IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 26 January 2022, accepted 10 February 2022, date of publication 15 February 2022, date of current version 22 February 2023.

Digital Object Identifier 10.1109/ACCESS.2022.3151900

Adverse Drug Reaction Detection From Social Media Based on Quantum Bi-LSTM With Attention

XUQI WANG, XIANFENG WANG, AND SHANWEN ZHANG
School of Information Engineering, Xijing University, Xi'an, Shaanxi 710123, China

Corresponding author: Shanwen Zhang (zhanshanwen@xijing.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 62172338 and Grant 62072378, and in part by the Xijing University High-Level Talent Special Fund Project under Grant XJ21B14.

ABSTRACT Drug combination is very common in the course of disease treatment. However, it inevitably increases the overall risk of adverse drug reactions (ADRs). It is very important to early and accurately detect and identify the potential ADRs for combined medication safety and public health. Social media is an important pharmacovigilance data source for ADR detection. But the data are complex, mass, clutter, highly sparse, so it is difficult to detect the ADR information from these data. Deep learning stands out in terms of increased accuracy. However, it takes a lot of training time and requires a lot of computing power. Quantum computing has strong parallel computing capability, and requires less computing power. By introducing attention mechanism and quantum computing into Bi-directional Long Short-Term Memory (Bi-LSTM), a quantum Bi-LSTM with attention (QBi-LSTMA) model is constructed for ADR detection from social media big data. QBi-LSTMA is composed of 6 variable component subcircuits (VQC) stacked. Under the condition that the main topology of Bi-LSTM remains unchanged, the biases of QBi-LSTMA in input gate, forgetting gate, candidate memory unit and output gate are removed to simplify the network structure, and the weight and active value qubits of the model are used to update the network weight. The performance of the proposed method is evaluated on the SMM4H dataset, comparing with one traditional ADR detection method and three deep learning based ADR detection approaches. The experiment results show that the proposed method has great potential in ADR detection.

INDEX TERMS Social media big data, adverse drug reactions (ADRs), bi-directional long short-term memory (Bi-LSTM), quantum Bi-LSTM with attention (QBi-LSTMA).

I. INTRODUCTION

Drug combination is very common in medication and clinical practice [1]. However, it often causes various unexpected ADRs, and the more drugs are combined, the more likely they are to interact with each other in terms of pharmacological or physicochemical properties and thus the greater the possibility of ADRs [2], [3]. ADRs are one of the great problems facing the medical field. The occurrence of ADRs greatly increases the length of hospital stay, economic burden and mortality of patients. Most ADRs are mild to moderate and can be controlled with adequate supervision and monitoring, but few serious ADRs may result in deterioration, shock and even death [4]–[6]. ADRs have become the fourth cause of death in the United States and in similar countries, after heart disease, diabetes and AIDS [7]–[9]. Margraff and Bertram [9] found that direct patient reporting systems exist in 44 countries and represent 9% of total reports, the rest coming from healthcare professionals. Dorji *et al.* [10] investigated the knowledge level of both ADRs and ADR reporting among healthcare professionals (HCPs). The survey consisted of 12 questions pertaining to ADRs and 10 questions pertaining to knowledge of ADR reporting. The results show that the clinical doctors and pharmacists have better knowledge of ADRs than nurses and traditional medicine practitioners, while knowledge of ADR reporting is also low for all HCPs surveyed.

Since ADRs are a serious health problem and a leading cause of death, it is very important to detect and identify ADRs correctly and timely. But it is hard and challenging

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

to discover many ADRs, because they happen to certain groups of people in certain conditions and they may take a long time to expose. In fact, the direct or indirect causes of ADRs are diverse, ranging from pharmacological, immune and genetic factors to race, age, sex, social factors and factors related to drugs and diseases. The traditional methods of ADR detection by the ADR related databases and the reports of ADR events rely on manual case and ADR event reviews by clinical/pharmacological experts [11].

Recently the text unstructured data such as social media networks are abundant and generated rapidly, and have been used to mine ADRs [11], [12]. Motivated by limitations of ADR detection in clinical trials and passive post-market drug safety surveillance systems, a number of methods have been presented for potential ADR detection from social media data [12], [13]. Ho *et al.* [14] collected a large number of literatures about ADR in the past 20 years, summarized the existing ADR detection methods and established three tables to provide brief information on the research for ADR detection and prediction. They pointed out that the data-driven approach is powerful in ADR detection and prediction. Liu *et al* [15] presented a feature-based ADR extraction method by utilizing various lexical, syntactic, and semantic features, and compared with four well-known kernel-based approaches (i.e., subset tree kernel, tree kernel, shortest dependency path kernel, and all-paths graph kernel), and tested these methods on three data sets: two health-related discussion forums and one general social media site (i.e., Twitter). Azadeh *et al.* [16] reviewed ADR detection methods in social media and their application in pharmacovigilance, and classified the existing research results according to ADR detection methods, data sources, corpus sizes, and availability and evaluation criteria.

Deep learning has gained considerable attention and achieved great success in big data classification and identification, including Person Re-identification [17] and computer vision objection detection [18] and ADR detection [19]. Masino *et al* [20] framed the ADR detection problem as a binary classification task, and developed a convolutional neural network (CNN) model for tweet ADR classification. The results show the feasibility of detection of infrequent ADR mentions in large-scale media data. The approach reduces manual data-labeling requirements and is scalable to large social media datasets. Tang *et al* [21] proposed a LSTM-CRF based ADR recognition method by combining LSTM and conditional random fields (CRFs) from social media. The results on a benchmark corpus show that LSTM-CRF achieves better F-score than CRF. Anne *et al* [22] developed a recurrent neural network (RNN) model for labeling ADRs in Twitter posts. The only input features are word-embedding vectors, which are formed through task-independent pretraining or during ADR detection training. The results show that the ADR detection performance in social media is significantly improved by using a contextually aware model and word embedding formed from large and unlabeled datasets. Fan *et al* [23] proposed an adverse drug event detection and extraction based deep learning by utilizing Bidirectional Encoder Representations from Transformers (BERT) and compared with the standard deep learning models and current state-of-the-art extraction models. The proposed model can be applied to the medical entity extraction and entity recognition.

Although deep learning based methods outperform the traditional approaches, the general deep learning models require a sharp increase in computational power as the amount of data increases, and the training process of many deep learning models based on the existing computational power is very long, and even cannot be applied in practice. By now, quantum machine learning models experience interesting results compared to the traditional machine learning models, and have shown promising results in natural language processing tasks [24]. Quantum process neural network takes advantage of the advantages of quantum computing, especially the parallel computing characteristic of quantum computing, and has stronger parallel computing capability and larger data processing capability than classical neural network. Jia *et al* [25] briefly reviewed the classical neural networks and many crucial aspects of quantum neural network states, and illustrated how to use neural networks to represent quantum states and density operators. Cong *et al* [26] introduced and analyzed a quantum CNN (QCNN) model motivated by CNN, explicitly illustrated its potential with two examples, and discussed the potential experimental realization and generalizations of QCNN. Shekhar *et al* [22] demonstrated that quantum LSTM (QLSTM) model is capable of learning and accurately predicting the languages used in social media texts. This work paves the way for future applications of deep learning methods in quantum dynamics without relying on the explicit form of the Hamiltonian. Overall, the existing ADR detection approaches are limited with shallow models and heavily engineered features. There is a lack of an end-to-end network model that relies on redundancy of unannotated and annotated data.

Aiming at the hard-problem to predict ADRs from the complex, sparse social media data, inspired by QCNN and QLSTM, making use of the advantages of Bi-LSTM and quantum computing in natural language processing, a quantum Bi-LSTM with attention (QBi-LSTMA) model is constructed for ADR detection based on Bi-LSTM, attention mechanism and quantum computing. It is an attempt to introduce quantum computing into the deep learning to greatly accelerate model training and reduce the demands on computational power.

Compared with most of the deep learning based ADR detection approaches, the proposed method does not heavily rely on the quality of input instance representation and does not require any linguistic knowledge and high computational power. The main contributions are as follows:

- A QBi-LSTMA model is constructed by introducing attention mechanism and quantum computing into Bi-LSTM.

- QBi-LSTMA can extend the performance on Bi-LSTM based ADR extraction method.
- A lot of extensive experiments are conducted to show the reasonable performance of QBi-LSTMA.

The rest of paper is arranged as follows. Section 2 briefly introduces the related works, including variational quantum circuits (VQC) and quantum Long Short-Term Memory (QLSTM). The ADR detection method based QBi-LSTMA is described in detail in Section 3. The experiments and results are presented in Section 4. Section 5 summarized the paper and points out the future works.

II. RELATED WORKS

A. VARIATIONAL QUANTUM CIRCUITS (VQC)

VQC is a kind of quantum circuits that has tunable parameters subject to iterative optimizations. It is a hybrid quantum-classical approach which leverages the strengths of quantum and classical computation, but it is more expressive than classical neural networks with a limited number of parameters, which can be optimized in an iterative manner by a classical computer. Its general structure is shown in Fig.1,

where $U(x)$ is the quantum routine for encoding the classical input data x into the quantum state of the circuit and is not subject to optimization, and $V(\theta)$ is the variational circuit block with learnable parameter θ that is optimized through gradient method.

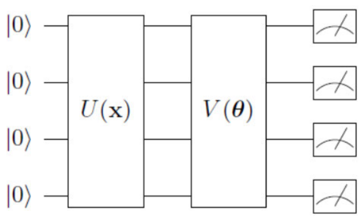


FIGURE 1. The general structure of VQC.

N-qubit state is denoted by

$$|\psi\rangle = \sum_{(q_1,q_2,\cdots,q_N\in\{0,1\})} c_{q_1,q_2,\cdots,q_N} |q_1\rangle \otimes |q_2\rangle \otimes \ldots \otimes |q_N\rangle \tag{1}$$

where $C_{q_1,q_2,\ldots q_N}$ is the complex amplitude of each basis state and each quantum $q_i \in \{0, 1\}$, the square of the amplitude $C_{q_1,q_2,\ldots q_N}$ is the measurement probability of the measured state $|q_1\rangle \otimes |q_2\rangle \ldots \otimes |q_N\rangle$, and the total probability should sum to 1, i.e., $\sum_{(q_1,q_2,\ldots q_N)\in\{0,1\}} \left\| C_{q_1,q_2,\ldots,q_N} \right\|^2 = 1$.

To convert the initial state, $|0\rangle \otimes \cdots \otimes |0\rangle$ to an unbiased state,

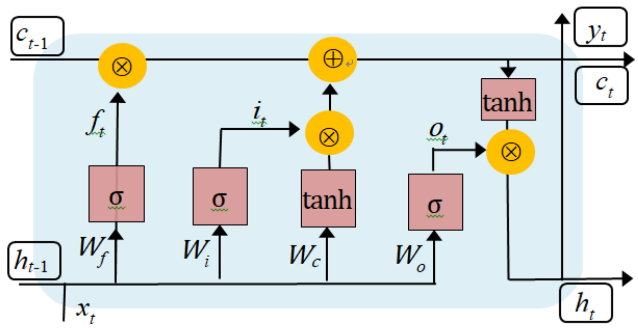
$$\begin{aligned}(H|0\rangle)^{\otimes N} &= \frac{1}{\sqrt{2^N}}(|0\rangle \otimes \cdots \otimes |0\rangle + \cdots + |1\rangle \otimes \cdots \otimes |1\rangle) \\ &\equiv \frac{1}{\sqrt{2^N}} \sum_{i=0}^{2^N-1} |i\rangle \end{aligned} \tag{2}$$

where i is the decimal number marking the corresponding bit string.

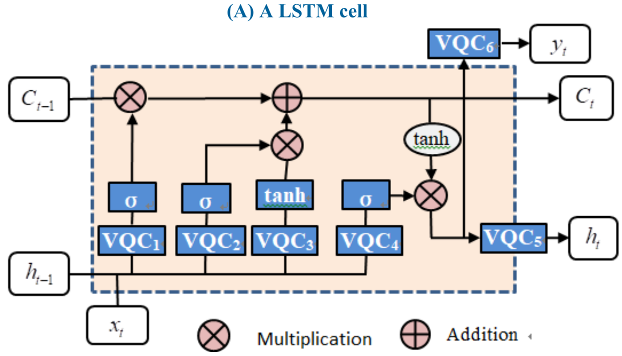
B. QUANTUM LONG SHORT-TERM MEMORY(QLSTM)

LSTM introduces an input gate, a forgetting gate and an output gate. The input gate controls determines how much of the network input is saved to the cell state at the current moment. The forgetting gate determines how much of the cell state at the previous time is retained to the current time. Output gate controls how much the unit state outputs to the current output value of LSTM, thus enabling semantic long-term and short-term memory for longer sequences. In LSTM and Bi-LSTM, there are a lot of parameters that need to be trained. The architecture of a LSTM cell is shown in Fig. 2A.

Similar to LSTM, quantum LSTM (QLSTM) is constructed by introducing quantum computing into the LSTM. The architecture of a QLSTM cell is shown in Fig.2B, consisting of six VQCs stacked, where σ and tanh are the sigmoid and the hyperbolic tangent nonlinear activation functions, respectively, x_t is the input at time t, ht is for the hidden state, c_t is for the cell state, and y_t is the output, $\otimes$ and $\oplus$ are element-wise multiplication and addition, respectively.



(A) A LSTM cell



(B) A QLSTM cell

FIGURE 2. The architecture of LSTM and QLSTM.

In QLSTM, VQC1 controls v_t and outputs a vector f_t with values in the interval [0; 1] through σ, VQC2 deals with vt and then passes through σ so as to determine which values will be added to the cell state, VQC3 deals with the same concatenated input and goes through tanh to generate a new cell state candidate $\tilde{C}_t$, the result from VQC2 is multiplied

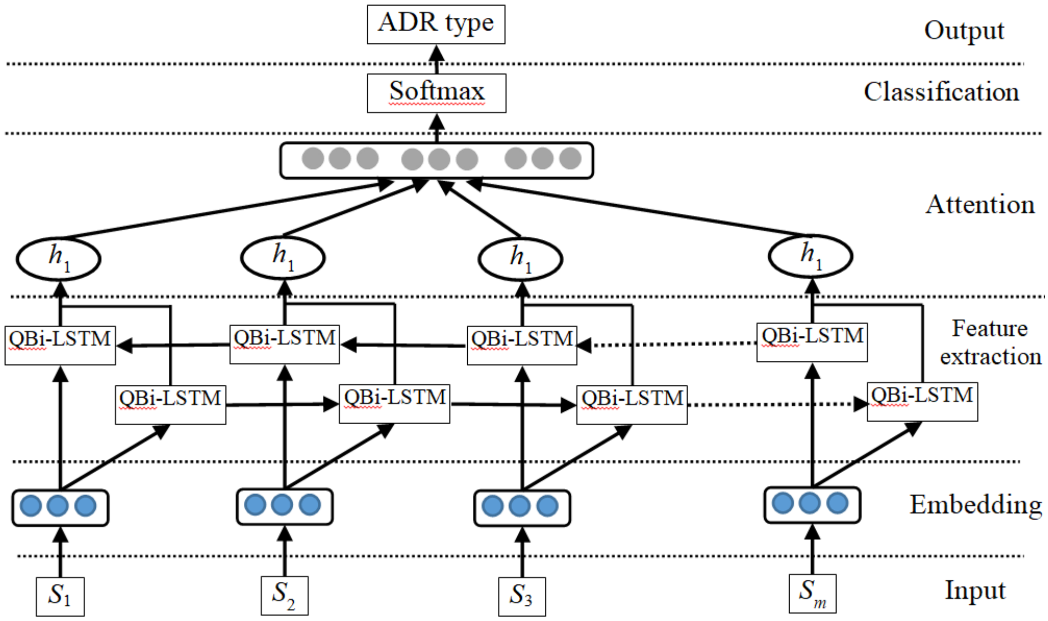


FIGURE 3. The flowchart of the proposed method.

element wisely by $\tilde{C}_t$, and the resulting vector is then used to update the cell state, VQC4 processes v_t and goes through σ to determine which values in the cell state ct are relevant to the output. The cell state itself goes through tanh and then is multiplied element-wisely by the result from VQC4. Finally this value of the above block is further processed with VQC5 to get the hidden state h_t or VQC6 to get the output y_t. That is to say, VQC5 is used to transform c_t to h_t, and likewise VQC6 to transform c_t to y_t. For VQC1 to VQC4, the input is the concatenation v_t of the hidden state h_{t-1} from the previous time step and the current input vector xt, and the output is four vectors obtained from the measurements at the end of each VQCs. The measured values are Pauli Z expectation values of each qubit by design, then go through σ and tanh. The feature extraction process are given as follows,

$$\begin{aligned} f_t &= \sigma\left(VQC_1\left(v_t\right)\right); i_t = \sigma\left(VQC_2\left(v_t\right)\right) \\ \tilde{C}_t &= \tanh\left(VQC_3\left(v_t\right)\right) \\ c_t &= f_t * c_{t-1} + i_t * \tilde{C}_t, o_t = \sigma\left(VQC_4\left(v_t\right)\right) \\ h_t &= VQC_5\left(o_t * \tanh\left(c_t\right)\right) \\ y_t &= VQC_6\left(o_t * \tanh\left(c_t\right)\right) \end{aligned} \tag{3}$$

Different from LSTM, in QLSTM, the iterative transfer relationship in the network is realized through four parts: weighting, activation, aggregation and excitation. It extends the classical LSTM into the quantum realm by replacing the classical neural networks in the LSTM cells with VQCs. To simplify the network structure, under the condition that the main topology of LSTM remains unchanged, the biases in input gate, forgetting gate, candidate memory unit and output gate are removed, and the weight and active value qubits of the network are updated to update the network weight. In certain cases, QLSTM requires fewer parameters than normal neural networks, making them promising for modeling complex environments, and would play the roles of both feature extraction and data compression.

III. QUANTUM BI-LSTM WITH ATTENTION (QBI-LSTMA) FOR ADR DETECTION

The ADR detection task can be treated as a binary classification problem and classify an entity pair as related or not. Aiming at the difficult problem of ADR detection, QBi-LSTMA is constructed by combining Bi-LSTM, attention mechanism and quantum computing, and then an ADR detection method is proposed based on QBi-LSTMA, consisting of input and output, embedding, feature extraction by QBi-LSTM, attention mechanism and ADR classification by Softmax classifier. Its flowchart is shown in Fig.3. The model assumes that a potentially related entity pair of drugs can be supported by the relations between co-existing pairs in the same sentence. From Fig.3, the steps to implement QBI-LSTMA can be divided into five parts, introduced in detail as follows.

A. INPUT

The original text collected from Twitter contains lots of meaningless words or characters, such as URLs, some non-alphabetic characters like '∗@,?&/)(', which may decrease the detection accuracy of ADR. We break it into fine-grained tokens. After screening the noise, there are still some words not related to drugs or ADRs in the data, such as prepositions and verbs. This grammar form human-readable texts, but previous studies have shown that bypassing these words can reduce the solution space and improve performance of the model. Consequently, the dictionary of stop words is compiled to filter these words.

Input the sentence set $\{S_1, S_2, \ldots, S_m\}$ into the model. Build a crawler platform based on Scrapy package, obtain corresponding posts according to the ID number of user posts in Twitter, and store the crawled data in a text file uniformly.

B. EMBEDDING

Typically, classification and clustering algorithms require text input to be represented as a fixed-length vector. The common models that meet this requirement are bag-of-words and bag-of-n-grams. To better represent the words from social media, beyond the contextual information of words, the morphological and shape information are take into consideration. We map each word into a low-dimensional dimension vector, including word segmentation and word embedding. Each word obtained and preprocessed by the input layer is mapped into a word vector, entity type vector and part of speech (POS) vector, and then are concatenated as a vector. Word segmentation is to divide the input sentence by word unit, and word embedding is a process of transforming words into vectors. The word segmentation operation is carried out, and then the vector E of each word in the sentence is obtained by combining the word sequence dictionary and One-Hot coding, whose length is the hyperparametric word embedded dimension d. Then, the whole sentence-word embedding matrix $[E_1, E_2, \ldots, E_n]$ is obtained, where N is the maximum length of all sentences in the set S after word segmentation. Finally, the same operation is carried out for each sentence in the set to obtain a three-dimensional word-embedding matrix with the size of $m \times n \times d$.

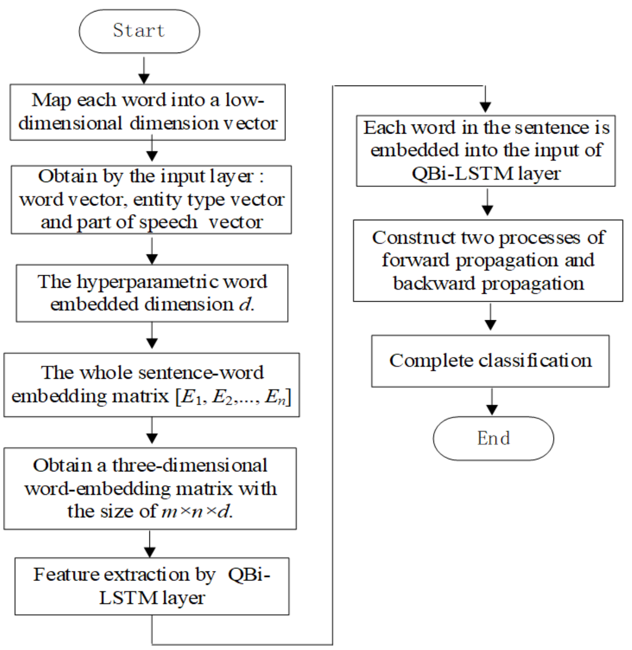


FIGURE 4. The detail flow diagram for feature extraction.

C. FEATURE EXTRACTION BY QBi-LSTM

Use QBi-LSTM to obtain advanced features from Section B. For a sentence text data EW, each word in the sentence is embedded into the input of QBi-LSTM layer for feature extraction. QBi-LSTM can make full use of the information of the whole text sequence and thus enhance the memory capability of Bi-LSTM, including the information of the relationship between each word, and use this information for the processing of each word. It constructs two processes of forward propagation and backward propagation. The complete classification is obtained by fusing the feature forward propagation feature and the backward propagation feature. The detail flow diagram for feature extraction is shown in FIGURE 4.

D. ATTENTION MECHANISM

Generate a weight vector and multiply word-level function and weight vector of each time step to merge them into a sentence-level feature vector. The attention carries out weighted transformation on the state information sequence extracted by QBi-LSTM to highlight the contribution of important state information and effectively improve the accuracy of the ADR detection. The process of attention is described as follows,

$$\begin{aligned} M &= \tanh(H), \quad \alpha = soft\max\left(w^T M\right) \\ r &= H\alpha^T, \quad H^* = \tanh(r) \end{aligned} \tag{4}$$

where H is the feature set of all words in a sentence extracted by QBi-LSTM, w is the parameter vector of training learning.

E. CLASSIFICATION AND OUTPUT

The output of the attention are concatenated and input into Softmax classifier to calculate the ADR probability $(y = c)$ of the candidate drugs as follows,

$$\begin{aligned} P(y = c) &= soft\max(w \cdot \gamma + b) \\ \bar{y} &= \arg\max_{y \in c} P(y = c) \end{aligned} \tag{5}$$

where W and b are weight matrix and bias, C is the set ADR type label.

The probability biggest category label $\bar{y}$ is the ADR type of the candidate drug. Finally, there are a lot of parameters in QBi-LSTM which need to be trained during the model training. The gradient descent based method is adopted to learn the model parameters. In each training time, for L input samples $\langle x_i, y_i \rangle$, the gradient (using the chain rules) of each parameter relative to loss is calculated and then updated each parameter with learning rate λ:

$$Loss = \sum_{i=1}^{L} -\log p(x_i | y_i), \theta = \theta - \lambda \frac{\partial Loss}{\theta} \tag{6}$$

where θ is the super parameter.

It is notable that fixed learning rate λ would lead to unstable loss in training.

TABLE 1. Statistical Information of two Corpuses.

Corpus set	Samples at publication	Available samples	Positive	Negative
TwiMed	1000	608	234	374
TwitterADR	10 822	6 966	765	6 201

IV. EXPERIMENTS AND RESULTS

To verify the effectiveness of the QBi-LSTM based ADR detection method, a number of comparative experiments are conducted using social medical data, and compared the experimental results with the existing ADR methods, i.e., linear neighborhood similarity (LNS) [32], CNN and Word Embedding Features (CNNWEF) [20], recurrent neural network (RNN) [22], and attention-based recurrent neural networks (ATT-RNN) [28], where LNS is a traditional method, CNNWEF is CNN, and ATT-RNN is to use attention mechanism to incorporate RNN.

A. DATASET DESCRIPTION

Twitter is a particularly attractive platform because it has a large, diverse user community. The challenges faced in applying Twitter data to ADR detection are the ADR related data are highly sparse relative to the overall number of user posts and human review of all posts is impractical. In the following experiments, two corpus sets TwiMed [29] and TwitterADR [30], [31] are used to validate the proposed method, which were all tagged from Twitter with the corresponding ID and category tag for users. The statistical information of two corpuses is shown in Table 1. The proportion of the available samples on corpus set is about 60%. The TwiMed corpus takes into account the proportion of positive and negative samples in labeling, which is relatively more balanced, about 1:1.6. The TwitterADR corpus size is larger than that of TwiMed, and the ratio of positive and negative samples is about 1:8.1.

B. EXPERIMENTAL SETTING

The experiments are conducted on 32G memory, with Intel Core i5-4200U CPU @2.30 GHz, GPU GEFORCE GTX 1080ti, Ubuntu14.0. The deep learning architecture is Tensorflow1.7.0 and Keras, including LSTM. During the training, the learning rate of the model is initialized as 0.001, the attenuation rate is 0.1, the training cycle is set as 300, the batch size is set as 10, the dropout coefficient is set as 0.5, and other parameters are randomly initialized.

C. SYSTEM CONFIGURATION

Ten-fold-cross validation criterion is used to carry out the experiments. For each fold, all data are split into the training and test sets, and the training dataset is split into training and validating sets, and the ratio of training set to verification set is 9:1. The word embedding dimension used in the experiment is 100, the embedding dimension is 50. The cross entropy loss is used as the objective function of classification problem. Stochastic gradient descent (SGD) algorithm is used to optimize the training the model. The training stops when there is no performance improvement on the validation set after 5 consecutive epochs.

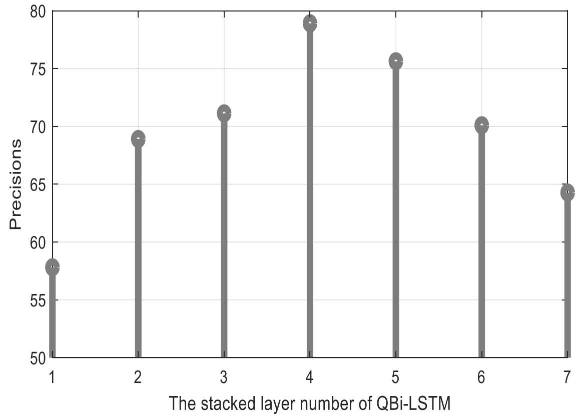


FIGURE 5. Precision versus the stacked layer number of QBi-LSTM.

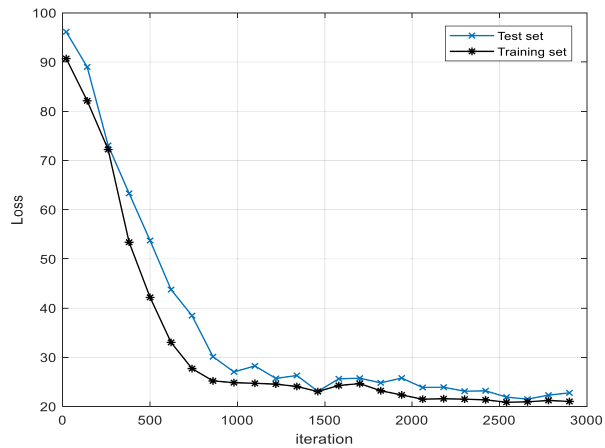


FIGURE 6. Losses with the different number of iterations.

D. EVALUATION METRICS

Precision (P), Recall (R) rate and F1-score (F1) of the positive class (instances labeled as containing the description of ADRs) are used as evaluation metrics to test the performance of QBi-LSTMA in the experiment. F-score is defined as $F1 = (2PR)/(P+)$, where F1 can play a balancing role between P and R.

E. RESULTS

First, we evaluate the impact of the number of the stacked layers of QBi-LSTM on TwiMed dataset. FIGURE 5 show the Precisions versus the number of the stacked layers of QBi-LSTM. From FIGURE 4, it is found the optimal number is 4, more stacked layers usually mean more learnable features, but further increasing its number does not improve the performance, and more than 4 layers may cause the model difficult to train. In the following experiments, the stacked layer number of QBi-LSTM is set as 4.

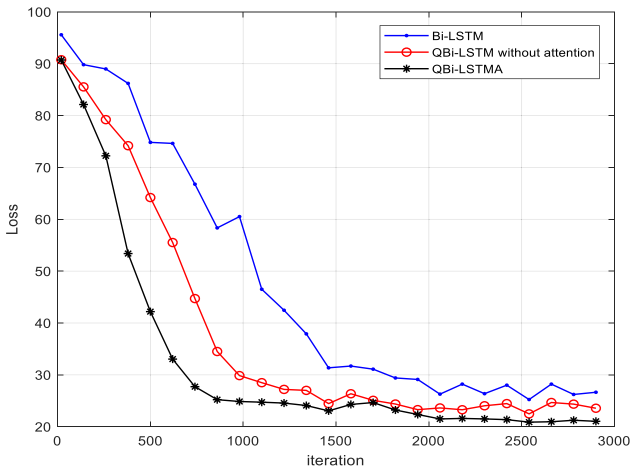


FIGURE 7. Losses with the different number of iterations by QBi-LSTMA, Bi-LSTM and QBi-LSTM without attention.

To test the converge performance with the different iteration times, FIGURE 6 shows the losses on the set training and test set of the TwiMed dataset with the different number of iterations. From FIGURE 6, it is seen that the training and test losses of QBi-LSTMA decrease as the number of iterations processed increased, at the initial stage of learning, two losses are high, and then reduce down very quickly before the 1000^{th} iterations, after 2000 iterations, the losses become slowly stable, and we select the trained QBi-LSTMA at the 2500th iterations. It is also seen that the performance in training set is slightly better than that in test set overall.

Attention mechanism equivalently appends additional restrictions to the model and requires semantic meanings to match strictly. To test the effect of attention mechanism, FIGURE 6 shows the losses by QBi-LSTM with attention (QBi-LSTMA) and QBi-LSTM without attention on the TwiMed dataset with the different number of iterations, comparing with the existing Bi-LSTM. From FIGURE 7, it is obviously seen that the performance of QBi-LSTMA is distinctly superior to that of Bi-LSTM and QBi-LSTM without attention. The reason is that attention mechanism

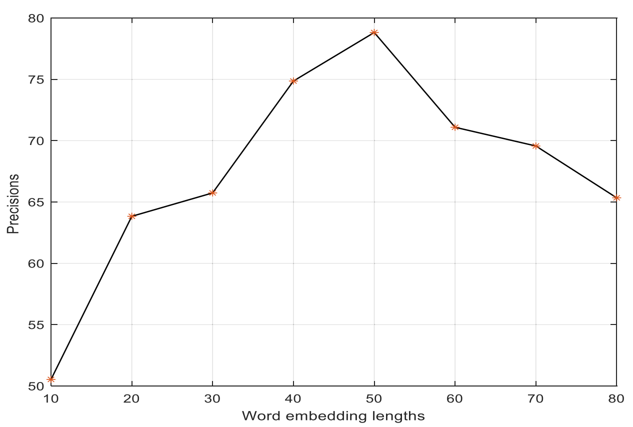


FIGURE 8. Precisions by QBi-LSTMA versus word embedding lengths.

TABLE 2. The ADR detection results on TwiMed by LNS [32], CNNWEF [20], RNN [22], ATT-RNN [28] and QBi-LSTMA.

Method / Results	LNS	CNNWEF	RNN	ATT-RNN	QBi-LSTMA
P	32.37	58.54	62.35	75.10	78.41
R	66.30	61.35	71.24	71.12	69.38
F1	43.50	59.91	66.50	72.65	73.62

embedding lengths

TABLE 3. The ADR detection results on TwitterADR by LNS [32], CNNWEF [20], RNN [22], ATT-RNN [28] and QBi-LSTMA.

Method / Results	LNS	CNNWEF	RNN	ATT-RNN	QBi-LSTMA
P	58.23	63.05	61.21	67.13	66.25
R	54.11	61.37	51.49	65.64	65.71
F1	56.09	62.20	56.01	65.67	65.98
Parameter	--	42.71MB	2.70MB	4.04MB	1.83MB

TABLE 4. The characteristics of Bi-LSTM, Bi-LSTMA and QBi-LSTMA.

Method / Merit & demerit	LNS	CNNWEF	RNN	ATT-RNN	QBi-LSTMA
Accuracy	Lowest	Low	Low	High	Highest
Training time	Least	Most	Medium	Less	Least
Memory	Weakest	Weak	Medium	Strongest	Strong
Feature learning ability	Weakest	Strongest	Medium	Strong	Strong
Generalization	Weakest	Weak	Medium	Strong	Strong

can fully extract character-level, word-level, sentence-level and even inter-sentence relationship features, so that the pre-trained word vector can better represent syntactic and semantic information in different contexts and improve the performance of entity recognition. The performance of QBi-LSTMA and QBi-LSTM without attention is much better than that of Bi-LSTM, because quantum networks have excellent nonlinear approximation ability, ideal generalization performance and fast convergence rate.

To decide the embedding word length, we additionally combine different word embedding lengths and utilize QBi-LSTMA to deal with the challenge that encountered in data representation of multiple sources, and further identify text containing ADR information. FIGURE 8 shows the Precisions versus word embedding lengths. From FIGURE 8, it is known that the word embedding length affects the performance of QBi-LSTMA. The reason is that low-length vectors do not contain enough semantic information, while increasing the embedding length brings much more noise despite their richer semantics. The appropriate dimension is 50.

From FIGURE 5 to FIGURE 8, we obtained a trained QBi-LSTMA, where the stacked layer number of QBi-LSTM is 4, the number of iterations is 2500, and the word embedding length is 50.

The ten-fold-cross validation experiments are repeated 10 times, and their average results are regarded as the ADR detection result of each method, as shown in Table 2 and Table 3. To validate the training performance, the parameter scales of four networks are also given in Table 2.

From the above experiments and result analysis, the merits & demerits of LNS, CNNWEF, RNN, ATT-RNN and QBi-LSTMA are listed in TABLE 4.

Through contrast experimental results in Tables 2 to 4, it can be seen that the QBi-LSTMA based ADR detection method outperforms the other methods. It is also found that deep learning models based methods are is far better than the traditional method LNS, and RNN and ATT-RNN are better than CNNWEF. The reasons are concluded as follows.

LNS is a traditional method, has few parameters to be determined by training, and none memory, so its accuracy is lowest. CNN in CNNWEF has strong feature extraction ability, but poor memory ability. It is well suited for image recognition, but not for text classification, including ADR detection. Compared with LSTM, RNN can only handle short-term dependencies. ATT-RNN is to use attention mechanism to incorporate RNN. RNN is relatively simple with only one parameter matrix and has a very short memory and is prone to the long dependence problem of gradient disappearance, while QBi-LSTMA is more complex with four parameter matrices and can solve this problem by having a longer memory. RNN has no cell state; QBi-LSTMA similar to LSTM, remembers information through cellular states. The biases of QBi-LSTMA in input gate, forgetting gate, candidate memory unit and output gate are removed to simplify the network structure, and the weight and active value qubits of the model are used to update the network weight. CNN in CNNWEF assigns a weight to each word when making classification decision usually needs to copy the feature detector, which will reduce the model building efficiency, and CNN is often insensitive to spatial position, it is difficult to effectively encode information such as position information and semantic information in text sentences. RNN is better than CNN since RNN has long and short time memory ability. QBi-LSTMA is the best, because it integrates the advantages of Bi-LSTM, attention and quantum computing, and the nonlinear approximation ability and generalization performance of the original LSTM and Bi-LSTM are improved. From Table 2, it is found that the parameter scale of QBi-LSTMA is minimal, since six VQCs are used instead of the traditional neural networks.

On the whole, QBi-LSTMA outperforms the other models.

V. CONCLUSION AND FUTURE WORK

Adverse Drug Reactions (ADRs) are potentially dangerous to patients and are amongst the top causes of morbidity and mortality, and are one of main concerns in the drug discovery, which gains wide attentions, however it is unfeasible to investigate all possible ADDIs. It is known that a great number of computational methods have been proposed for ADR detection, but it is still hard to detect and predict all ADRs, because many ADRs hide in the massive social media big data, which contain a lot of ill-grammatical sentences and short forms. A QBi-LSTMA based ADR detection method is proposed to detect the possible ADRIs from social media data. QBi-LSTMA takes the advantages of additional Bi-LSTM, attention and quantum computing and thus enhance the detection performance, training ability and robust. Experimental analysis indicates that the proposed method achieves a good overall performance on the ADR task. Future research is to optimize QBi-LSTMA so as to be applied to the computationally constrained mobile device, and we will consider using a multi-stage classifier to handle each type of data in a targeted manner. The generalization ability and computational complexity of the proposed QBi-LSTMA need to be further studied.

REFERENCES

[1] A. Poleksic and L. Xie, "Database of adverse events associated with drugs and drug combinations," *Sci. Rep.*, vol. 9, no. 1, p. 20025, Dec. 2019, doi: 10.1038/s41598-019-56525-5.

[2] P. Crits-Christoph, M. G. Newman, K. Rickels, R. Gallop, M. B. C. Gibbons, J. L. Hamilton, S. Ring-Kurtz, and A. M. Pastva, "Combined medication and cognitive therapy for generalized anxiety disorder," *J. Anxiety Disorders*, vol. 25, no. 8, pp. 1087–1094, Dec. 2011.

[3] R. Harpaz, H. S. Chase, and C. Friedman, "Mining multi-item drug adverse effect associations in spontaneous reporting systems," *BMC Bioinf.*, vol. 11, no. S9, pp. 1–8, Oct. 2010, doi: 10.1186/1471-2105-11-S9-S7.

[4] H. Yang and C. C. Yang, "Discovering drug-drug interactions and associated adverse drug reactions with triad prediction in heterogeneous healthcare networks," in *Proc. IEEE Int. Conf. Healthcare Informat. (ICHI)*, Oct. 2016, pp. 244–254.

[5] J. Strandell, O. Caster, J. Hopstadius, I. R. Edwards, and G. N. Norén, "The development and evaluation of triage algorithms for early discovery of adverse drug interactions," *Drug Saf.*, vol. 36, no. 5, pp. 371–388, May 2013.

[6] M. Liu, Y. Hu, and B. Tang, "Role of text mining in early identification of potential drug safety issues," in *Biomedical Literature Mining*. New York, NY, USA: Humana Press, 2014, pp. 227–251, doi: 10.1007/978-1-4939-0709-0_13.

[7] S. V. Iyer, R. Harpaz, P. LePendu, A. Bauer-Mehren, and N. H. Shah, "Mining clinical text for signals of adverse drug-drug interactions," *J. Amer. Med. Inform. Assoc.*, vol. 21, no. 2, pp. 353–362, Mar. 2014.

[8] D. A. Nguyen, C. H. Nguyen, and H. Mamitsuka, "A survey on adverse drug reaction studies: Data, tasks and machine learning methods," *Briefings Bioinf.*, vol. 22, no. 1, pp. 164–177, Jan. 2021, doi: 10.1093/bib/bbz140.

[9] F. Margraff and D. Bertram, "Adverse drug reaction reporting by patients: An overview of fifty countries," *Drug Saf.*, vol. 37, no. 6, pp. 409–419, Jun. 2014.

[10] C. Dorji, P. Tragulpiankit, A. Riewpaiboon, and T. Tobgay, "Knowledge of adverse drug reaction reporting among healthcare professionals in Bhutan: A cross-sectional survey," *Drug Saf.*, vol. 39, no. 12, pp. 1239–1250, Dec. 2016.

[11] Y. Zhou, V. Miller, M. Hogan, and L. Callahan, "An overview of adverse drug reaction monitoring in China," *Int. J. Pharmaceutical Med.*, vol. 20, no. 2, pp. 79–85, 2006.

[12] A. Sarker, R. Ginn, A. Nikfarjam, K. O'Connor, K. Smith, S. Jayaraman, T. Upadhaya, and G. Gonzalez, "Utilizing social media data for pharmacovigilance: A review," *J. Biomed. Informat.*, vol. 54, pp. 202–212, Apr. 2015.

[13] I. Korkontzelos, A. Nikfarjam, M. Shardlow, A. Sarker, S. Ananiadou, and G. H. Gonzalez, "Analysis of the effect of sentiment analysis on extracting adverse drug reactions from tweets and forum posts," *J. Biomed. Informat.*, vol. 62, pp. 148–158, Aug. 2016.

[14] T.-B. Ho, L. Le, D. T. Thai, and S. Taewijit, "Data-driven approach to detect and predict adverse drug reactions," *Current Pharmaceutical Des.*, vol. 22, no. 23, pp. 3498–3526, Jun. 2016.

[15] J. Liu, S. Zhao, and X. Zhang, "An ensemble method for extracting adverse drug events from social media," *Artif. Intell. Med.*, vol. 70, pp. 62–76, Jun. 2016.

[16] A. Nikfarjam, A. Sarker, K. O'Connor, R. Ginn, and G. Gonzalez, "Pharmacovigilance from social media: Mining adverse drug reaction mentions using sequence labeling with word embedding cluster features," *J. Amer. Med. Inf. Assoc.*, vol. 22, no. 3, pp. 671–681, May 2015.

[17] S. U. Khan, T. Hussain, A. Ullah, and S. W. Baik, "Deep-ReID: Deep features and autoencoder assisted image patching strategy for person re-identification in smart cities surveillance," *Multimedia Tools Appl.*, pp. 1–22, Jan. 2021. [Online]. Available: https://link.springer.com/article/10.1007/s11042-020-10145-8#citeas, doi: 10.1007/s11042-020-10145-8.

[18] S. U. Khan, I. U. Haq, S. Rho, S. W. Baik, and M. Y. Lee, "Cover the violence: A novel deep-learning-based approach towards violence-detection in movies," *Appl. Sci.*, vol. 9, no. 22, p. 4963, Nov. 2019.

[19] E. Florez, F. Precioso, R. Pighetti, and M. Riveill, "Deep learning for identification of adverse drug reaction relations," in *Proc. Int. Symp. Signal Process. Syst.*, 2019, pp. 149–153.

[20] A. J. Masino, D. Forsyth, and A. G. Fiks, "Detecting adverse drug reactions on Twitter with convolutional neural networks and word embedding features," *J. Healthcare Inform. Res.*, vol. 2, nos. 1–2, pp. 25–43, Jun. 2018.

[21] B. Tang, J. Hu, X. Wang, and Q. Chen, "Recognizing continuous and discontinuous adverse drug reaction mentions from social media using LSTM-CRF," *Wireless Commun. Mobile Comput.*, vol. 2018, pp. 1–8, Apr. 2018.

[22] A. Cocos, A. G. Fiks, and A. J. Masino, "Deep learning for pharmacovigilance: Recurrent neural network architectures for labeling adverse drug reactions in Twitter posts," *J. Amer. Med. Inform. Assoc.*, vol. 24, no. 4, pp. 813–821, Jul. 2017.

[23] B. Fan, W. Fan, and C. Smith, "Adverse drug event detection and extraction from open data: A deep learning approach," *Inf. Process. Manage.*, vol. 57, no. 1, Jan. 2020, Art. no. 102131, doi: 10.1016/j.ipm.2019.102131.

[24] A. Sordoni, J.-Y. Nie, and Y. Bengio, "Modeling term dependencies with quantum language models for IR," in *Proc. 36th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Dublin, Ireland, Jul. 2013, pp. 653–662.

[25] Z. A. Jia, B. Yi, R. Zhai, Y.-C. Wu, G.-C. Guo, and G.-P. Guo, "Quantum neural network states: A brief review of methods and applications," *Adv. Quantum Technol.*, vol. 2, nos. 7–8, 2019, Art. no. 1800077, doi: 10.1002/qute.201800077.

[26] I. Cong, S. Choi, and D. L. Mikhail, "Quantum convolutional neural networks," *Nature Phys.*, vol. 15, no. 12, pp. 1273–1278, 2019.

[27] S. Shekhar, D. K. Sharma, and M. M. S. Beg, "Language identification framework in code-mixed social media text based on quantum LSTM—The word belongs to which language?" *Modern Phys. Lett. B*, vol. 34, no. 6, Feb. 2020, Art. no. 2050086, doi: 10.1142/S0217984920500864.

[28] C. Du and L. Huang, "Text classification research with attention-based recurrent neural networks," *Int. J. Comput. Commun. Control*, vol. 13, no. 1, p. 50, Feb. 2018.

[29] N. Alvaro, Y. Miyao, and N. Collier, "TwiMed: Twitter and PubMed comparable corpus of drugs, diseases, symptoms, and their relations," *JMIR Public Health Surveill.*, vol. 3, no. 2, p. e24, May 2017.

[30] A. Sarker and G. Gonzalez, "Portable automatic text classification for adverse drug reaction detection via multi-corpus training," *J. Biomed. Informat.*, vol. 53, pp. 196–207, Feb. 2015.

[31] A. Sarker, A. Nikfarjam, and G. Gonzalez, "Social media mining shared task workshop," in *Proc. Pacific Symp. Biocomputing*, 2016, pp. 581–592, doi: 10.13140/RG.2.1.3886.7924.

[32] W. Zhang, X. Yue, F. Liu, Y. Chen, S. Tu, and X. Zhang, "A unified frame of predicting side effects of drugs by using linear neighborhood similarity," *BMC Syst. Biol.*, vol. 11, no. S6, pp. 23–34, Dec. 2017, doi: 10.1186/s12918-017-0477-2.

XUQI WANG received the B.S. degree in computer science and technology from Xidian University, Shaanxi, China, in 1999, the M.S. degree in software engineering from Xi'an Jiaotong University, Shaanxi, in 2005, and the Ph.D. degree in communication and information system from the School of Mechanical Electronic and Information Engineering, China University of Mining and Technology (Beijing), Beijing, China, in 2020.

He is currently an Associate Professor with the Department of Information and Engineering, Xijing University, Xi'an, China. His research interests include wireless sensor networks and artificial intelligence.

XIANFENG WANG received the Bachelor of Science degree from the Department of Mathematics, Northwest University, in 1988, and the Master of Operations Research degree from the Air Force Missile Academy, in 1994.

From July 1988 to June 2008, he taught with the Missile College, Air Force Engineering University, as an Associate Professor and a Master's Supervisor. He is currently an Associate Professor with the College of Science, Xijing University, Xi'an, China. His research interest includes pattern recognition.

SHANWEN ZHANG was born in Shaanxi, China. He received the B.S. degree in mathematics from Northwest University, China, in 1988, the M.S. degree in applied mathematics from Northwest Polytechnic University, China, in 1995, and the Ph.D. degree in electromagnetic field and microwave from Air Force Engineering University, China, in 2001. He is currently a Professor with the School of Information Engineering, Xijing University, Xi'an, China, and a Visiting Scholar with the Department of Computer Science, Virginia Tech. His research interests include machine learning and its application in data mining, including machine learning, leaf image processing, data reduction, data mining, feature selection, wavelet transforms, and their application in the sonar image recognition.