prediction results. The results indicate that the model has a high probability of correctly labeling the positive question pair cases and can well understand and extract related features of the similar statements to perform the representation calculations.

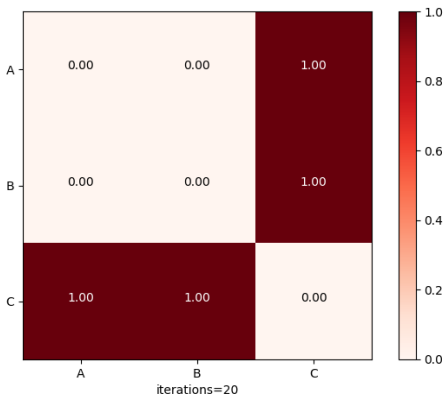## B. EXPERIMENTAL RESULTS AND ANALYSIS ON IPC-QA DATASET

Based on the research work of CCKS2018 dataset experiment, the model was verified and tested on IPC-QA in this paper. The preset parameters are shown in Section 4.2. Firstly, the iterations were set to 10, 20, and 30 respectively, and three question statements A, B and C were selected from the test set to test model's prediction ability with different iterations. A: How does LQR in the inverted pendulum work? B: What is the working state of the inverted pendulum? C: The application methods of LQR in the inverted pendulum. If a pair of positive cases or a pair of negative cases is predicted correctly, the label is 1 and vice versa, the same statement not predicted. The test results are shown in Fig. 7.

In order to further test the effects of the proposed methods, experiments are performed on the five models:
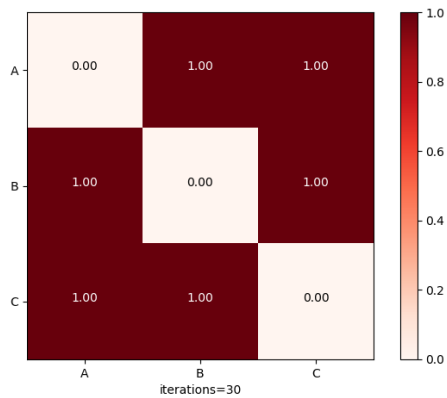
---

[1]https://www.biendata.com/competition/CCKS2018_3/leaderboard/

(a). The model's ability of label prediction with 10 iterations.

(b). The model's ability of label prediction with 20 iterations.

(c). The model's ability of label prediction with 30 iterations.

FIGURE 7. The model's ability of label prediction with different iterations.

BiLSTM, stacked BiLSTM (SBiLSTM), stacked BiLSTM + harmonic cosine similarity and Euclidean distance (SBiLSTM-CE), stacked BiLSTM + coattention mechanism + harmonic cosine similarity and Euclidean distance (SBiLSTM-coA), CNN + stacked BiLSTM + coattention mechanism + harmonic cosine similarity and Euclidean distance (CNN-SBiLSTM-coA). The results are shown in Table 5.

(1) In Fig. 7, it can be observed that when the iteration is 10, the model can correctly predict the positive cases A and C.

TABLE 5. Experimental results of the model on the IPC-QA dataset.

	Models	Precision	Accuracy	Recall	F1
1	BiLSTM	0.7487	0.7597	0.7609	0.7548
2	SBiLSTM	0.7613	0.7726	0.7785	0.7698
3	SBiLSTM-CE	0.7725	0.7882	0.7913	0.7818
4	SBiLSTM-coA	0.7960	0.8124	0.8183	0.8070
5	CNN-SBiLSTM-coA	**0.8031**	0.8196	**0.8214**	0.8121

However, at 20 iterations, the negative cases A and B still cannot be correctly predicted. The reason may be that the A and B sentences contain the same words, and the semantic features extracted by the model have not reached a certain depth. At 30 iterations, the model has been able to accurately classify the positive and negative cases composed of the three sentences.

(2) In Table 5, comparing lines 1 and 2, it can be found that the experimental results of the internally stacked BiLSTM network model are better than those of the single-layer BiLSTM. Generally, a properly stacked LSTM network helps the model to understand the relationship between the words in the mapping statement at a deep level.

(3) In Table 5, compared with the simple stacked BiLSTM, the precision and recall of the model with the coattention mechanism are increased by 2.35% and 2.7%, respectively. These data strongly proved that the coattention mechanism can effectively bridge the lexical gap between question pair sentences. A reasonable structure of the attention mechanism can help the NLP task to extract the key information of the sentence.

(4) In this paper, the optimal precision and recall are obtained through a model combining CNN, stacked BiLSTM, and the coattention mechanism, which are 0.8031 and 0.8214, respectively. This not only indicates the feasibility and accuracy of the model, but also shows that the harmonic cosine similarity and Euclidean distance can balance the angle and distance relationship between the vectors to achieve better matching of the sentence vectors.

(5) In Table 5, the precision of the model added with CNN network is 0.71% higher than the original model. At the same time, recall of the two models are 2.23% and 1.83% higher than precision respectively. The data indicates that the model has the certain ability of semantic analysis, and there are good experimental results in the understanding and matching of similar sentences. However, the final experimental precision is not good enough, of which one reason may be that the IPC-QA dataset corpus is not ample, resulting that the semantic understanding of the negative cases of the question pairs is not sufficiently differentiated by the model.

VI. CONCLUSION

In this paper, we focus on the difficulty of Chinese sentence semantic understanding and the lack of valid question answering datasets. Using convolutional neural network, long and short-term memory network, coattention mechanism and

attention mechanism to analyze the sentence information, we validated and analyzed the model on CCKS2018 dataset and the self-built Chinese question answering dataset for the inverted pendulum experiment. The experimental results verified the analytical expression ability of the model on Chinese corpus. The idea of question pair matching and the construction of inverted pendulum dataset have laid a solid foundation for further research on intelligent question answering. However, the model of this paper still has some defects in the expansion of professional field and dataset. In addition, although the model has achieved good results in experiments, it lacks practical applications in reality.

In the future, providing knowledge solutions based on professional restricted domains will become a general trend of future intelligent question answering. Deep learning can predict effective answers based on the summarized and reorganized knowledge databases in specific fields. However, the existing large-scale question answering corpora are oriented to the open domains. Researchers will focus on expanding datasets of restricted domains and designing subtler models corresponding to these datasets. With the development of data transmission speed, the application of restricted-domain intelligent question answering to the real scene environment such as intelligent teaching in virtual environment will also become the focus of future development.

REFERENCES

[1] A. Uyar and F. M. Aliyu, "Evaluating search features of Google knowledge graph and bing satori: entity types, list searches and query interfaces," *Online Inf. Rev.*, vol. 39, no. 2, pp. 197–213, Apr. 2015, doi: 10.1108/OIR-10-2014-0257.

[2] M.-D. Olvera-Lobo and J. Gutiérrez-Artacho, "Open-vs. restricted-domain QA systems in the biomedical field," *J. Inf. Sci.*, vol. 37, no. 2, pp. 152–162, Apr. 2011, doi: 10.1177/0165551511398575.

[3] M. Heilman and N. A. Smith, "Tree edit models for recognizing textual entailments, paraphrases, and answers to questions," in *Proc. NAACL HLT*, Los Angeles, CA, USA, Jun. 2010, pp. 1011–1019.

[4] C. Monz, "Machine learning for query formulation in question answering," *Natural Lang. Eng.*, vol. 17, no. 4, pp. 425–454, Oct. 2011, doi: 10.1017/S1351324910000276.

[5] G. Zhou, Y. Zhou, T. He, and W. Wu, "Learning semantic representation with neural networks for community question answering retrieval," *Knowl.-Based Syst.*, vol. 93, pp. 75–83, Feb. 2016, doi: 10.1016/j.knosys.2015.11.002.

[6] T. Mikolov, I. Sutskever, C. Kai, G. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Proc. NIPS*, Lake Tahoe, CA, USA, Dec. 2013, pp. 3111–3119.

[7] J. Song, Y. Guo, L. Gao, X. Li, A. Hanjalic, and H. T. Shen, "From deterministic to generative: Multimodal stochastic RNNs for video captioning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 10, pp. 3047–3058, Oct. 2019, doi: 10.1109/TNNLS.2018.2851077.

[8] R. Socher, C. C.-Y. Lin, A. Y. Ng, and C. D. Manning, "Parsing natural scenes and natural language with recursive neural networks," in *Proc. ICML*, Bellevue, WA, USA, Jun./Jul. 2011, pp. 129–136.

[9] N. Kalchbrenner, E. Grefensttete, and P. Blunsom, "A convolutional neural network for modelling sentences," in *Proc. ACL*, Baltimore, MD, USA, vol. 1, Jun. 2014, pp. 655–665.

[10] H. Y. Fang, F. Wu, Z. Zhao, X. Y. Duan, and Y. T. Zhuang, "Community-based question answering via heterogeneous social network learning," in *Proc. AAAI*, Phoenix, AZ, USA, Feb. 2016, pp. 122–128.

[11] Z. Zhao, "Community-based question answering via asymmetric multi-faceted ranking network learning," in *Proc. AAAI*, San Francisco, CA, USA, Feb. 2017, pp. 3532–3538.

[12] J. Song, H. Zhang, X. Li, L. Gao, M. Wang, and R. Hong, "Self-supervised video hashing with hierarchical binary auto-encoder," *IEEE Trans. Image Process.*, vol. 27, no. 7, pp. 3210–3221, Jul. 2018, doi: 10.1109/TIP.2018.2814344.

[13] Y.-P. Nie, Y. Han, J.-M. Huang, B. Jiao, and A.-P. Li, "Attention-based encoder-decoder model for answer selection in question answering," *Frontiers Inf. Technol. Electron. Eng.*, vol. 18, no. 4, pp. 535–544, Apr. 2017, doi: 10.1631/FITEE.1601232.

[14] S. D. Cheng and Y. Hu, "Limited domain knowledge question answering system based on BI-LSTM-CRF," *Comput. Mod.*, no. 7, pp. 53–57, 2018, doi: 10.3969/j.issn.1006-2475.2018.07.011.

[15] L. Zhang and L. Ma, "Coattention based BiLSTM for answer selection," in *Proc. IEEE Int. Conf. Inf. Autom. (ICIA)*, Macau, China, Jul. 2017, pp. 1005–1011, doi: 10.1109/ICInfA.2017.8079049.

[16] T. Chen, R. Xu, Y. He, and X. Wang, "Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN," *Expert Syst. Appl.*, vol. 72, pp. 221–230, Apr. 2017, doi: 10.1016/j.eswa.2016.10.065.

[17] X.-Y. Duan, S.-L. Tang, S.-Y. Zhang, Y. Zhang, Z. Zhao, J.-R. Xue, Y.-T. Zhuang, and F. Wu, "Temporality-enhanced knowledgememory network for factoid question answering," *Frontiers Inf. Technol. Electron. Eng.*, vol. 19, no. 1, pp. 104–115, Jan. 2018, doi: 10.1631/FITEE.1700788.

[18] W. Yin, H. Schütze, B. Xiang, and B. Zhou, "ABCNN: Attention-based convolutional neural network for modeling sentence pairs," *Trans. Assoc. Comput. Linguistics*, vol. 4, pp. 566–567, Dec. 2016, doi: 10.1162/tacl_a_00244.

[19] S. Pathak and N. Mishra, "Context aware restricted tourism domain question answering system," in in *Proc. NGCT*, Dehradun, India, Oct. 2016, pp. 534–539.

[20] C. Sun, Y. Guan, X. Wang, Q. Wang, and T. Liu, "InsunTourQA: A restricted-domain question answering system," *J. Comput. Inf. Syst.*, vol. 3, no. 4, pp. 1581–1590, Apr. 2007.

[21] P. Moreda, H. Llorens, E. Saquete, and M. Palomar, "Combining semantic information in question answering systems," *Inf. Process. Manage.*, vol. 47, no. 6, pp. 870–885, Nov. 2011, doi: 10.1016/j.ipm.2010.03.008.

[22] X. H. Zhu, Q. H. Cao, and F. F. Su, "A Chinese intelligent question answering system based on domain ontology and sentence templates," *Int. J. Digit. Content Technol. Appl.*, vol. 5, no. 11, pp. 158–165, Nov. 2011, doi: 10.4156/jdcta.vol5.issue11.20.

[23] M. Y. Kim, X. Ying, R. Goebel, and K. Satoh, "Answering yes/no questions in legal bar exams," in *Frontiers in Artificial Intelligence* (Lecture Notes in Computer Science), vol. 8417. 2014, pp. 199–213, doi: 10.1007/978-3-319-10061-6_14.

[24] T. Hao and E. Agichtein, "Finding similar questions in collaborative question answering archives: Toward bootstrapping-based equivalent pattern learning," *Inf. Retr.*, vol. 15, nos. 3–4, pp. 332–353, Jun. 2012, doi: 10.1007/s10791-012-9188-x.

[25] D. Palomera and A. Figueroa, "Leveraging linguistic traits and semi-supervised learning to single out informational content across how-to community question-answering archives," *Inf. Sci.*, vol. 381, pp. 20–32, Mar. 2017, doi: 10.1016/j.ins.2016.11.006.

[26] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. NIPS*, Lake Tahoe, CA, USA, Dec. 2012, pp. 1097–1105.

[27] F. Y. Zhou, L. P. Jin, and J. Dong, "Review of convolutional neural network," *Chin. J. Comput.*, vol. 40, no. 6, pp. 1229–1251, Jun. 2017, doi: 10.11897/SP.J.1016.2017.01229.

[28] L. J. Jin, Y. B. Fu, and Q. W. Dong, "The auto-question answering system based on convolution neural network," *J. East China Normal Univ. (Natural Sci.)*, no. 5, pp. 66–79, 2017, doi: 10.3969/j.issn.1000-5641.2017.05.007.

[29] X. Zhou, B. Hu, Q. Chen, and X. Wang, "Recurrent convolutional neural network for answer selection in community question answering," *Neurocomputing*, vol. 274, pp. 8–18, Jan. 2018, doi: 10.1016/j.neucom.2016.07.082.

[30] H. Palangi, L. Deng, Y. Shen, J. Gao, X. He, J. Chen, X. Song, and R. Ward, "Deep sentence embedding using long short-term memory networks: Analysis and application to information retrieval," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 24, no. 4, pp. 694–707, Apr. 2016, doi: 10.1109/TASLP.2016.2520371.

[31] L. Gao, X. Li, J. Song, and H. T. Shen, "Hierarchical LSTMs with adaptive attention for visual captioning," *IEEE Trans. Pattern Anal. Mach. Intell.*, to be published, doi: 10.1109/TPAMI.2019.2894139.

[32] D. Bahdanau, K. H. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," in *Proc. ICLR*, San Diego, CA, USA, May 2015, pp. 1–15.

[33] H. Fu, Z. Niu, C. Zhang, J. Ma, and J. Chen, "Visual cortex inspired CNN model for feature construction in text analysis," *Frontiers Comput. Neurosci.*, vol. 10, p. 64, Jul. 2016, doi: 10.3389/fncom.2016.00064.

[34] M. Tan, C. D. Santos, B. Xiang, and B. Zhou, "Improved representation learning for question answer matching," in *Proc. ACL*, Berlin, Germany, vol. 1, Aug. 2016, pp. 464–473.

[35] Y. Xiang, Q. Chen, X. Wang, and Y. Qin, "Answer selection in community question answering via attentive neural networks," *IEEE Signal Process. Lett.*, vol. 24, no. 4, pp. 505–509, Apr. 2017, doi: 10.1109/LSP.2017.2673123.

[36] B. Zhang, D. Xiong, J. Su, and M. Zhang, "Learning better discourse representation for implicit discourse relation recognition via attention networks," *Neurocomputing*, vol. 275, pp. 1241–1249, Jan. 2018, doi: 10.1016/j.neucom.2017.09.074.

[37] X. P. Li, "Beyond RNNs: Positional self-attention with co-attention for video question answering," in *Proc. AAAI*, Honolulu, HI, USA, Jan./Feb. 2019, pp. 8658–8665.

[38] C. Yang, M. Jiang, B. Jiang, W. Zhou, and K. Li, "Co-attention network with question type for visual question answering," *IEEE Access*, vol. 7, pp. 40771–40781, 2019, doi: 10.1109/ACCESS.2019.2908035.

[39] M. J. Er, Y. Zhang, N. Wang, and M. Pratama, "Attention pooling-based convolutional neural network for sentence modelling," *Inf. Sci.*, vol. 373, pp. 388–403, Dec. 2016, doi: 10.1016/j.ins.2016.08.084.

[40] Y. Kim, "Convolutional neural networks for sentence classification," in *Proc. EMNLP*, Doha, Qatar, Oct. 2014, pp. 1746–1751.

[41] A. Hassan and A. Mahmood, "Convolutional recurrent deep learning model for sentence classification," *IEEE Access*, vol. 6, pp. 13949–13957, 2018, doi: 10.1109/ACCESS.2018.2814818.

[42] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2673–2681, Nov. 1997, doi: 10.1109/78.650093.

[43] D. Wang and E. Nyberg, "A long short-term memory model for answer sentence selection in question answering," in *Proc. ACL-IJCNLP*, Beijing, China, Jul. 2015, pp. 707–712.

[44] Z. J. Liu, "Entity recognition from clinical texts via recurrent neural network," *BMC Med. Informat. Decis. Making*, vol. 17, no. 2, p. 67, Jul. 2017, doi: 10.1186/s12911-017-0468-7.

[45] T. Liu, S. Yu, B. Xu, and H. Yin, "Recurrent networks with attention and convolutional networks for sentence representation and classification," *Appl. Intell.*, vol. 48, no. 10, pp. 3797–3806, Oct. 2018, doi: 10.1007/s10489-018-1176-4.

[46] L. Cai, S. Zhou, X. Yan, and R. Yuan, "A stacked BiLSTM neural network based on coattention mechanism for question answering," *Comput. Intell. Neurosci.*, vol. 2019, pp. 1–12, Aug. 2019, doi: 10.1155/2019/9543490.

[47] Y. Liu, "An enhanced ESIM model for sentence pair matching with self-attention," in *Proc. CCKS-Tasks*, Tianjin, China, vol. 2242, Aug. 2018, pp. 52–62.

MIN WEI was born in Sichuan, China, in 1995. She received the B.E. degree in automation from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2018. She is currently pursuing the M.E. degree in control engineering with the Chongqing University of Posts and Telecommunications, Chongqing, China. Her research interests include natural language processing, intelligent question answering, and text sentiment analysis. She was a recipient of the Second Prize in the sixteenth China post-graduate mathematical contest in modeling.

SI-TONG ZHOU was born in Yunnan, China, in 1994. She received the B.E. degree in electrical engineering and automation from the Chongqing University of Posts and Telecommunications, Chongqing, China, in 2017, where she is currently pursuing the M.S. degree in control science and engineering. Her main research interests include natural language processing and image recognition. During the period, she has published one article of the SCI journal and two patents. Besides, she presided over a project of research and innovation in Chongqing. She was a recipient of the First Prize in the fifteenth China post-graduate mathematical contest in modeling.

LIN-QIN CAI (Member, IEEE) was born in Sichuan, China, in 1973. He received the Ph.D. degree in pattern recognition and intelligent system from the University of Science and Technology of China, Hefei, China, in 2007. He is currently a Professor with the School of Automation, Chongqing University of Posts and Telecommunications, Chongqing, China. His current research interests include artificial intelligence, human–computer interaction, and pattern recognition. He has published more than 40 articles in related international journals and conferences. He has served as a reviewer for several international journals.

XUN YAN was born in Hunan, China, in 1994. He received the B.E. degree in automation from the School of Information Engineering, Henan University of Science and Technology, Luoyang, China, in 2016, and the M.S. degree in control theory and control engineering from the School of Automation, Chongqing University of Posts and Telecommunications, Chongqing, China, in 2019. His research interests include artificial intelligence, natural language processing, and virtual learning environment.

• • •