IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 6 June 2022, accepted 11 July 2022, date of publication 18 July 2022, date of current version 22 July 2022.
Digital Object Identifier 10.1109/ACCESS.2022.3191678

RESEARCH ARTICLE

Building a Question Answering System for the Manufacturing Domain

LIU XINGGUANG[1,2], CHENG ZHENBO[1,2], SHEN ZHENGYUAN[1,2], ZHANG HAOXIN[3], MENG HANGCHENG[3], XU XUESONG[1,2], AND XIAO GANG[1,2]
[1]College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou, Zhejiang 310032, China
[2]College of Software Engineer, Zhejiang University of Technology, Hangzhou, Zhejiang 310032, China
[3]College of Mechanical Engineering, Zhejiang University of Technology, Hangzhou, Zhejiang 310032, China

Corresponding author: Cheng Zhenbo (czb@zjut.edu.cn)

This work was supported in part by the National Science Foundation of China under Grant 61976193, in part by the Zhejiang Natural Science Foundation of China under Grant LY19F020034, and in part by the Science and Technology Key Research Planning Project of Zhejiang Province of China under Grant 2021C03136.

ABSTRACT The design or simulation analysis of special equipment products must follow the national standards, and hence it may be necessary to repeatedly consult the contents of the standards in the design process. However, it is difficult for the traditional question answering system based on keyword retrieval to give accurate answers to technical questions. Therefore, we use natural language processing techniques to design a question answering system for the decision-making process in pressure vessel design. To solve the problem of insufficient training data for the technology question answering system, we propose a method to generate questions according to a declarative sentence from several different dimensions so that multiple question-answer pairs can be obtained from a declarative sentence. In addition, we designed an interactive attention model based on a bidirectional long short-term memory (BiLSTM) network to improve the performance of the similarity comparison of two question sentences. Finally, the performance of the question answering system was tested on public and technical domain datasets.

INDEX TERMS Question answering system, BiLSTM, interactive attention, similarity comparison, design standard.

I. INTRODUCTION

The term special equipment refers to boilers, pressure vessels (including gas cylinders), pressure pipelines, elevators, hoisting machinery, passenger ropeways, large amusement facilities, and special motor vehicles on sites that involve threats to human safety and high risk. The design process of special equipment products often needs to follow various national standards [1]. Taking China's elevator products as an example, the whole life cycle of an elevator product must comply with the related equipment safety regulations. In addition, elevator products must follow 12 safety technical specifications and 35 national and industrial standards. These regulations, technical specifications, and standards ensure the safety of products with respect to the design, manufacture, installation, and maintenance of elevators. Therefore, many decisions in the design and manufacturing process of these products need to be made by consulting various standards and technical specifications frequently.

The design and manufacture of mechanical products involve many types of decision-making problems, and artificial intelligence technology is often used to assist these decision-making processes to improve the efficiency of decision-making and the accuracy of the decision-making results [2]–[6]. Early decision-making systems mainly consisted of parametric systems and expert systems. A parametric system is primarily used to determine the selection of parameters in the design process [7]. The parameters can be automatically obtained under certain constraints through an intelligent algorithm. Systems that use intelligent algorithms to assist parameter selection have been successfully applied in aircraft design, automobile design, and elevator design [2]. In an expert system, a process of formalizing human expert knowledge is used to form rules, and then the

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

rules are used to assist decision-making. Expert systems have been successfully applied in the design and manufacture of mechanical products [8]. For example, Bojan *et al.* proposed an expert system for finite element mesh generation [9]. The system collected 2000 mesh generation rules to help users select appropriate mesh types and mesh setting parameters. El-Ghany *et al.* established an expert system for finite element analysis in mechanical manufacturing [10]. The system can answer users' questions about geometry, element selection, and boundary condition settings in mechanical manufacturing. However, both parametric and expert systems, which are difficult to use for decision-making in certain domains, can only assist decision-making for specific products.

One of the most promising ways to make decisions within the technical domain is through the use of a question answering (QA) system [11]. A QA system aims to provide accurate answers to users' questions in natural language. The initial idea of the QA system began with the Turing test in 1950, and structured QA systems were first proposed in the 1960s [12], [13]. The main process of a structured QA system is to analyze the questions, turn them into database queries, and finally search for the answers in a database. The most well-known system from this period is the Baseball system, which can answer questions related to the American Basketball League in a certain season [14]. In the 1990s, due to the rapid development of the Internet, users produced a large amount of text data. The knowledge of QA systems comes from various fields, and the QA system in this period gradually entered the open domain [12], [15], [16]. The goal of QA systems is to give the final answer to a question directly instead of returning a list of relevant fragments or hyperlinks, thus providing better user-friendliness and efficiency than search engines.

With the development of deep learning and large-scale pretrained language models, domain-oriented QA systems have made great progress [14]–[19]. In terms of question analysis, many studies have developed a classifier based on neural networks to classify questions. For example, the classification of a given question can be realized by a convolutional neural network [18] or LSTM [19]. For document retrieval in QA systems [20], [21], to reduce the difficulty of accurately matching terms, problems or documents can be encoded into a vector space to avoid the decrease in retrieval performance caused by mismatching terms. For example, [22], [23] proposed measuring the similarity of documents or questions by training the encoder to encode documents or questions into vectors and then directly calculating the inner product of the vectors.

The pretrained model of natural language processing has led to remarkable breakthroughs in various natural language tasks [24]. Pretrained models such as bidirectional encoder representations from transformers (BERT) have also been applied in QA systems [25]. The question-answer pairs can be constructed in advance, and then the answer to the new question can be obtained by comparing the similarity between the new question and existing questions. This method obtains the answers to questions through an end-to-end training process. Their performance mainly depends on the quality of the question-answer pairs and the precise similarity comparison between questions. BERT can be used to obtain a vector representation of the questions using masked language modeling in the pretraining stage, and then the similarity between the questions can be obtained by fine-tuning the model parameters in the fine-tuning stage [26]–[28].

Recently, given the many successful applications of QA system in the open domain, researchers in the manufacturing field have also begun to pay attention to the use of QA systems [29]. For example, Li and Ding integrated ontology and a mechanical product design scenario into the semantic similarity calculation process, and their QA system can answer users' questions about mechanical product function, structure, principle, index parameters, calculation formula, and engineering material data [30]. Li *et al.* built a knowledge-atlas QA system. The QA system can answer the task-based requirements of the machinery industry, assist in analysis, and support decision-making [31]. Wang *et al.* discuss the evolvement of deep learning technology and their advantages over traditional machine learning, and propose a deep learning-based computational approach aimed at improving system performance in manufacturing [32]. Shi constructed a QA system for mechanical intelligent manufacturing based on a knowledge graph, and used natural language processing technologies such as entity recognition, entity links, and relationship extraction to improve the accuracy of the QA system [33].

Generally speaking, these QA systems need to construct a question-answer pair dataset in advance. When a new question is given, the answer to the question that is most similar to the new question can be taken as the answer to the new question by comparing the similarity between the new question and each question in the dataset. Therefore, the main factors affecting the performance of QA systems are the size of question-answer pair dataset and the question similarity comparison algorithm. However, question-answer pair datasets in the mechanical manufacturing field tend to be small in scale, and it is difficult for existing sentence similarity comparison algorithms to capture the semantics of sentences associated with the application context. As a result, it is difficult to achieve the expected results using current QA systems in the mechanical manufacturing field.

To address the aforementioned challenges, we propose a method to generate question sentences according to a declarative sentence from several semantic dimensions so that multiple question-answer pairs can be accumulated from one declarative sentence. In addition, we propose an interactive attention model based on a BiLSTM network to improve the performance of the similarity comparison of two question sentences. The interactive attention model can capture the semantic associations of words and parameters in sentences. Experiments show our proposed model can provide superior performance than other state-of-the-art methods. Finally, a QA system based on the proposed methods was developed. The system can enable pressure vessel designers to quickly obtain the answers to common problems in the design process.

II. METHODS

This section has three parts. In the first part the framework of technical QA system was proposed. In the second part, we put forward a method to generate question sentences according to a declarative sentence from several semantic dimensions so that multiple question-answer pairs can be accumulated from one declarative sentence. In the third part, an interactive attention model based on a BiLSTM network was proposed to improve the performance of the similarity comparison of two question sentences.

A. FRAMEWORK OF TECHNICAL QA SYSTEM

We present our proposed framework for the technical QA system. As shown in Fig.1, the system consists of two parts: the front end and back end. The front end handles the input of user questions and the output of answers. The back end includes three modules: the question-answer pairs generating module, the knowledge library module and similarity comparison module. The knowledge library module is composed of question-answer pairs. The question-answer pairs generating module manually and automatically constructs question-answer pairs. The similarity comparison module finds a set of questions similar to the user's question in the knowledge library. The similarities between user's question and the questions in knowledge library are sorted. The answer i_1 of the question i_1 with the highest similarity to the user input question will be used as the candidate answer of the user input question.

B. TEMPLATE-BASED QUESTION GENERATION ALGORITHM

The knowledge library module stores question-answer pairs and provides basic data services for the technical QA system. Question-answer pairs are constructed using two methods. One method is to have them generated manually by technical experts according to technical documents.

In this approach, technical experts manually design the basic question-answer pairs $P = \{(q_i, a_i)|i = 1, 2, \cdots, m$ according to the standard manual, and then the annotator expands the questions accordingly to obtain $P = \{(q_i^j, a_i)|i = 1, 2, \cdots, m; j = 1, 2, \cdots, k$. For example, if the question given by technical experts is "标准适用核能装置的容器吗? (Is the standard applicable to containers for nuclear power plants?)" and the answer is "不能。(No.)," the additional questions created by the annotator include "核能装置的容器适用于本标准吗? (Is the container of nuclear power plant applicable to this standard?)" and "核能装置的容器是否适用于本标准? (Is this standard applicable to the container of a nuclear power plant?)," and the answer to these two extended questions is still "不能。(No.)."

The other approach is to automatically generate question-answer pairs using a question word replacement algorithm. We designed a template-based question generation algorithm to automatically generate multiple question-answer pairs from a statement sentence. As shown in Fig. 2, the algorithm includes the following steps:

- Carry out named entity recognition (NER) on the standard document, and select the sentences containing entities as candidate statement sentences for generating question-answer pairs.
- Select an appropriate interrogative word according to the entity type, and then replace the entity in the candidate sentence with the interrogative word.
- Adjust the order of words in the sentence containing the interrogative word to form a question. The entity replaced in the previous step will be the answer to the question.

We use the Bert-BiLSTM-CRF model to realize NER. First, the labeled corpus is transformed into a word vector through BERT pretrained language model. Then input the word vector into the BiLSTM module for further processing. The conditional random field (CRF) module is used to decode the output result of the BiLSTM module to obtain a predictive annotation sequence. Finally, each entity in the sequence is extracted and classified to complete the whole process of NER.

We use the BIO mode to label entities, where B (Begin) represents the starting position of an entity, I (Inside) indicates that the word is inside the entity, and O (Outside) indicates that the word does not belong to any entity. For each entity, we also designed types to describe it. All entity types are listed in Table 1. The "Property" indicates the dimension description method of a mechanical product, such as "Geometric dimension." The "Condition" represents the working state of a product in what environment or under what conditions, such as "Corrosive environment." The "Category" refers to the category of the product itself, such as "Evaporator" or "Reactor." The "Material" is the name of the production material of the product. The "Stage" is the stage of the product in the production process, such as "Failure analysis" or "Detection stage." The "Parameter" refers to a dimension of the product that can be calculated, such as "Length," "Density," or "Roundness."

B-PRO indicates that the entity is the starting word of the Property entity. For instance, for the sentence "本规定不适用于对焊法兰的颈部过渡段 (This provision is not applicable to the neck transition section of the butt welding flange.)," the entity labels are shown in Table 2.

After determining the entities in the sentence, each entity can be replaced with interrogative words to form a question sentence. Because six types of named entities are used, six interrogative words are used. The six interrogative words correspond to the English terms "what property," "what working condition," "what product category," "what material," "what stage," and "what parameter." The relationship

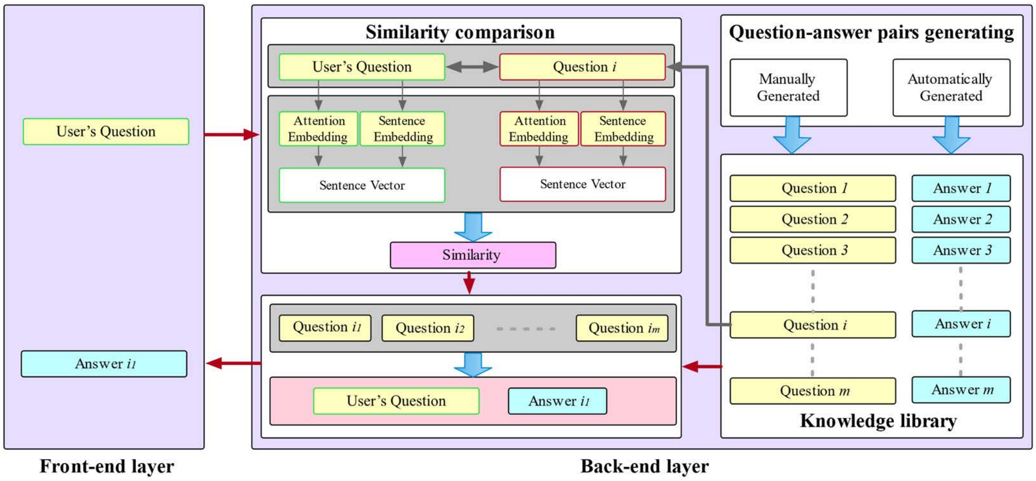

FIGURE 1. Architecture of the QA system.

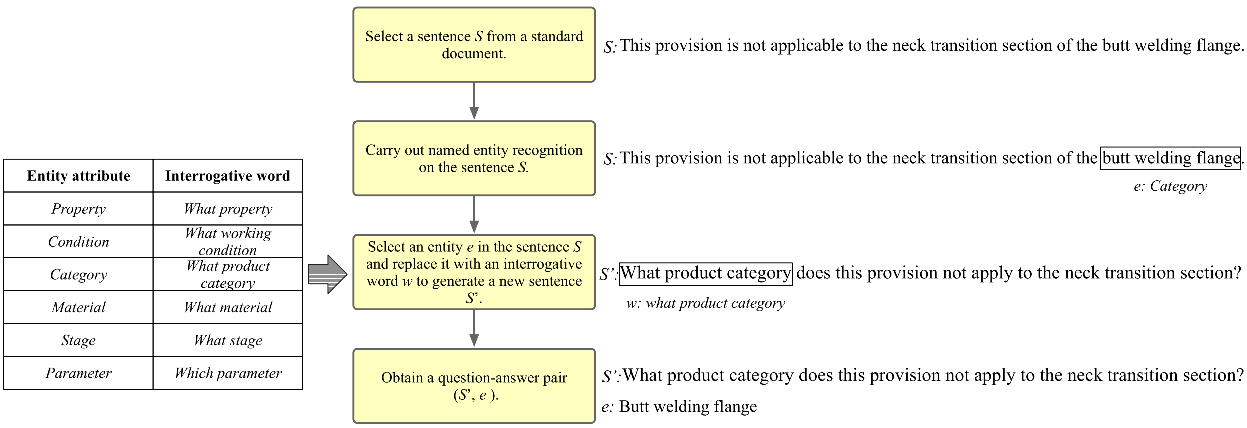

FIGURE 2. Flow chart of the template-based question generation algorithm.

between the named entities and corresponding interrogative words is shown in TABLE 1.

To place the interrogative words in the appropriate position in the question sentence, each word of the question sentence is in a position consistent with Chinese grammar. For the interrogative words "what working condition" and "what parameter" in TABLE 1, it is necessary to put the interrogative words in the first position of the question sentence. In other cases, according to Chinese grammar, it is possible to directly replace the entity with the interrogative word. For example, for the sentence "本规定不适用于对焊法兰的颈部过渡段 (This provision is not applicable to the neck transition section of the butt-welding flange.)." The entity of this sentence is "butt welding flange." After replacing the entity with the interrogative word according to Table 1, the question becomes "本规定不适用于什么产品类别的颈部过渡段? (What product category does this provision not apply to the neck transition section?)." After replacement, the formed question-answer pair is ("本规定不适用于什么产品类别的颈部过渡段" ~ "对焊法兰". "What product category does this provision not apply to the neck transition section?" ~ "Butt welding flange.").

We named the dataset of question-answer pairs generated using the methods in this section as "Pressure Vessel Manufacturing dataset (PVM)." We use PostgreSQL to store question-answer pairs in the knowledge library module. A PostgreSQL database supports JSON format data import. Hence, we can directly import question-answer pairs into PostgreSQL database to facilitate subsequent operations.

C. SEMANTIC SIMILARITY BASED ON INTERACTIVE ATTENTIONS

For the QA system, when a question q is given, it is necessary to compare q with each question in the question-answer pairs in turn and return the question s_i with the most similar semantics to q. The answer a_i of question s_i is used as an alternative answer to the question q. We proposed an interactive attention BiLSTM (IA-BiLSTM) model, as shown in Fig.3 to compare the similarity between two questions.

TABLE 1. Types of entities.

Attribute	Abbreviation	Interrogative words
Property	PRO	什么属性("what property")
Condition	CON	什么工况("what working condition")
Category	CAT	什么类别("what product category")
Material	MAT	什么材料("what material")
Stage	STA	什么阶段("what stage")
Parameter	PAR	什么参数("which parameter")

The input of this model is two questions (q_a, q_b), and the output is the similarity between them. The similarity value is between 0 and 1, where 0 means that the two sentences are completely different, and 1 means that the two sentences are exactly the same.

The question is segmented using the Jieba word segmentation tool. For example, if the question is "标准适用最大多少压力? (For how much pressure is the standard applicable?)" then the result after word segmentation is "标准/适用/最大/多少/压力." The word vector representation of each word in the sentence can be obtained according to

$$\begin{aligned} a_i &= embedding\left(q_a^i\right), \quad \forall i \in [1, \cdots, l_a], \\ b_j &= embedding\left(q_b^i\right), \quad \forall i \in [1, \cdots, l_b], \end{aligned} \tag{1}$$

where *embedding* represents the embedding function, such as word2vec, and l_a (l_b) is the length of words in sentence q_a (q_b). We then use BiLSTM to encode the word context in the sentence. A BiLSTM runs a forward and backward LSTM on a sequence starting from the left and right ends, respectively. The hidden states generated by these two LSTMs at each word are concatenated to represent a word and its context. We write $\bar{a}_i$ $(\bar{b}_j)$ to denote the output state generated by the BiLSTM at ith (jth) word over the input sequence $q_a(q_b)$.

$$\begin{aligned} \bar{a}_i &= BiLSTM(a, i), \\ \bar{b}_j &= BiLSTM\,(b, i)\,. \end{aligned} \tag{2}$$

After the word vector containing the context is obtained, the cross attention can be calculated according to the word vector. The semantic association between words and parameters in sentences can be captured using the interactive attentions. We calculate the dot product between two-word vectors to represent the attention weight between them, that is,

$$e_{ij} = \bar{a}_i^T \times \bar{b}_j. \tag{3}$$

For the ith word in sentence q_a, we accumulate each word vector in sentence q_b to form a new representation of word i, which is represented as $\tilde{a}_i$ with Eq.4. The word vector in sentence q_b is multiplied by the attention weight in Eq.3 so that the word in sentence q_b is most related to the ith word in sentence q_a. The same is applied to $\tilde{b}_j$ with Eq.4.

$$\begin{aligned} \tilde{a}_i &= \sum_{j=1}^{l_b} \frac{exp e_{ij}}{\sum_{k=1}^{l_b} \exp e_{ik}} \bar{b}_j, \\ \tilde{b}_j &= \sum_{i=1}^{l_b} \frac{exp e_{ij}}{\sum_{k=1}^{l_a} \exp e_{jk}} \bar{a}_i. \end{aligned} \tag{4}$$

The sentence vector generated by BiLSTM and the sentence vector obtained by interactive attention are combined to form the final vector representation of the sentence, that is,

$$\begin{aligned} v_a &= [\bar{a}, \tilde{a}], \\ v_b &= [\bar{b}, \tilde{b}]. \end{aligned} \tag{5}$$

The similarity of two sentences is calculated by cosine similarity. The cosine similarity formula of vector $\mathrm{v_a}$ and vector $\mathrm{v_b}$ is as follows:

$$d_{ab} = \frac{\sum_{i=1}^{n} v_a^i v_b^i}{\sqrt{\sum_{i=1}^{n} (v_a^i)^2}\sqrt{\sum_{i=1}^{n} (v_b^i)^2}}. \tag{6}$$

III. EXPERIMENTAL RESULTS AND DISCUSSION

We first evaluate the performance of the automatically generated question-answer pairs algorithm, and then compare the results of IA-BiLSTM with other common sentence similarity comparison methods in various benchmark datasets. We also develop a technical QA system for the pressure vessel product design standard to verify the effectiveness of the proposed technical QA framework with front-end and back-end layers.

A. PERFORMANCE OF QUESTION-ANSWER PAIR GENERATION

We first evaluate the performance of the question-answer pair generating algorithm. The proposed system has 3,649 question-answer pairs that were automatically generated according to our method and 7,601 manually generated question-answer pairs (see Section III A.). Technical experts and annotators were recruited to manually generate question-answer pairs. In this study, the technical experts were master students majoring in mechanical engineering, and the annotators were undergraduates majoring in mechanical engineering. All students were recruited from the Zhejiang University of Technology.

We manually evaluated these question-answer pairs using the four indicators of relevance, fluency, ambiguity, and instruction. The scoring rules for these four indicators are listed in Table 3. We recruited a total of 30 volunteers majoring in mechanical engineering to score the question-answer pairs, and the average scores are shown in Table 4. Compared with the manually generated question-answer pairs, automatically generated question-answer pairs have better relevance and instruction, but score less well with respect to fluency and ambiguity. In addition, the method of automatically generating question-answer pairs does not consider the relevance of context when replacing question words, resulting

TABLE 2. Example of label entities.

sentence	本	规定	不	适用	于	对焊	法兰	的	颈部	过渡段
label	O	O	O	O	O	B-CAT	I-CAT	O	O	O

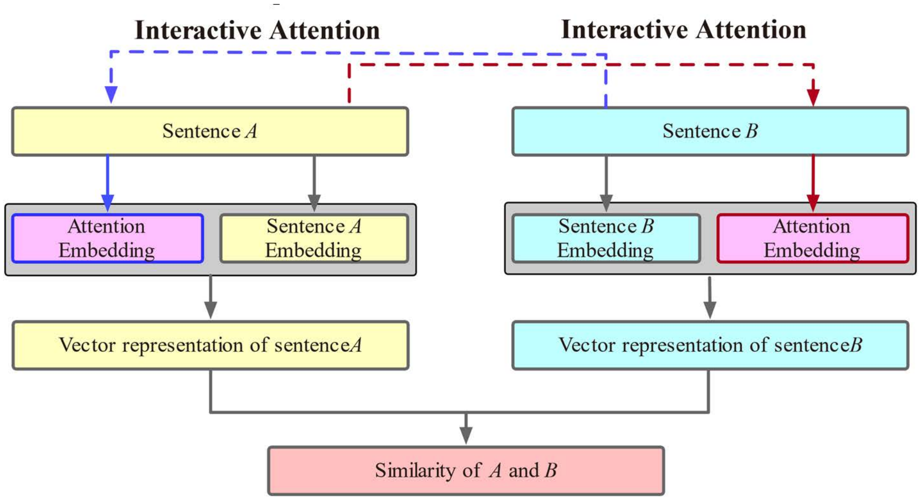

FIGURE 3. IA-BiLSTM model for computing semantic similarity.

TABLE 3. Scoring rules for the four indicators.

Score	Relevance	Fluency	Ambiguity	Instruction
1	Completely irrelevant.	Total Inconsistent.	Contradictions.	No help.
2	Partial correlation.	Partial Inconsistent.	Partial Contradictions.	Partial help.
3	Most related.	Most Consistent.	One ambiguous word.	Helps.
4	Completely related.	Fluency.	No ambiguous word.	Helps a lot.

TABLE 4. Average score of generated questions.

	Relevance	Fluency	Ambiguity	Instruction
Manually generated	3	3.3	3.4	2.8
Automatically generated	3.5	2.9	2.3	3.4

in a low score for the ambiguity indicator. Although the indicator results of automatically generated question-answer pairs are generally lower than those of manually generated question-answer pairs, considering the high cost of manually generated question-answer pairs, automatically generated question-answer pairs can balance the trade-off between the cost and size of a training dataset.

B. PERFORMANCE OF SEMANTIC SIMILARITY BETWEEN QUESTIONS

We compare the performance of the IA-BiLSTM model with existing models such as CNN, LSTM, and BiLSTM on the task of question semantic similarity matching using the Quora Question Pair (QQP) dataset, the LCQMC dataset, and the CCKS dataset. Among them, the QQP dataset is a collection of similar question pairs found in Quora, the question pairs in the training set are marked as similar or dissimilar, there are 363,870 pairs in the training set, 390,965 pairs in the test set, and 40,431 pairs in the validation set. The LCQMC dataset is provided by Harbin Institute of Technology, which cover many fields such as daily life, education, entertainment, computer games, social interaction, natural science and sports, there are 238,766 pairs in the training set, 8,802 pairs in the test set. CCKS is a dataset constructed in the real scene sentence intent matching task organized and implemented by the Intelligent Computing Research Center of Harbin Institute of Technology (Shenzhen), which contains common questions asked by users to customer service, filtered and matched with human annotation intent, there are 60,000 pairs in the training set, 12,000 pairs in the test set. We used the CNN, LSTM, and BiLSTM models for comparison with the proposed IA-BiLSTM model. Each model was trained for 20 iterations. After each training process, the test dataset was used to test the accuracy of the model.

Alibaba cloud's Platform of Artificial Intelligence (PAI) was used to run these experiments, and the neural network model training code was written in the Notebook Interactive Data Science Workshop. The hardware environment was the ecs.gn5-c4g1.xlarge instance on the PAI platform, which contains four virtual CPUs, 30G of memory, a NVIDIA P100 (graphics card), and 3 Gbps bandwidth. The experimental data were divided into training and test sets according to the ratio of 9:1. Nesterov adaptive motion estimation was used as the optimization function to adjust the weights of the neural networks, and cross entropy was used as the loss function. The TensorFlow library was used to build the neural network models, scikit-learn was used for the dataset segmentation, and the Jieba word splitter was used for sentence segmentation.

As shown in Table 5, the accuracy rates of CNN, LSTM and BiLSTM are all about 72% in the QQP dataset. Although the accuracy of the method IA-BiLSTM in this paper is slightly lower than that of the CNN model, but the IA-BiLSTM shows a higher accuracy rate (74.42%) and recall rate(70.81%). It shows that IA-BiLSTM has better performance in predicting the computation of truly similar problems. Because the content of the QQP dataset is full English short text data, and CNN is better at processing short text than LSTM (or BiLSTM network). But in contrast, the accuracy, precision and recall rate of IA-BiLSTM are higher than those using traditional attention mechanism LSTM or BiLSTM.

On the LCQMC dataset, IA-BiLSTM achieves an accuracy of 78.09%, a precision of 76.48, and a recall of 81.14%. On the CCKS dataset, IA-BiLSTM achieves 88.59% accuracy, 88.08% precision, and 89.25% recall. As shown in Table 5, the test performances of IA-BiLSTM are all better than CNN, LSTM or BiLSTM models. Compared with other models, IA-BiLSTM uses the interactive attention between sentences, which makes it easier to capture the semantic associations of words and parameters in sentences, so as to improve the performance of sentence similarity comparison.

C. QA SYSTEM FOR ANALYSIS AND DESIGN STANDARD

According to the methods proposed above, we developed a QA system based on the "JB4732 steel pressure vessels - analysis and design standard" document. JB4732 is a professional mandatory standard for pressure vessels, which is reviewed and approved by the National Technical Committee of China for pressure vessel standardization. The applicable scope of the standard includes vessels with a design pressure greater than or equal to 0.01 MPa and less than 100 MPa and vessels with a vacuum degree greater than or equal to 0.02 MPa. The standard considers the design, material selection, manufacturing, inspection, and acceptance of vessels as a system, and gives the methods of stress analysis and fatigue analysis. The standard has 11 chapters and 11 appendices.

We designed a QA system based on the JB4732 standard to help engineers quickly determine various common problems in the design process. Considering the modularity of the system implementation, we further divided the functional structure shown in Fig.1 into functional modules. The front end of the system includes an interaction module, and the back end of the system contains a function module, model module, and data module. The interaction module is mainly responsible for processing the input of administrators and users. The function module provides various services for the interaction module, such as log viewing, model training, user response, and QA knowledge library update. These services are implemented in the function module and encapsulated as

FIGURE 4. Interface of the QA system.

independent functions for retrieval by the interaction module. The model module mainly provides technical support for the function module, including data preprocessing, NER, and similarity matching sub-modules. The data module provides data storage services for the model module, including the standard question table, historical user question table, feedback table, and entity table.

The user interface of the system is shown in Fig.5. After the user enters the question in the search box, the system will give the corresponding candidate answer and the location of the answer in JB4732 standard. The user clicks the "jump" button to go the page in the JB4732 standard where the answer is located to help the user verify the answer.

For example, when the user enters a question belonging to the design method: "安全阀的排放面积如何计算? (How is the discharge area of the safety valve calculated?)," the similar question returned by the system is "怎么计算安全阀的排放面积? (How is the discharge area of the safety valve calculated?)." The system then outputs the location Chapter E.6.3 and a link button. Users click on the button to open the page that includes the method of calculating the discharge area. When the user enters a numerical question such as "制造单位的容器技术文件至少保存几年? (How many years must the container technical documents of the manufacturing unit be kept?)," the system directly returns the answer to this question: 5 years.

We used the system to test 1,785 problems in PVM dataset, which were categorized into four groups (design method, concept definition, numerical value, and calculation method), and the accuracy rate of the results was 98.15%. Although the system proposed in this study has substantially improved the performance of a technical QA system, the system cannot give an answer to a question entered by a user when no similar result can be found in the question-answer pairs. Therefore, in future research, we plan to integrate document understanding and reasoning technologies into the technical QA system to improve the flexibility of the system. In addition, this study has only verified the technical QA system's effectiveness in the design process of pressure vessels. Whether the QA framework proposed in this study can adapt to other products is undoubtedly worth further study.

TABLE 5. Performance comparison with other models.

Model	QQP Accuracy	QQP Precision	QQP Recall	LCQMC Accuracy	LCQMC Precision	LCQMC Recall	CCKS Accuracy	CCKS Precision	CCKS Recall
CNN	0.7291	0.7609	0.6354	0.7001	0.6742	0.7744	0.7591	0.7318	0.8175
LSTM	0.7238	0.6518	0.6234	0.7727	0.7559	0.8057	0.8661	0.8753	0.8537
BiLSTM	0.7223	0.6643	0.6358	0.7567	0.7229	0.8328	0.8404	0.8295	0.8565
IA-BiLSTM	0.7442	0.7043	0.7081	0.7809	0.7648	0.8114	0.8859	0.8808	0.8925

IV. CONCLUSION

This study proposed a template-based question generation algorithm to generate questions from multiple semantic dimensions according to a declarative sentence, so as to accumulate multiple question-answer pairs from a declarative sentence. The NER method was used to identify the entities in each declarative sentence, and then the entities were replaced with appropriate interrogative words to form question sentences from different semantic dimensions. We evaluated the generated questions from the four indicators of relevance, fluency, ambiguity and instruction. The results show that the automatically generated question answer pairs are helpful to improve the performance of the QA model. In addition, an IA-BiLSTM model integrating BiLSTM and interactive attention was designed to calculate the semantic similarity between two sentences. IA-BiLSTM can build the semantic association of words and parameters in sentences through the interactive attention between sentences, thus improving the accuracy of similarity comparison between sentences. We tested the performance of IA-BiLSTM on our own constructed data set, but also tested the performance of IA-BiLSTM on three benchmark data sets (QQP, LCQMC and CCKS). The test results indicate that IA-BiLSTM is better than CNN, LSTM and BiLSTM in accuracy, precision and recall. Finally, a QA system that integrated with the template-based question generation algorithm and the IA-BiLSTM based similarity comparison model was developed based on the ''JB4732 steel pressure vessels - analysis and design standard''. The system can enable pressure vessel designers to quickly determine the answer to common problems in the design process. Therefore, the proposed methods in this study could be applied as useful supporting tools for the construction of automatic QA systems in the mechanical manufacturing field.

The data source of the QA system supporting intelligent manufacturing in this study only includes the design standards, and ignores other real time data obtained from the Internet of Manufacturing Things (IoMT) [3], [34], [35]. In the future, the QA system incorporated with IoMT data is undoubtedly a direction worthy of in-depth research.

ACKNOWLEDGMENT

The authors would like to thank Dr. Kimberly Moravec from Liwen Bianji (Edanz) (www.liwenbianji.cn/) for editing the English text of a draft of this article.

REFERENCES

[1] ''Regulations on safety supervision of special equipmentequipment(order of the state council of the people's republic of China),'' vol. 19, no. 4, pp. 54–56, 2003.

[2] D.-H. Kim, T. J. Y. Kim, X. Wang, M. Kim, Y.-J. Quan, J. W. Oh, S.-H. Min, H. Kim, B. Bhandari, I. Yang, and S.-H. Ahn, ''Smart machining process using machine learning: A review and perspective on machining industry,'' *Int. J. Precis. Eng. Manuf.-Green Technol.*, vol. 5, no. 4, pp. 555–568, Aug. 2018, doi: 10.1007/s40684-018-0057-y.

[3] A. Smith and V. Machova, ''Internet of Things-based decision support systems, industrial big data analytics, and autonomous production processes in sustainable smart manufacturing,'' *J. Self-Governance Manage. Econ.*, vol. 9, no. 4, pp. 21–34, 2021.

[4] M. Higgins and J. Horak, ''Cyber-physical process monitoring systems, artificial intelligence-based decision-making algorithms, and sustainable industrial big data in smart networked factories,'' *Econ., Manage. Financial Markets*, vol. 16, no. 4, pp. 42–55, 2021.

[5] S. Rogers and E. Kalinova, ''Big data-driven decision-making processes, real-time advanced analytics, and cyber-physical production networks in industry 4.0-based manufacturing systems,'' *Econ., Manage. Financial Markets*, vol. 16, no. 4, pp. 84–97, 2021.

[6] A. Gordon, ''Internet of Things-based real-time production logistics, big data-driven decision-making processes, and industrial artificial intelligence in sustainable cyber-physical manufacturing systems,'' *J. Self-Governance Manage. Econ.*, vol. 9, no. 3, pp. 61–73, 2021.

[7] A. K. Sood, R. K. Ohdar, and S. S. Mahapatra, ''Parametric appraisal of mechanical property of fused deposition modelling processed parts,'' *Mater. Des.*, vol. 31, no. 1, pp. 287–295, Jan. 2010, doi: 10.1016/j.matdes.2009.06.016.

[8] K. S. Metaxiotis, D. Askounis, and J. Psarras, ''Expert systems in production planning and scheduling: A state-of-the-art survey,'' *J. Intell. Manuf.*, vol. 13, no. 4, pp. 253–260, 2002, doi: 10.1023/A:1016064126976.

[9] B. Dolšak, I. Bratko, and A. Jezernik, ''Knowledge base for finite-element mesh design learned by inductive logic programming,'' *Artif. Intell. Eng. Des., Anal. Manuf.*, vol. 12, no. 2, pp. 95–106, Apr. 1998, doi: 10.1017/S0890060498122023.

[10] K. M. A. El-Ghany and M. M. Farag, ''Expert system to automate the finite element analysis for non-destructive testing,'' *NDT E Int.*, vol. 33, no. 6, pp. 409–415, 2000, doi: 10.1016/S0963-8695(00)00009-8.

[11] W. Yu, L. Wu, Y. Deng, R. Mahindru, Q. Zeng, S. Guven, and M. Jiang, ''A technical question answering system with transfer learning,'' in *Proc. Conf. Empirical Methods Natural Lang. Processing, Syst. Demonstrations*, 2020, pp. 92–99, doi: 10.18653/v1/2020.emnlp-demos.13.

[12] Y. Lai, Y. Lin, J. Chen, Y. Feng, and D. Zhao, ''Open domain question answering system based on knowledge base,'' in *Natural Language Understanding and Intelligent Applications*. Cham, Switzerland: Springer, 2016, pp. 722–733.

[13] M. A. C. Soares and F. S. Parreiras, ''A literature review on question answering techniques, paradigms and systems,'' *J. King Saud Univ. Comput. Inf. Sci.*, vol. 32, no. 6, pp. 635–646, Jul. 2020, doi: 10.1016/j.jksuci.2018.08.005.

[14] B. F. Green, A. K. Wolf, C. Chomsky, and K. Laughery, ''Baseball: An automatic question-answerer,'' in *Proc. Papers Presented Western Joint IRE-AIEE-ACM Comput. Conf. IRE-AIEE-ACM (Western)*, May 1961, pp. 219–224, doi: 10.1145/1460690.1460714.

[15] L. Kodra and E. Kajo, ''Question answering systems: A review on present developments, challenges and trends,'' *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 9, pp. 217–224, 2017.

[16] T. Sultana and S. Badugu, ''A review on different question answering system approaches,'' in *Advances in Decision Sciences, Image Processing, Security and Computer Vision*. Cham, Switzerland: Springer, 2020, pp. 579–586.

[17] Z. Huang, S. Xu, M. Hu, X. Wang, J. Qiu, Y. Fu, Y. Zhao, Y. Peng, and C. Wang, "Recent trends in deep learning based open-domain textual question answering systems," *IEEE Access*, vol. 8, pp. 94341–94356, 2020, doi: 10.1109/ACCESS.2020.2988903.

[18] T. Lei, Z. Shi, D. Liu, L. Yang, and F. Zhu, "A novel CNN-based method for question classification in intelligent question answering," in *Proc. Int. Conf. Algorithms, Comput. Artif. Intell.*, Dec. 2018, pp. 1–6, doi: 10.1145/3302425.3302483.

[19] W. Xia, W. Zhu, B. Liao, M. Chen, L. Cai, and L. Huang, "Novel architecture for long short-term memory used in question classification," *Neurocomputing*, vol. 299, pp. 20–31, 2018, doi: 10.1016/j.neucom.2018.

[20] C. L. A. Clarke and E. L. Terra, "Passage retrieval vs. Document retrieval for factoid question answering," in *Proc. 26th Annu. Int. ACM SIGIR Conf. Res. Develop. Informaion Retr. (SIGIR)*, 2003, pp. 427–428.

[21] O. Khattab, C. Potts, and M. Zaharia, "Relevance-guided supervision for OpenQA with ColBERT," *Trans. Assoc. Comput. Linguistics*, vol. 9, pp. 929–944, Sep. 2021, doi: 10.1162/tacl_a_00405.

[22] A. McCallum, M. Zaheer, R. Das, and S. Dhuliawala, "Multi-step retriever-reader interaction for scalable open-domain question answering," in *Proc. Int. Conf. Learn. Represent.*, Sep. 2019.

[23] Y. Feldman and R. El-Yaniv, "Multi-hop paragraph retrieval for open-domain question answering," in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics*, 2019, pp. 2296–2309.

[24] X. Qiu, T. Sun, Y. Xu, Y. Shao, N. Dai, and X. Huang, "Pre-trained models for natural language processing: A survey," *Sci. China Technol. Sci.*, vol. 63, no. 10, pp. 1–26, 2020, doi: doi:10.1007/s11431-020-1647-3.

[25] M. Zaib, D. H. Tran, S. Sagar, A. Mahmood, W. E. Zhang, and Q. Z. Sheng, "BERT-CoQAC: BERT-based conversational question answering in context," in *Proc. Int. Symp. Parallel Archit., Algorithms Program.* Cham, Switzerland: Springer, 2010, pp. 47–57.

[26] W. Sakata, T. Shibata, R. Tanaka, and S. Kurohashi, "FAQ retrieval using query-question similarity and BERT-based query-answer relevance," in *Proc. 42nd Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Jul. 2019, pp. 1113–1116.

[27] A. Reusch, M. Thiele, and W. Lehner, "An ALBERT-based similarity measure for mathematical answer retrieval," in *Proc. 44th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Jul. 2021, pp. 1593–1597.

[28] C. Qu, L. Yang, M. Qiu, W. B. Croft, Y. Zhang, and M. Iyyer, "BERT with history answer embedding for conversational question answering," in *Proc. 42nd Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Jul. 2019, pp. 1133–1136.

[29] Q.-L. Guo, "A novel approach for agent ontology and its application in question answering," *J. Central South Univ. Technol.*, vol. 16, no. 5, pp. 781–788, Oct. 2009, doi: 10.1007/s11771-009-0130-3.

[30] H. Li and S. Ding, "Research on product design domain knowledge Qa system based on ontology and design scenario (in Chinese)," *Inf. Theory Pract.*, vol. 38, no. 1, pp. 130–134, 2015.

[31] S. Li, S. Zhao, B. Cheng, and H. Yang, "An end-to-end multi-task learning model for fact checking," in *Proc. 1st Workshop Fact Extraction VERification (FEVER)*, 2018, pp. 138–144.

[32] J. Wang, Y. Ma, L. Zhang, R. X. Gao, and D. Wu, "Deep learning for smart manufacturing: Methods and applications," *J. Manuf. Syst.*, vol. 48, pp. 144–156, Jul. 2018, doi: 10.1016/j.jmsy.2018.01.003.

[33] M. Shi, "Knowledge graph question and answer system for mechanical intelligent manufacturing based on deep learning," *Math. Problems Eng.*, vol. 2021, Feb. 2021, Art. no. 6627114, doi: 10.1155/2021/6627114.

[34] G. H. Popescu, S. Petreanu, B. Alexandru, and H. Corpodean, "Internet of Things-based real-time production logistics, cyber-physical process monitoring systems, and industrial artificial intelligence in sustainable smart manufacturing," *J. Self-Governance Manage. Econ.*, vol. 9, no. 2, pp. 52–62, 2021, doi: 10.22381/jsme9220215.

[35] E. Hopkins and A. Siekelova, "Internet of Things sensing networks, smart manufacturing big data, and digitized mass production in sustainable industry 4.0," *Econ., Manage. Financial Markets*, vol. 16, no. 4, pp. 2–41, 2021, doi: 10.22381/emfm16420212.

LIU XINGGUANG is currently pursuing the master's degree with the Zhejiang University of Technology. His research interests include natural language processing and deep learning.

CHENG ZHENBO received the Ph.D. degree in computer science and technology from Tsinghua University, in 2012. He is currently an Associate Professor with the School of Computer Science and Technology, Zhejiang University of Technology. His main research interests include natural inspired intelligent algorithms, reinforcement learning, and decision-making.

SHEN ZHENGYUAN received the master's degree from the Zhejiang University of Technology. He is works with Shenzhen Tencent Computer Systems Company Ltd. His research interests include natural language processing and deep learning.

ZHANG HAOXIN is currently pursuing the master's degree with the Zhejiang University of Technology. His research interests include mechanical intelligent manufacturing.

MENG HANGCHENG is currently pursuing the master's degree with the Zhejiang University of Technology. His research interest includes mechanical intelligent manufacturing.

XU XUESONG received the Ph.D. degree from the Zhejiang University of Technology. His main research interest includes mechanical intelligent manufacturing.

XIAO GANG is a Professor with the School of Computer Science and Technology, Zhejiang University of Technology. He is also a Vice President with China Jiliang University. His research interests include manufacturing informatization, intelligent information system, and graphics and image processing.