Disease-Pertinent Knowledge Extraction in Online Health Communities Using GRU Based on a Double Attention Mechanism

YANLI ZHANG\(^1,2\), XINMIAO LI\(^1\), AND ZHE ZHANG\(^3\)

\(^1\)School of Information Management and Engineering, Shanghai University of Finance and Economics, Shanghai 200433, China
\(^2\)School of Information Management, Zhengzhou University of Aeronautics, Zhengzhou 450015, China
\(^3\)China Financial Computerization Corporation, Beijing 100076, China

Corresponding authors: Yanli Zhang (zhangyanli_sufe@126.com) and Xinmiao Li (xinmiaoli@126.com)

The work of Yanli Zhang was supported in part by the National Natural Science Foundation of China under Grant 91646205, and in part by the Shanghai University of Finance and Economics (SHUFE), under Grant CXJJ-2018-396.

ABSTRACT

Relationship extraction among diseases, symptoms and tests has always been a concerning research issue in the biomedical field. Disease-pertinent relationship extraction for user-generated content in the online health community represents a research trend. By training the word embedding vectors for the medical-health field, conducting entity recognition and relationship annotation, and using deep learning technology, we construct a relation extraction model for extracting the relationships among diseases, symptoms and tests. Our relationship extraction model of the bidirectional gate recurrent unit (BiGRU) network based on character-level and sentence-level attention mechanisms achieved the best results on question-answer data in the online health community. Our research results can not only help physician diagnoses but also help patients perform health management, which has important industrial application value.

INDEX TERMS

Knowledge extraction, relationship extraction, attention mechanism, deep learning, bidirectional gated recurrent unit (BiGRU), online health community.

I. INTRODUCTION

There are shortages and imbalances in medical resources around the world. The emergence of online health communities (OHCs) has led to an increasing number of users using OHCs for question-and-answer (Q&A) consultation, which has partly alleviated the problem of resource shortages and imbalances to some extent [1]. Simultaneously, the existing disease-pertinent medical knowledge bases cannot be invariable and need to be updated over time [2]. There is a considerable quantity of user-generated content in OHCs: users’ health consultation and physicians’ replies, etc. Knowledge extraction among diseases, symptoms and tests on these Q&A data is of great significance for supplementing and improving the existing knowledge base [3]. However, the physician-patient Q&A data in OHC contains considerable redundant information. This unextracted medical information has also caused resource waste. How to extract valuable information from these data has become an important issue for researchers. During patient visits, one of the greatest concerns is the disease, symptoms and test. Knowledge extraction from the Q&A OHC among diseases, symptoms and tests has increased access for obtaining health knowledge, which can help patients understand more medical information before going to the hospital, improve medical treatment efficiency, and improve the existing knowledge bases of diseases, symptoms and tests to assist physicians in decision-making.

There are many studies on relationship extraction in the biomedical health domain, but most of them are based on electronic medical records [4]–[6], discharge abstracts [7]–[10] and medical literature abstracts [5], [11], [12]. The words and sentences in these corpora are relatively rigorous and standardized. However, few scholars study the relationship extraction in texts with word colloquialisms and are poorly structured in physician-patient OHC Q&A. Therefore, we conducted relationship extraction among diseases, symptoms, and tests for the Q&A data in OHCs. Based on popular deep learning techniques and attention mechanisms for identifying better semantic relationships, we have developed an attention-based BiGRU (bidirectional gated recurrent unit network) model for relationship extraction among diseases, symptoms and tests in OHCs.
II. RELATED WORK

Medical knowledge is very vast. How to extract valuable medical knowledge from different kinds of medical texts is a research hotspot and one of the problems to be solved in practice, and it is very important for improving public health and knowledge management. Biomedical relationship extraction is an important task in biomedical text mining. It has important applications in the fields of gene-disease and protein-protein interactions, such as generated disease-pertinent treatment vocabulary from MEDLINE citations [13], chemical-induced disease relationship extraction from PubMed articles [11], and relationship extraction between proteins [14], [15].

Diseases, symptoms and tests are the primary concerns of healthcare and have aroused widespread concern. Scholars have made many explorations by extracting the relationship among diseases, symptoms and tests from medical texts. We first carried out entity recognition (for the unlabeled text (diseases, symptoms and tests), the system had difficulty recognizing these entities), and then the relationship was extracted based on entity recognition. The I2B2/VA challenge task in 2010 organized to extract the relationship among the medical concepts of patients’ clinical records, annotated medical problems (diseases, symptoms), tests and treatments, and machine learning were used to extract the following three relationships: medical problems and tests; medical problems and treatment; medical problems and medical problems [9].

The methods of pattern matching and machine learning are commonly used for relationship extraction among medical texts. The relationship extraction method of pattern matching refers to expert-defined rules based on syntactic analysis; however, the extraction result usually has a lower recall. Song et al. [16] implemented a comprehensive text mining system, extracting multiple relationships based on rules from biomedical literature of the MEDLINE database. Khoo et al. [12] developed a system for identifying and extracting disease-pertinent causal relationships in the literature abstract of the MEDLINE database and constructed a set of graphical patterns using syntactic analysis trees to match causal relationships in sentences. In addition, the methods for supporting vector machines and kernel functions have been applied to relationship extraction in the field of biomedicine. Peng et al. [11] extracted the relationship between chemicals and diseases from the titles and abstracts of PubMed literature using the rich features of support vector machines. Bunescu and Mooney [14] innovatively applied the kernel function method to extract the relationships of protein interactions. Frunza et al. [17] used a variety of machine learning algorithms to extract the three semantic relationships among disease and treatment from the titles and abstracts of MEDLINE literature: cure, prevention, and side effects. Uzuner et al. [10] studied the semantic relationship classifications of the discharge summaries, defined and extracted the relationships related to the patient’s medical problems (disease, symptom, test and treatment), and achieved better effects in relationship classification. Shen et al. [6] used the deep learning method to extract the relationships between diseases and complications, and the relationships between diseases and symptoms in electronic health records. The relationship extraction corpus of the abovementioned research comes from relatively structured texts such as discharge summaries or medical literature abstracts. However, few scholars have studied how to extract the relationship among diseases, symptoms and tests from the word colloquialisms and less structured texts in OHCs.

The machine learning method requires a certain amount of domain knowledge and a large number of artificial features in the relationship extraction task and requires higher manual effort. However, the deep learning method can automatically extract features in relation classification, eliminating considerable manual effort, and has achieved better results in many fields. Socher et al. [18] used a recurrent neural network (RNN), which performed better on three different relational classification tasks (emotional classification, causality, etc.). Zeng et al. [19] applied remote supervision piecewise convolutional neural networks (CNNs) to the relationship classification on the NYT corpus, which solved the problem of lack of labeled data and incorrect labels to some extent. Santos et al. [20] used the classification ranking CNN (CR-CNN) method with a new pairwise ranking loss function to classify the relationship on the SemEval-2010 Task 8 dataset, which yielded good results. Zhou et al. [21] used the attention-based long short-term memory (LSTM) model to show better classification results on the SemEval-2010 dataset. The GRU model simplifies the LSTM model with higher execution efficiency and is increasingly used in relation extraction tasks. Li et al. [22] used the GRU model to extract the Bacteria Biotope event from the biomedical literature on the BioNLP’16 corpus, and the results confirm the architecture’s validity. Luo et al. [23] used the GRU model to extract the geological data relations and achieved a satisfactory result. Shen et al. [6] used the GRU framework to extract the relationship between disease and complications and the relationship between disease and symptoms in electronic health records. However, few or no scholars have studied how to use deep learning technology to extract the relationship among diseases, symptoms and tests from these word colloquialisms and unstructured texts in OHCs.

Research on OHCs has focused on community theme analysis, sentiment analysis, gender analysis, post classification, physicians’ service charges, value creation from urban to rural regions, and so on [24]–[26]. There are also fewer studies that performed relationship extraction in OHCs, such as using the MedDRA (Medical Dictionary for Regulatory Activities) and the SpanishDrugEffectDB database to extract the relationship between drugs and effects [27]. Eftimov et al. [28] used the rule-based method to extract dietary recommendations knowledge from online health websites. However, the relationship extraction between the medical entities related to disease diagnosis is mostly based on more specialized, relatively normalized terms, such as electronic medical records, discharge abstracts and
medical literature abstracts, the corpus size is mostly small, and the obtained knowledge by relationship extraction is relatively limited; the application effect is not very satisfactory on the large sample corpus with word colloquialisms and unstructured text. However, there are massive physician-patient Q&A data generated by hundreds of millions of users in OHCs. If these data can be used well, valuable knowledge can be extracted, which will be very helpful for improving the existing medical knowledge base and assist clinical decision support.

III. METHODOLOGY

Our relationship extraction network framework of 2ATT-BiGRU is shown in Figure 1 and was derived based on previous research [21], [22]. The main modules are described as follows, containing input layer, word embedding layer, BiGRU with character attention layer, BiGRU with sentence attention layer, and output layer.

A. TEXT REPRESENTATION AND WORD EMBEDDINGS

Word embeddings are word distributed representations that map each word in the corpus into a low-dimensional vector, which contains rich semantic information considering the contextual semantic environment of the words. Words are represented by word embedding before relationship classification. We use the physician-patient Q&A data in OHCs and tens of thousands of biomedical studies to train the word embedding vectors of word representations, which includes a wealth of entity information of the biomedical domain.

In the relationship extraction task, words that are close to the entity can more clearly highlight the relationship between two entities. Thus, to more accurately express the semantic information, the position embedding is derived from the relative distance of each word to target entity 1 and target entity 2 in a sentence [29]. The distance is converted to an offset value, taking the following sentence as an example:

[Chest pain] symptom is a common clinical manifestation of [cardiovascular disease] disease.

The relative distances of “is” to entity 1 “chest pain” and entity 2 “cardiovascular disease” are −1 and 6, respectively. The relative distance is mapped to a low-dimensional vector and initialized. The obtained position embedding is incorporated into the word representation vectors.

B. BIGRU

The gated recurrent unit (GRU) was proposed by Cho et al. [30], which merges the forget gates and input gates of LSTM into an update gate; thus, the GRU only has the update gate and reset gate. LSTM architecture is good at handling relationship classification tasks of text sequences and machine translations. GRU simplifies the LSTM model [22], [31] and performs more efficiently than GRU, as shown in Figure 2.

The update gate \( z_t \) determines the extent to which information of the previous state is to be forgotten and which information of the new content is to be added. The reset gate \( r_t \) controls the extent to which the previous hidden state and the current input is ignored. At the same time, the GRU merges the cell state with the hidden state.

C. ATTENTION MECHANISM

By analyzing the corpus, the influence of different input sequences on the output is different in semantic information. Some words have an important influence on the output, whereas some words are irrelevant. We use the attention mechanism to find words that have a significant impact on the output, giving it a higher weight, so that its semantic information can be fully obtained. In our study, we used the character-level and sentence-level attention mechanism to extract the relationship among diseases, symptoms and tests [21], [23].

Character-level attention mechanism: \( H = \frac{D'_f + D'_b}{2} \) (1)

\( f \) and \( b \) represent the forward and backward networks, respectively.

FIGURE 1. The model of 2ATT-BiGRU.

FIGURE 2. GRU structure model.
BiGRU layer, and T represents the sentence’s length. The sentence’s eigenvector $r$ is obtained by the sum of the output vectors multiplied by the weight.

\[
\begin{align*}
M &= \tanh(H) \quad (2) \\
\alpha &= \text{softmax}(\omega^T M) \quad (3) \\
\gamma &= H\alpha^T \quad (4)
\end{align*}
\]

where $M$ is the state after the activation function; $\alpha$ is the obtained attention weight; $\gamma$ is the output vector after weight summation; $H \in \mathbb{R}^{d \times T}$, $d^\omega$ is the word embedding, $\omega$ is the parameter vector after training; and the dimensions of $\omega$, $\alpha$, $\gamma$ are $d^\omega$, $T$, $d^\omega$.

Finally, the sentence representation for character-level attention is obtained from:

\[
h^* = \tanh(\gamma) \quad (5)
\]

Sentence-level attention mechanism: To prevent excessive noise data from affecting the relationship extraction, the 2ATT-BiGRU model adds the sentence-level attention mechanism based on the BiGRU character-level structure and assigns a higher weight to the sentences that truly reflect the relationship between entities. Set $S$, as the n sentences covered by the relational entity pair, $S = \{x_1, x_2, \ldots, x_n\}$.

\[
s = \sum_{i} a_i x_i \quad (6) \\
\alpha_i = \frac{\exp(e_i)}{\sum_k \exp(e_k)} \quad (7) \\
e_i = x_i A r \quad (8)
\]

$s$ is the weighted sum of the sentences’ vector set; $\alpha_i$ is the attention weight of sentence-level vector; $e_i$ is a function that scores the input sentence $x_i$ and predicting relationship $r$; $A$ is a weighted diagonal matrix; $r$ is a vector that indicates the relationship $r$. Finally, the sentence representation used for the relationship classification is obtained from the following equation:

\[
h^*_s = \sum_i h^* a_i \quad (9)
\]

**D. THE EXECUTION PROCESS OF OUR MODEL**

1) INPUT LAYER
Enter the sentences of the physician-patient Q&A data, such as the annotated relationship disease-has-symptom: [Chest pain]symptom is a common clinical manifestation of [cardiovascular disease]disease.

2) EMBEDDING LAYER
Words after sentence segmentation are represented by word embedding concatenation. The word embedding is trained by a domain-related word representation model, and position embedding is concatenated to obtain more semantic information.

3) BiGRU LAYER
Low-dimensional vectors are trained in the bidirectional (forward and backward) GRU network, and more semantic features are extracted to obtain multidimensional vectors.

4) CHARACTER-LEVEL ATTENTION LAYER
Generates a weight vector by a function, and the sentence representation for character-level attention is obtained by multiplying by the weight vector.

5) SENTENCE-LEVEL ATTENTION MECHANISM
On the basis of the character-level attention level, a sentence-level attention mechanism is introduced, which assigns a higher weight to the sentence that truly reflects the relationship classification and a lower weight for the noise sentence.

6) OUTPUT LAYER
We regard relationship extraction as a classification task. The sentence vector after the character-level and sentence-level attention mechanism is classified by the softmax function classifier to output the probability of the predicted relationship type of the entity pairs.

\[
\hat{P}(y|S) = \text{softmax} \left( W^{(o)} h^*_s + b^{(o)} \right) \quad (10) \\
\hat{y} = \arg\max P(y|S) \quad (11)
\]

Taking the above sentence as an example, in that sentence, entity 1 is “chest pain”, and entity 2 is “cardiovascular disease”. The input sentences are converted to word embeddings and position embeddings of word representation through the bidirectional GRU layer and by the weighted distribution of attention mechanisms. Then, the relationship between two entities is classified by the softmax function. Finally, our classification model determined that the relationship type of the maximum probability is disease-has-symptom. To avoid the loss of important information in the process of automatic feature extraction, the 2ATT-BiGRU model is applied to the problem of medical relationship extraction in the physician-patient Q&A data, combining the character-level and sentence-level attention mechanism with the BiGRU model, considering the correlation between the input sequence and the output sequence, which adds more semantic information to the extracted feature vector because it uses the bidirectional context and improves the performance of the relation extraction model in accuracy and stability.

To evaluate the experimental results, we define the prediction value and annotated value as 1 and the inconsistency as 0. We use F (F-score), R (recall) and P (precision) to evaluate our results, calculated as follows:

\[
P = \frac{TP}{TP + FP} \quad (12) \\
R = \frac{TP}{TP + FN} \quad (13) \\
F - score = \frac{2 \times P \times R}{P + R} \quad (14)
\]
TP stands for true positives, FP stands for false positives, and FN stands for false negatives.

For the overall performance of the model, we use the microaverage evaluate:

\[
\text{microP} = \frac{\sum_c TP(c)}{\sum_c TP(c) + \sum_c FP(c)} \tag{15}
\]

\[
\text{microR} = \frac{\sum_c TP(c)}{\sum_c TP(c) + \sum_c FN(c)} \tag{16}
\]

\[
\text{microF-score} = \frac{2 \times \text{microP} \times \text{microR}}{\text{microP} + \text{microR}} \tag{17}
\]

IV. EXPERIMENT SETUP AND RESULTS DISCUSSION

A. DATA DESCRIPTION

There are two datasets in our relationship extraction task. One is the Ask and Answer website (www.120ask.com, abbreviated as 120ask), China’s most popular question-and-answer (Q&A) OHC, and more than 160 million Q&A data provided by 52 million users for the three years from 2014 to 2016 were preprocessed. The other is the obtained 17 million question-and-answer pairs generated by 5 million users in 2016 on the Seeking Medical Advice website (www.xywy.com, abbreviated as xywy, another popular Q & A health community in China). Taking cardiovascular disease as an example, after executing the task of entity recognition for the physician-patient Q&A data, four relationships are annotated for model training, testing, and predictions. The relationships are (1) disease-has-symptom (DhSYM); (2) disease-suits-test (DsTES); (3) symptom-suits-test (SsTES); and (4) others, i.e., no relationship or fuzzy relationship between entities. The diagram is shown in Figure 3. Among these entities, diseases refer to an unhealthy state or physician’s diagnosis, symptoms are discomfort signs caused by illness, and tests are equipment or item examinations taken to confirm an illness.

For the unstructured physician-patient Q&A data, due to insufficient annotation data, we attempt to use the bootstrapping method to obtain more seed relationships or seed patterns between entity pairs through iterative extraction, which is shown in Fig. 4. The bootstrapping method has the phenomena of higher recall, lower precision, and semantic drift; thus, we need to control the pattern, which is used to find these relationship types to be extracted among diseases, symptoms and tests. After manual review, the relationships are annotated as our experimental corpus of relationship extraction, and then the 2ATT-BiGRU model is used to train and predict the relationship type.

We take the Q&A data with cardiovascular disease as an example of training and testing. There are 520,966 cardiovascular disease Q&A records generated by 46,673 users in the three years in the 120ask Q&A dataset and 165,380 cardiovascular disease Q&A records generated by 22,168 users in the xywy Q&A dataset. We annotate 8,820 relationships in the 120ask Q&A dataset and annotate 8,038 relationships in the xywy Q&A dataset, while dividing the training set and test set according to the proportion of 3:1. The training data are input into the model for training, and then the parameters of the model are obtained. Finally, we obtain the results on the test set. Our annotation data are shown in Table 1.

B. PARAMETERS SETUP

Our experiment uses the program development language of Python3 and the open source learning library of TensorFlow for relationship extraction. Word embeddings are obtained by the word presentation model. Finally, our 2ATT-BiGRU sequence model attempts to obtain a classification probability by model training and parameter tuning. We adjust these parameters in the appropriate range according to experience and perform a superparametric training process: the dimension of word embedding ∈ (100, 200, 300); batch size ∈ (32, 64, 128); epoch ∈ (10, 20, 30); the number of hide nodes ∈ (120, 240, 360) and hidden layer number 1; learning rate ∈ (0.0005, 0.001, 0.002); dropout ∈ (0.3, 0.5, 0.8). We obtained the best results for the parameters as follows:
word embedding 200, position embedding 50, batch size 64, epoch 30, hidden layer node number 240, learning rate 0.002, dropout 0.5 and hidden layer number 1. The effect of each parameter on the relationship classification is shown in Figure 5. The word embedding, epoch and learning rate on the relationship classification effects show a positive correlation effect. The batch size and dropout values show a negative correlation effect. The relationship between the number of hidden layer nodes and the result is not clear.

![Figure 5. The difference in parameter effects on the results.](image)

C. EXPERIMENT RESULTS

Using the 2ATT-BiGRU model, the relationship extraction results are shown in Table 2.

| Category | Precision | Recall | F-Score |
|----------|-----------|--------|---------|
| 120ask dataset | | | |
| DhSYM | 86.31% | 84.22% | 85.25% |
| DsTES | 66.25% | 60.51% | 63.25% |
| SsTES | 60.82% | 57.45% | 59.08% |
| Others | 68.52% | 66.49% | 67.49% |
| Total | 78.35% | 76.89% | 77.61% |
| xywy dataset | | | |
| DhSYM | 84.87% | 85.69% | 85.28% |
| DsTES | 70.19% | 65.27% | 67.64% |
| SsTES | 64.35% | 61.52% | 62.90% |
| Others | 71.85% | 67.24% | 69.47% |
| Total | 79.16% | 78.02% | 78.59% |

According to the experimental results, the precision, recall and F-Score of our BiGRU-2ATT model performs best compared with other models in relational classification: the precision reaches 78.35%, the recall reaches 76.89%, and the F-Score reaches 77.61%. Compared with the baseline (LSTM) model, our 2ATT-BiGRU model is approximately 8% higher than the F-Score of the baseline in the 120ask dataset; the precision reaches 79.16%, the recall reaches 78.02%, and the F-Score reaches 78.59%. Compared with the baseline (LSTM) model, our 2ATT-BiGRU model is approximately 8% higher than the F-Score of the baseline in the xywy dataset.

1) INFLUENCE OF THE DOUBLE ATTENTION MECHANISM

All hidden states are treated equally before the attention mechanism; thus, it is unreasonable. Some words or sentences play an important decision-making role in the relationship classification, and they should be given higher weights in the relationship classification judgment, making relationship classification more accurate; therefore, the weight distribution using the character-level and sentence-level attention mechanism plays an important role in relationship classification. Our experimental results show that the model with the double attention mechanism showed better performance: the F-Score of the 2ATT-BiGRU model is approximately 6% higher than the GRU model, which is 3% higher than the BiGRU-CharLevel model and is 5% higher than the BiGRU-SentLevel model in the 120ask dataset; the F-Score of the 2ATT-BiGRU model is approximately 7% higher than the GRU model, which is approximately 5% higher than the BiGRU-CharLevel model and is approximately 6% higher than the BiGRU-SentLevel model in the 120ask dataset.

TABLE 3. Comparison results with other relationship classification models.

| Category | Precision | Recall | F-Score |
|----------|-----------|--------|---------|
| 120ask dataset | | | |
| LSTM (Baseline) | 68.90% | 70.75% | 69.81% |
| GRU | 72.92% | 70.55% | 71.72% |
| BiGRU-CharLevel | 75.62% | 73.25% | 74.42% |
| BiGRU-SentLevel | 72.72% | 71.71% | 72.21% |
| 2ATT-BiGRU | 78.35% | 76.89% | 77.61% |
| xywy dataset | | | |
| LSTM (Baseline) | 68.69% | 71.21% | 69.93% |
| GRU | 73.71% | 70.12% | 71.87% |
| BiGRU-CharLevel | 74.11% | 73.32% | 73.71% |
| BiGRU-SentLevel | 73.16% | 72.12% | 72.64% |
| 2ATT-BiGRU | 79.16% | 78.02% | 78.59% |
TABLE 4. Some examples of annotation and prediction results on the Q&A data relationship extraction task in OHC.

| Head entity | Tail entity | Q&A sentences | True relation | The top 3 prediction relations(probability) |
|-------------|-------------|---------------|---------------|------------------------------------------|
| 冠心病       | 胸闷心悸     | 根据你的描述，是 冠心病 所致。有 胸闷心悸 等表现，严重时表现为心绞痛，在发作时心率大约为 70——100。发作时下心率加快，建议：平时可以使用阿司匹林，美托洛尔等治疗。有心绞痛时用 硝酸甘油 舌下含服缓解。有条件的话可行冠状动脉搭桥术，祝您健康！ | DhSYM | 1. DhSYM(0.965542) 2. StTES (0.268821) 3. DeTIES (0.112534) |
| 冠状动脉疾病 | 胸闷心悸     | 根据您的描述，是 冠状动脉疾病 所致。有 胸闷心悸 等表现，严重时表现为心绞痛，在发作时心率大约为 70——100。发作时下心率加快，建议：平时可以使用阿司匹林，美托洛尔等治疗。有心绞痛时用 硝酸甘油 舌下含服缓解。有条件的话可行冠状动脉搭桥术，祝您健康！ | DhSYM | 1. DhSYM(0.965542) 2. StTES (0.268821) 3. DeTIES (0.112534) |

According to your description, it is caused by 冠状动脉疾病. Symptoms such as 胸闷心悸 and 心悸, are approximately 70-100 when there is no attack. In the case of an attack, the heart rate is accelerated. Suggestions: usually, you can use aspirin, metoprolol, etc. When you have angina pectoris, you can take nitroglycerin under your tongue to gain relief. If conditions persist, coronary artery bypass surgery is available. I wish you health! |

2) GRU FRAMEWORK EFFECTS
The GRU model simplifies the LSTM model and is more effective, which is approximately 2% higher than baseline both in the 120ask dataset and in the xywy dataset. On the basis of the GRU model, the character-level attention mechanism and the sentence-level attention mechanism are added separately; the F-score rises on the basis of previous models, and the character-level attention mechanism is more effective. The results show that the 2ATT-BiGRU framework is very effective in feature discovery and relationship classification, which is mainly due to domain-related word embedding and the proposed attention mechanism.

D. 2ATT-BiGRU MODEL APPLICATION
We randomly select a few examples and use our model to predict the relationship between entity pairs of our dataset in OHC. The obtained results are shown in Table 4. As shown in Table 4, our relation classification model of 2ATT-BiGRU has better predictive effects on relationship classification. Therefore, it can be used to extract knowledge in OHC and improve existing knowledge bases.
Taking hypertension as an example, the final relation extraction results among the diseases, symptoms, and tests are shown in Figure 6. Hypertension is associated with multiple symptoms and tests. Hypertension, symptoms, and tests are represented by orange, yellow, and green, respectively. The straight line represents the relationship between entities. Symptoms and tests associated with hypertension also have relationships with other entities, such as certain symptoms related to other diseases and other tests. Taking the symptom “chest pain” as an example, the relationship among symptoms, diseases, and tests is shown in Figure 7.

Our relationship extraction model can capture new, evolving knowledge from ever-changing Q&A data in OHCs, thus enabling continuous improvement and enhancement of existing knowledge bases. Therefore, it has important implications for improving existing knowledge bases.

V. CONCLUSION AND FUTURE WORK

In our research, we proposed a 2ATT-BiGRU network architecture based on character-level and sentence-level attention mechanisms for relationship extraction among diseases, symptoms, and tests. This model obtains important contextual grammatical and semantic feature information from the biomedical field by domain-related word embedding training and bidirectional GRU, which does not require complex manual design features. Finally, our model has a microaverage of 78.35% in precision, 76.89% in recall, 77.61% in F-score on relational extraction, and the F-score is approximately 8% higher than the baseline in the 120ask dataset, which has a microaverage of 79.16% in precision, 78.02% in recall, 78.59% in F-score on relational extraction, and the F-score is approximately 9% higher than the baseline in the xywy dataset. The experimental results confirm our model’s effectiveness.

Since the existing disease-pertinent relationship extraction is mostly based on electronic medical records [4]–[6], discharge abstracts [7]–[10] and medical literature abstracts [5], [11], [12], the greatest advantage of our research is the unstructured, large-scale physician-patient Q&A data in OHCs, which realized the best results by our model. Additionally, the domain-related word embeddings that we trained and the attention mechanisms based on character-level and sentence-level play an important role in relation extraction.

Our relationship extraction results complement the existing knowledge base, and the extracted knowledge provides assistance for physicians’ medical diagnosis, which also helps patients with self-diagnosis before medical treatment and helps them conduct beneficial daily health management. Therefore, it has important practical significance.

The relation extraction model in our study achieved good results, but it still has some limitations. Our research does not consider coreference resolution in relation extraction among diseases, symptoms, and tests in the Q&A data in OHC. In future research, we will consider the technology of coreference resolution to further improve the effectiveness of disease-pertinent relationship extraction in online healthy communities.

REFERENCES

[1] H. Wu and N. Lu, “Service provision, pricing, and patient satisfaction in online health communities,” Int. J. Med. Inform., vol. 110, pp. 77–89, Feb. 2018.
[2] Y. Wang, L. Wang, M. Rastegar-Mojarad, S. Moon, F. Shen, N. Afzal, S. Liu, Y. Zeng, S. Mehrabi, S. Sohn, and H. Liu, “Clinical information extraction applications: A literature review,” J. Biomed. Inform., vol. 77, pp. 34–49, Jan. 2018.
[3] X. Lv, Y. Guan, J. Yang, and J. Wu, “Clinical relation extraction with deep learning,” Int. J. Hybrid Inf. Technol., vol. 9, no. 7, pp. 237–248, 2016.
[4] F. Li and H. Yu, "An investigation of single-domain and multidomain medication and adverse drug event relation extraction from electronic health record notes using advanced deep learning models," J. Amer. Med. Inform. Assoc., vol. 26, no. 7, pp. 646–654, Jul. 2019.

[5] G. Atri, R. Safdari, H. Rezaeezadeh, A. Abbassian, M. Mokhtarani, and M. H. Ayati, "A systematic literature review and classification of knowledge discovery in traditional medicine," Comput. Methods Programs Biomed., vol. 168, pp. 39–57, Jan. 2019.

[6] Y. Shen, L. Zhang, J. Zhang, M. Yang, B. Tang, Y. Li, and K. Lei, "CBN: Constructing a clinical Bayesian network based on data from the electronic medical record," J. Biomed. Inf., vol. 88, pp. 1–10, Dec. 2018.

[7] R. Patel and S. Tanwani, "Application of machine learning techniques in clinical information extraction," in Smart Techniques for a Smarter Planet (Studies in Fuzziness and Soft Computing), M. Mishra, B. Mishra, Y. Patel, and R. Misra, Eds., Cham, Switzerland: Springer, 2019, pp. 145–165.

[8] Y. Tang, J. Yang, P. S. Ang, S. R. Dorajoo, B. Foo, S. Soh, S. H. Tan, M. Y. Tham, Q. Ye, L. Shek, C. Sung, and A. Tung, "Detecting adverse drug reactions in discharge summaries of electronic medical records using readdeep," Int. J. Med. Inform., vol. 128, pp. 62–70, Aug. 2019.

[9] O. Uzuner, B. R. South, S. Shen, and S. L. DuVall, "2012 i2b2/VA challenge on concepts, assertions, and relations in clinical text," J. Amer. Med. Inform. Assoc., vol. 18, no. 5, pp. 552–556, Sep./Oct. 2011.

[10] O. Uzuner, J. Mailoa, R. Ryan, and T. Sibanda, "Semantic relations for problem-oriented medical records," Artif. Intell. Med., vol. 50, no. 2, pp. 63–73, Oct. 2010.

[11] Y. Peng, C.-H. Wei, and Z. Lu, "Improving chemical disease relation extraction with rich features and weakly labeled data," J. Cheminform., vol. 8, no. 1, p. 53, Dec. 2016.

[12] C. S. G. Kho, S. Chan, and Y. Niu, "Extracting causal knowledge from a medical database using graphical patterns," presented at the 38th Annu. Meet. Assoc. Comput. Linguistics, Oct. 2002.

[13] L. Wang, G. Del Fiol, B. E. Bray, and P. J. Haug, "Generating disease-pertinent treatment vocabularies from MEDLINE citations," J. Biomed. Inform., vol. 65, pp. 46–57, Jan. 2017.

[14] R. C. Bunescu and R. J. Mooney, "Subsequence kernels for relation extraction," in Proc. Adv. Neural Inf. Process. Syst., Jan. 2005, pp. 171–178.

[15] M. Miwa, R. Satre, Y. Miyao, and J. Tsujii, "Protein–protein interaction extraction by leveraging multiple kernels and parsers," Int. J. Med. Inform., vol. 78, no. 12, pp. e39–e46, Dec. 2009.

[16] M. Song, W. C. Kim, D. Lee, G. E. Heo, and K. Y. Kang, "PKDE4l: Entity and relation extraction for public knowledge discovery," J. Biomed. Inf., vol. 57, pp. 320–332, Oct. 2015.

[17] O. Frunza and D. Inkpen, "Extraction of disease-treatment semantic relations from biomedical sentences," in Proc. Workshop Biomed. Natural Lang. Process. Assoc. Comput. Linguistics, Jul. 2010, pp. 91–98.

[18] R. Socher, B. Huvall, C. D. Manning, and A. Y. Ng, "Semantic compositionality through recursive matrix-vector spaces," in Proc. Joint Conf. Empirical Methods Natural Lang. Process. Comput. Natural Lang. Learn., Jul. 2012, pp. 1201–1207.

[19] D. Zeng, K. Liu, Y. Chen, and J. Zhao, "Distant supervision for relation extraction via piecewise convolutional neural networks," in Proc. Conf. Empirical Methods Natural Lang. Process., Feb. 2015, pp. 1753–1762.

[20] C. dos Santos, B. Xiang, and B. Zhou, "Classifying relations by ranking with convolutional neural networks," in Proc. 53rd Annu. Meet. Assoc. Comput. Linguistics 7th Int. Joint Conf. Natural Lang. Process., May 2015, pp. 626–634.

[21] P. Zhou, W. Shi, J. Tian, Z. Qi, B. Li, H. Hao, and B. Xu, "Attention-based bidirectional long short-term memory networks for relation classification," in Proc. 54th Annu. Meet. Assoc. Comput. Linguistics, Aug. 2016, pp. 207–212.

[22] L. Li, J. Wan, J. Zheng, and J. Wang, "Biomedical event extraction based on GRU integrating attention mechanism," BMC Bioinf., vol. 19, no. 59, pp. 177–184, Aug. 2018.

[23] X. Luo, W. Zhou, W. Wang, Y. Zhu, and J. Deng, "Attention-based relation extraction with bidirectional gated recurrent unit and highway network in the analysis of geological data," IEEE Access, vol. 6, pp. 5705–5715, 2018.

[24] H. Hao, K. Zhang, W. Wang, and G. Gao, "A tale of two countries: International comparison of online doctor reviews between China and the United States," Int. J. Med. Inform., vol. 99, pp. 37–44, Mar. 2017.

[25] J. M. Goh, G. Gao, and R. Agarwal, "The creation of social value: Can an online health community reduce rural-urban health disparities?" MIS Quart., vol. 40, no. 1, pp. 247–263, Jan. 2016.

[26] A. Cohan, S. Young, A. Yates, and N. Goharian, "Triaging content severity in online mental health forums," J. Assoc. Inf. Sci. Technol., vol. 68, no. 11, pp. 2675–2689, Nov. 2017.

[27] P. Martinez, J. L. Martinez, I. Segura-Bedmar, J. Moreno-Schneider, A. Luna, and R. Revert, "Turning user generated health-related content into actionable knowledge through text analytics services," Comput. Ind., vol. 78, pp. 43–56, May 2016.

[28] T. Efimov, B. K. Seljak, and P. Korosce, "A rule-based named-entity recognition method for knowledge extraction of evidence-based dietary recommendations," PLoS ONE, vol. 12, no. 6, Jun. 2017, Art. no. e0179488.

[29] D. Zeng, K. Liu, S. Lai, G. Zhou, and J. Zhao, "Relation classification via convolutional deep neural network," in Proc. the 25th Int. Conf. Comput. Linguistics, Tech. Papers (COLING), Aug. 2014, pp. 2335–2344.

[30] K. Cho, B. van Merrienboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder-decoder approaches," 2014, arXiv:1409.1259 [Online]. Available: http://arxiv.org/abs/1409.1259

[31] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," IEEE Trans. Neural Netw., vol. 5, no. 2, pp. 157–166, Mar. 1994.