

Two End-to-End Quantum-Inspired Deep Neural Networks for Text Classification

Jinjing Shi[✉], Zhenhuan Li[✉], Wei Lai, Fangfang Li, Ronghua Shi,
Yanyan Feng[✉], and Shichao Zhang[✉], *Senior Member, IEEE*

Abstract—In linguistics, the uncertainty of context due to polysemy is widespread, which attracts much attention. Quantum-inspired complex word embedding based on Hilbert space plays an important role in natural language processing (NLP), which fully leverages the similarity between quantum states and word tokens. A word containing multiple meanings could correspond to a single quantum particle which may exist in several possible states, and a sentence could be analogous to the quantum system where particles interfere with each other. Motivated by quantum-inspired complex word embedding, interpretable complex-valued word embedding (ICWE) is proposed to design two end-to-end quantum-inspired deep neural networks (ICWE-QNN and CICWE-QNN representing convolutional complex-valued neural network based on ICWE) for binary text classification. They have the proven feasibility and effectiveness in the application of NLP and can solve the problem of text information loss in CE-Mix [1] model caused by neglecting the important linguistic features of text, since linguistic feature extraction is presented in our model with deep learning algorithms, in which gated recurrent unit (GRU) extracts the sequence information of sentences, attention mechanism makes the model focus on important words in sentences and convolutional layer captures the local features of projected matrix. The model ICWE-QNN can avoid random combination of word tokens and CICWE-QNN fully considers textual features of the projected matrix. Experiments conducted on five benchmarking classification datasets demonstrate our proposed models have higher accuracy than the compared traditional models including CaptionRep BOW, DictRep BOW and Paragram-Phrase, and they also have great performance on F1-score. Especially, CICWE-QNN model has higher accuracy than the quantum-inspired model CE-Mix as well for four datasets including SST, SUBJ, CR and MPQA. It is a meaningful and effective exploration to design quantum-inspired deep neural networks to promote the performance of text classification.

Index Terms—Complex-valued word embedding, text classification, deep neural network, deep learning, natural language processing

1 INTRODUCTION

TEXT classification [2], [3], [4] as a basic task in NLP has been researched for a long time, where a large number of superior deep neural networks like AC-BLSTM [5] are applied with remarkable performance on experiments. It benefits from the powerful capabilities of deep learning methods such as recurrent neural network (RNN) [6] and convolution neural network (CNN) [7] for feature extraction. In 2015, Zhang *et al.* showed the application of

character-level convolutional networks [8] on text classification. In 2017, Hughes *et al.* developed deep convolutional neural networks for medical text classification [9]. In 2020, Lan *et al.* proposed SRCLA [10] for filtering different types of linguistic features and obtaining more semantic features. Unfortunately, there is a common problem for lack of sufficient interpretability in above text classification schemes with deep neural networks, which means that it is hard to explain how the “black box” works in classification tasks, and how to enhance the model interpretability is a recognized question worth exploring in recent years [11].

Since mathematical foundations of quantum theory are very similar to those of compositional NLP with applied category theory [12], quantum computers may provide a promising natural setting for compositional NLP tasks, and it is potential to enhance the model interpretability with quantum theory. In linguistics, the polysemy phenomenon that a word may contain multiple meanings brings a challenge to determine the emergent meaning [1] of the words, and the concrete meaning of a word could be captured in a certain context. Polysemy is quite similar to the superposition in quantum mechanics that a quantum particle may possess several states at the same time and interact with other particles in physical space. Thus a word containing distinct meanings in different contexts can be expressed with a single quantum particle existing in several quantum states simultaneously. For example, the word “Spring” shown in Fig. 1 which is analogous to a single particle (denoted as $\alpha|0\rangle + \beta|1\rangle$) firstly means the season corresponding to the state $|0\rangle$, while it can also denote

- Jinjing Shi, Zhenhuan Li, Wei Lai, Fangfang Li, Ronghua Shi, and Shichao Zhang are with the School of Computer Science and Engineering, Central South University, Changsha 410083, China. E-mail: {shijinjing, lifangfang, shirrh}@csu.edu.cn, {743879362, 1091697485}@qq.com, zhangsc@mailbox.gxnu.edu.cn.
- Yanyan Feng is with the School of Computer Science and Engineering, Central South University, Changsha 410083, China, and also with the College of Computer and Information Engineering, Central South University of Forestry and Technology, Changsha 410004, China. E-mail: fengyanyanhenu@163.com.

Manuscript received 28 June 2021; revised 12 November 2021; accepted 21 November 2021. Date of publication 25 November 2021; date of current version 7 March 2023.

This work was supported in part by the National Natural Science Foundation of China under Grants 61972418, 61872390, U1736113, and 61836016, in part by the Natural Science Foundation of Hunan Province under Grant 2020JJ4750, in part by the Special Foundation for Distinguished Young Scientists of Changsha Grant kq1905058, in part by the Changsha Municipal Natural Science Foundation Grant kq2014134, and in part by CCF-Baidu Open Fund under Grant CCF-BAIDUOF2021031.

(Corresponding authors: Yanyan Feng and Shichao Zhang.)

Recommended for acceptance by M. Wang.

Digital Object Identifier no. 10.1109/TKDE.2021.3130598

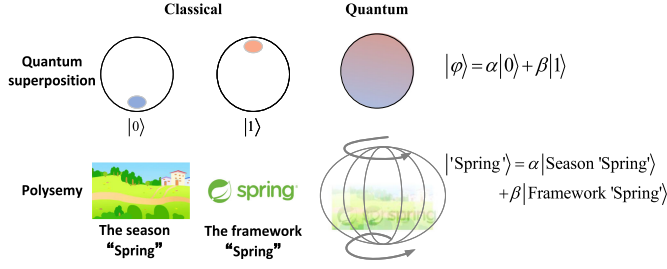


Fig. 1. The similarity between quantum superposition state and polysemy.

Java framework in another specific context corresponding to the state $|1\rangle$. In addition, a phrase, sentence or document composed of words could be analogized to a quantum mixed system containing multiple particles, which can be also represented by the form of a density matrix. In 2018, motivated by the similarity of polysemy and the superposition of quantum states [13], quantum-like and quantum-inspired models like NNQLM [14] and complex neural network [1] were proposed to implement common tasks such as text classification and question-answering [15], [16], which created a precedent for enhancing the model interpretability with quantum physics. Two end-to-end neural networks based on quantum theory named CE-Sup and CE-Mix were proposed in Ref. [1] to imply a combination of NLP with quantum computing, which can obtain superior performance on accuracy in handling text classification task [10], [17] in comparison with several non-quantum models like DictRep BOW [18], since it can improve the model interpretability to a certain extent. However, the sequence information among word-level tokens is ignored in their networks, where the density matrix representation of the sentence-level text is just formed by the linear combination of several matrices for word representation. It may affect the learning ability and accuracy for lack of consideration on the characteristics of human language.

Therefore, we involve deep learning methods including RNN and GRU into the quantum theory-based neural network aiming to obtain the positional information of sentences as textual features, which may enhance the learning ability of the model on the basis of ensuring interpretability. RNN which has excellent performance on dealing with sequential data such as sentence-level text composed of a series of words appearing in order, is applied to capturing sequence information or textual feature. GRU as a variant of standard RNN [19], is employed for solving the long-term dependency problems existing in the case of normal RNN receiving and handling long length text data. Two end-to-end quantum-inspired deep neural networks for text classification are proposed in this paper, and experiments are conducted on five benchmarking for binary text classification. The major work we have achieved can be summarized as follows:

- A novel word embedding method named interpretable complex-valued word embedding (ICWE) is proposed to improve the model interpretability of text classification. We specifically use GRU and self-attention layer [20] to update the amplitude word vectors for extracting more semantic features [21] and position information, where the updated amplitude word vectors and phase vectors together form ICWE.

TABLE 1
Notations

Notation	Description
$ \phi\rangle$	Quantum superposition state
D	Density matrix
M	Projection matrix
M^*	Complex conjugate of matrix M
M^\dagger	Hermitian conjugate of projection matrix M , $M^\dagger = (M^T)^*$
R	Result matrix or projected matrix
I	Identity matrix
x_t	Word vector as the first input of GRU at time t .
h_{t-1}	Input vector contains the text information before time t
h_t	Hidden vector

- On the basis of ICWE, we design and construct the first quantum-inspired deep neural network named ICWE-QNN for binary text classification. Compared with the models completely based on quantum theory such as CE-Sup and CE-Mix [1], it has better performance on accuracy for classification task with injected sequence information.
- The second model called CICWE-QNN we propose is presented for more remarkable performance on classification task by applying convolutional layer [22] on the projected matrix for considering the complete information and capturing local textual features, motivated by the extraction for the joint representation of question-answer pairs [14].

Effectiveness of methods is further illustrated and verified in following parts, and the rest of our paper is organized as follows: Related basic theory about quantum mechanics and deep learning algorithms are introduced in Section 2. We describe the details of complex-valued word embedding and propose two quantum-inspired deep neural networks in Section 3. Experimental results are shown in Section 4. In Section 5, we draw a conclusion on the achieved work and introduce the prospect for future research.

2 RELATED WORK

In this section, we briefly review the related work about NLP based on quantum theory or quantum-like theory. Elemental notations used in our paper are shown in the Table 1 at the beginning of this section.

2.1 Preliminary

In quantum mechanics, a single particle is often represented by a superposition state $|\phi\rangle = \alpha|0\rangle + \beta|1\rangle$, where $|\alpha|^2 + |\beta|^2 = 1$, $|\alpha|^2$ ($|\beta|^2$) denotes the probability for the state $|0\rangle$ ($|1\rangle$) with $0 \leq |\alpha|^2 \leq 1$, $0 \leq |\beta|^2 \leq 1$, while a quantum physical system containing multiple particles can be represented as a mixed state. In addition, quantum states including superposition state and mixed state can be observed or determined as a concrete state by projection measurement, where a density matrix is exploited for representing a mixed state and a projection matrix is applied into projection measurement.

Density Matrix. In order to describe the amplitude and phase of quantum particles, a quantum superposition state can be denoted as follows:

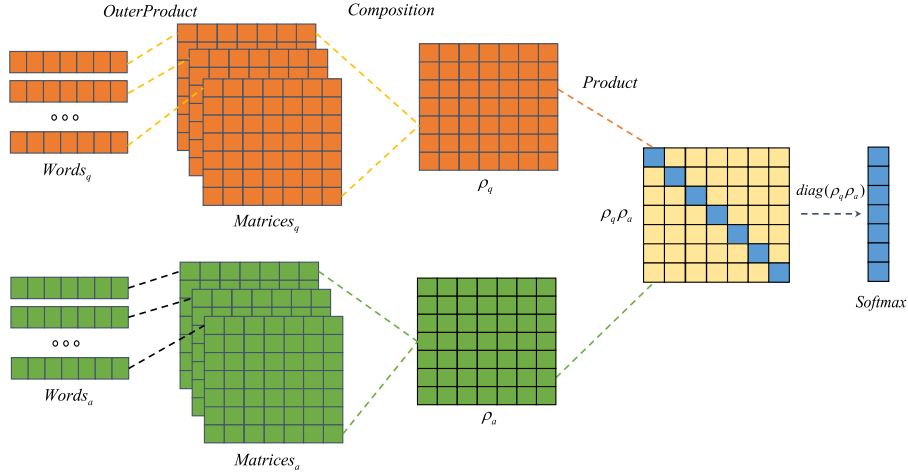


Fig. 2. NNQLM-I. Density matrix including ρ_q and ρ_a could be obtained by the outer product of real word vectors and the composition of word matrices. Then the similarity of question and answer pair could be calculated by the diagonal numbers of result matrix $\rho_q \rho_a$.

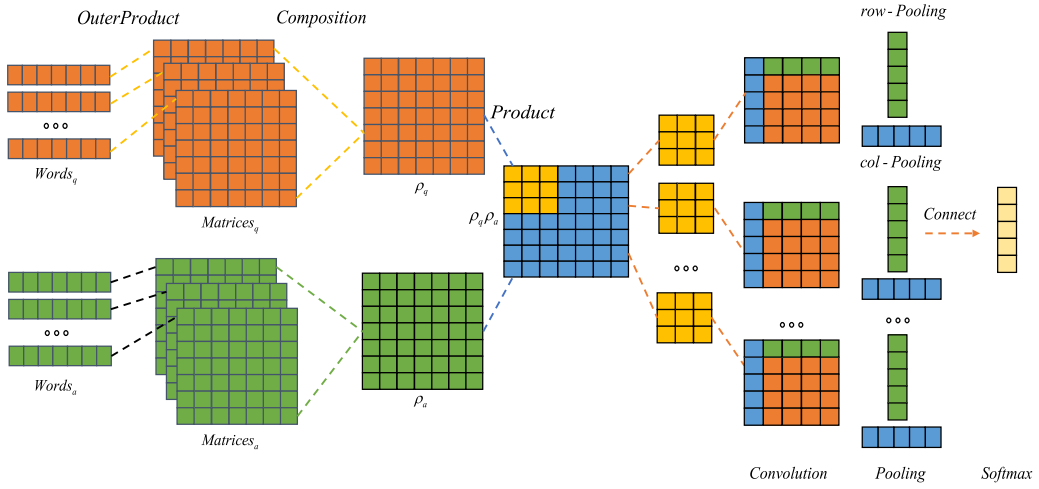


Fig. 3. NNQLM-II. It applies the convolution layer on the result matrix for extracting more abstract features on the basis of NNQLM-I.

$$|\phi\rangle = \sum_{j=1}^n r_j e^{i\psi_j} |e_j\rangle, \quad (1)$$

where i means the imaginary unit, r_j and ψ_j stand for the amplitude and phase of a single particle respectively, and $|e_j\rangle$ represents the basic state from the Hilbert space. Then the density matrix representing a quantum physical system can be described as

$$D = \sum_{i=1}^m p_i |\phi_i\rangle \langle \phi_i|, \quad (2)$$

where p_i satisfying $\sum_{i=1}^m p_i = 1$ represents the probability property, $|\phi_i\rangle$ is the superposition state in Eq. (1) and $\langle \phi_i|$ is the transpose of $|\phi_i\rangle$.

Projection Measurement. According to Gleason's theory [23], a matrix R as the measured result can be obtained with the equation $R = DM$, where D and M stand for the density and projection matrix respectively. Projection matrix can be described as $M_m = |x\rangle \langle x|$, where M_m satisfies $\sum_m M_m^\dagger M_m = I$, $|x\rangle$ comes from the orthonormal basis states in the space

of the observed system, M_m^\dagger represents the Hermitian [24] conjugate of M_m , and I stands for an identity matrix.

Activation Functions. Activation functions such as ReLU and sigmoid are commonly used to enhance the nonlinearity of neural networks. In this paper, we exploit sigmoid function described as $\sigma = \frac{1}{1+e^{-x}}$ for activation and obtaining the output value between 0 and 1 in our work.

2.2 Neural Network Based Quantum-Like Language Model (NNQLM)

Neural Network based Quantum-like Language Model (NNQLM) as a cornerstone of combination for NLP and quantum mechanics theory has been proposed for researching the linguistic subtask—question answering [25]. Question answering as a basic task in NLP is aimed at selecting the most accurate answer of the proposed problem from the candidates. The study of NNQLM for applying quantum theory into NLP for solving the fundamental task provides an original perspective. In Ref. [14], Zhang *et al.* build a close connection between quantum particles and word tokens by expressing the text sentence including question answer pairs as density matrices like ρ_q and ρ_a which are embedded into the end-to-end neural networks NNQLM-I (Fig. 2) and NNQLM-II (Fig. 3). The most significant difference between

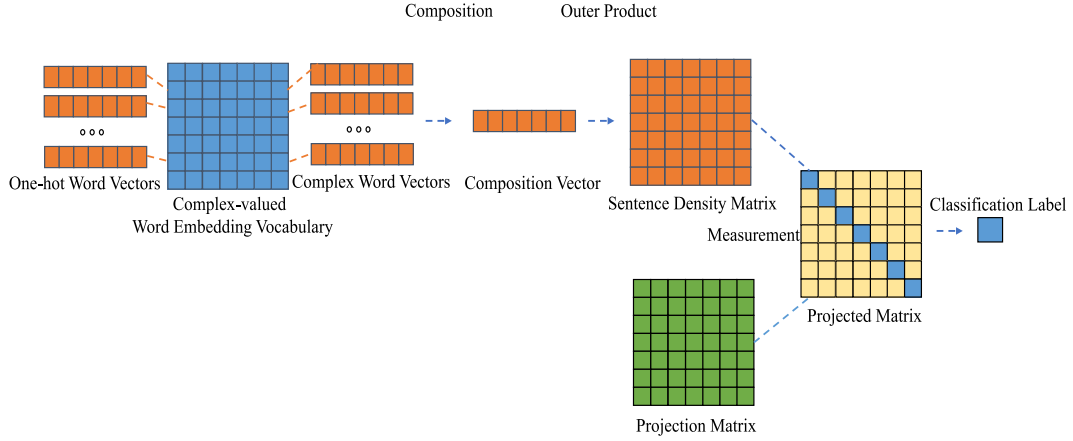


Fig. 4. CE-Sup. Complex word vectors representing amplitude and phase are mapped through the complex-valued word embedding vocabulary. Classification label as a predictor for input sentence could be obtained by the diagonal values of projected matrix.

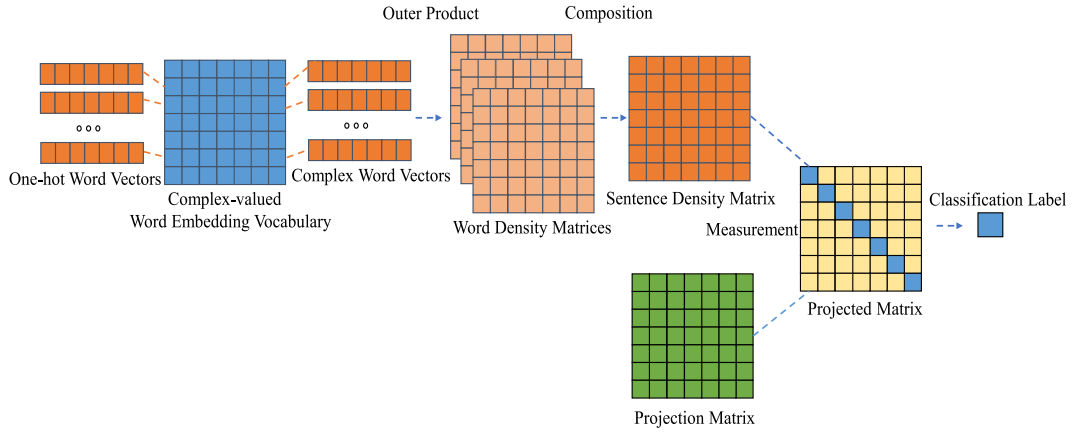


Fig. 5. CE-Mix. Different from CE-Sup, CE-Mix adopts the manner of mixed state for representing sentence density matrix.

the two models is that NNQLM-II applies the CNN on the result matrix for local feature extraction [26]. With the assistance of CNN, NNQLM-II performs better than NNQLM-I on two benchmarking QA datasets including TREC-QA and WIKIQA. The study of NNQLM is novel and oriented, but it only depends on quantum-like theory and cannot represent the complete application of quantum theory in NLP. Therefore, it is necessary to apply the complex-valued word embedding for representing the amplitude and phase of quantum particles for sake of truly simulating the quantum states.

2.3 Complex Embedding Network for Text Classification

Compared with NNQLM, complex embedding networks including Complex Embedding Superposition (CE-Sup) and Complex Embedding Mixture (CE-Mix) [1] are proposed on the basis of complex word vectors for text classification. The two models preserve the quantum properties with the complex-valued word vectors where the real and imaginary parts correspond to the amplitude and phase of quantum particles respectively [1]. Therefore, the construction of density matrix for sentence-level text is also based on complex word vectors. From Figs. 4 and 5, we can observe the unique difference between the two models is the construction of sentence density matrix.

CE-Sup. According to the concept of superposition state in quantum mechanics, CE-Sup establishes the density matrix $D_{sup} = |S\rangle\langle S|$ which is the outer product of a sentence-level vector generated by the linear combination of word vectors, where $|S\rangle = \frac{\sum_{i=1}^m \lambda_i |t_i\rangle}{\|\sum_{i=1}^m \lambda_i |t_i\rangle\|}$ ($\sum_{i=1}^m \lambda_i = 1$).

CE-Mix. The density matrix D_{mix} of CE-Mix is constructed on the basis of mixed state instead, whose building method is described in Eq. (2).

Experiments show that CE-Mix performs better than CE-Sup on the classification task. Specifically in CE-Mix, projection matrix initialized from the Hilbert space [11] is used for measuring the sentence density matrix viewed as a mixed state and determining the polarity of the sentence text, which describes the overall mathematical structure of CE-Mix model shown in Fig. 5. CE-Mix completely fits quantum theory by applying complex word vectors for simulating quantum states and performs better than CE-Sup which maybe result from the construction of sentence density matrix. However, there are still two defects existing in CE-Mix:

Firstly, there is no feature extraction before constructing the complex density matrix for input sentence where the sentence density matrix is just the sum of several word matrices and it is lack of the sequence information of language. We can use standard or variant RNN to capture

semantic features and inject the positional information into the network, where architecture conforms with grammatical structures of language, leading to the novel network achieving better performance on sentence classification.

Secondly, although it is rather original and creative to apply quantum probability [11] for predicting the polarity of sentences in CE-Mix, only diagonal values of the projected matrix are utilized for calculating the predictive value which may cause the feature information loss. Thus we apply CNN which is capable of fully considering text features on projected matrix for local feature extraction in our work.

2.4 CFN: A Complex-Valued Fuzzy Network for Sarcasm Detection in Conversations

Sarcasm detection in conversations called SDC aimed at discovering ironic emotion in daily conversations [27] has received more and more attention in recent years. Classical approaches including machine or deep learning tend to ignore the linguistic characteristics of human language like vagueness and uncertainty. In order to describe the fuzzy characteristics precisely, Zhang *et al.* propose CFN [27] composed of four parts as follows:

- Complex-valued embedding is used for representing utterance on the basis of the similarity of superposition of quantum states and the both characteristics of language mentioned above.
- Density matrix is adopted for modeling text data including utterance and its context.
- Quantum measurement layer is applied for sarcastic feature extraction.
- Dense layer is used for classification task.

In their work, amplitude word vectors are set to BERT [28] representation. Experiments are conducted on two benchmarking datasets named MUsTARD and the Reddit track. Compared with several classical models such as RCNN-RoBERTa [29], CFN obtains superior performance on the both datasets. At the same time, the authors validate the effectiveness of complex-valued word embedding and quantum measurement by ablation study.

3 QUANTUM-INSPIRED DEEP NEURAL NETWORKS ICWE-QNN AND CICWE-QNN

In this section, we present the interpretable complex-valued word embedding method based on quantum states and propose two end-to-end quantum-inspired deep neural networks.

3.1 Interpretable Complex-Valued Word Embedding (ICWE)

ICWE is a quantum-inspired complex-valued word embedding method based on GRU-attention manner, which can dramatically enhance the ability of feature extraction of neural network models on the basis of guaranteeing the model interpretability. Word embedding techniques are accepted methods used in NLP tasks [30]. Simultaneously, amplitude and phase are both required to describe quantum states in quantum mechanics. We adopt two embedding layers where one is used for generating amplitude vectors for representing

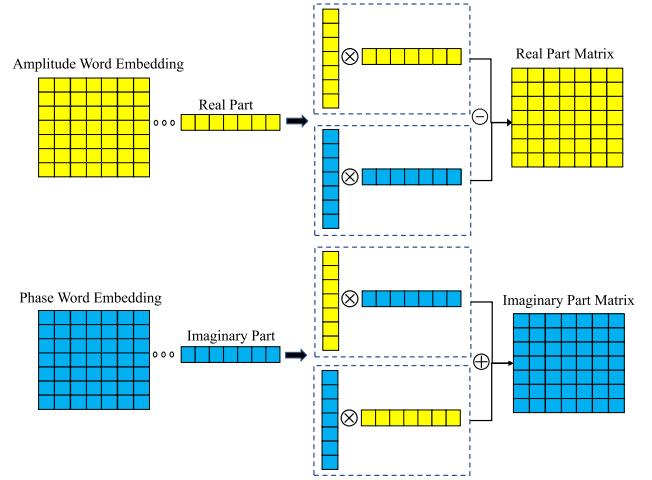


Fig. 6. Construction of density matrix based on complex-valued word embedding. Each word in the sentence can be mapped to amplitude and phase word vectors through embedding layer respectively, where \dots stands for the equation conversion by Euler formula. Operators \otimes , \oplus and \ominus represent outer product, addition and subtraction calculation respectively. Therefore, a word-level matrix actually consists of two matrices representing the real and imaginary parts of a word token.

the lexical meaning of the words and the other produces phase vectors for higher-level features representation including polarity and emotion [11]. Complex-valued word embedding consists of amplitude and phase embedding together. In fact, we use Euler formula $e^{i\psi} = \cos \psi + i \sin \psi$ to equivalently replace the related part of Eq. (1) to represent the quantum states in the calculation process for word-level matrix, which is shown in Fig. 6. Here we also give the brief introduction of deep learning methods we use in ICWE as follows:

Gated Recurrent Unit. GRU is frequently applied to capturing semantic features when processing long length text data. Compared to LSTM [31], GRU uses less gated units to selectively forget the previous text information and remember the information from the current input, leading to less trainable parameters and higher calculation speed [32]. Fig. 7 shows a standard GRU structure where x_t represents the word vector as the first input of GRU at time t . h_{t-1} representing another input contains the text information before t , and h_t is a hidden vector. The principal computation flow of GRU is presented as follows,

$$r_t = \sigma(W_{ir}x_t + b_{ir} + W_{hr}h_{t-1} + b_{hr}), \quad (3)$$

$$z_t = \sigma(W_{iz}x_t + b_{iz} + W_{hz}h_{t-1} + b_{hz}), \quad (4)$$

$$n_t = \tanh(W_{in}x_t + b_{in} + r_t \odot (W_{hn}h_{t-1} + b_{hn})), \quad (5)$$

$$h_t = (1 - z_t) \odot n_t + z_t \odot h_{t-1}, \quad (6)$$

where r_t and z_t in Eqs. (3,4) corresponding to reset and update gates respectively serve as the major functional units, $W_{ir}, W_{iz}, W_{in} \in R^{h \times d}$, $W_{hr}, W_{hz}, W_{hn} \in R^{h \times h}$ are trainable weight matrices (h notes hidden size, and d notes embedding size), and $b_{ir}, b_{iz}, b_{in}, b_{hr}, b_{hz}, b_{hn} \in R^h$ are trainable bias vectors. σ and \tanh represent different activation functions. \odot stands for the hadamard product. From Eq. (4), we know that h_t , the updated word vector at time t , combines the previous information h_{t-1} with the current information n_t from the reset gate r_t . Thus we can use the

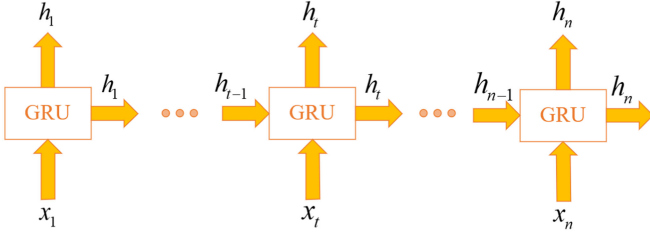


Fig. 7. GRU Structure.

updated vectors including more semantic features with GRU for constructing the density matrix in the next layer. Specifically, we could apply GRU on amplitude embedding for lexical meaning extraction. Generated model could obtain sequence information for avoiding random combination of words and become strong in handling text data, which means that sequence “*I eat apples*” would not become “*Apples eat I*” or “*Eat I apples*”.

Self-Attention Mechanism. Self-attention can help neural networks focus on the key points of sentences. For example, in this sentence “*The girl looks beautiful*”, the words “*girl*” and “*beautiful*” are apparently beneficial for network models to judge the polarity so that both of the words above should be paid more attention. In this paper, we adopt the scaled dot-product attention [20] into the designed complex-valued network. The related computation flow about self-attention is shown in Fig. 8.

After obtaining the word-level representation, we can construct a density matrix for sentence text representation prepared for classification task according to Eq. (2), where the probability $p_i = 1/m$ and m stands for the length of a sentence. The density matrix standing for a sentence-level document in our work is a novel representation made up of double matrixes containing real and imaginary information. We apply the sentence-level density matrix in the process of constructing two quantum-inspired neural networks detailed in the following subsections. In addition, before building the sentence matrix, we could use GRU-attention manner for collecting positional information and dynamically updating complex-valued word vectors. For sake of modeling quantum state of a single particle, we normalize the complex word vectors to 2-norm form.

3.2 Complex-Valued Neural Network Based on ICWE (ICWE-QNN)

In this section, we illustrate the first model we propose, named Complex-valued Neural Network based on ICWE (ICWE-QNN) for linguistic feature extraction. The complete

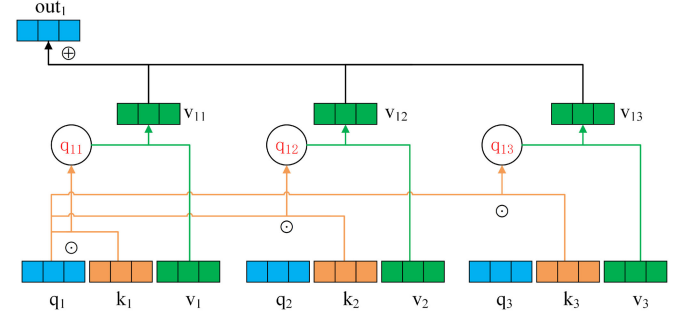


Fig. 8. Scaled dot-product self-attention. For sake of simplicity, the figure merely shows complete computation process of one output vector. Each input word vector produces an output vector after passing the self-attention layer, where each word vector can be mapped to three vectors named query (q), key (k), value (v) respectively. The query vector of each word is multiplied by all the key vectors for corresponding similarity coefficients which are shown in the circles. Finally, the output vector can be obtained by the sum of all the value vectors multiplied by relevant coefficients.

model based on the quantum mechanics and two kinds of deep learning methods is presented in Fig. 9.

GRU which is capable of capturing the relations and positional information of words in a sentence is applied into quantum theory-based model before constructing the sentence density matrix in the next layer for enhancing the extraction capabilities of quantum theory-based model for semantic features. It is equivalent to strengthen the connection among particles in quantum mixed system from our point of view. In addition, GRU is specifically applied after amplitude embedding layer because corresponding vectors in this layer represent rich lexical meaning of words in the amplitude-phase manner named complex word embedding.

We also exploit self-attention into the amplitude of designed neural network model expected for focusing on key words of text, after applying GRU into complex-valued neural network for extracting more linguistic features.

The connections between each word and the remaining words in the sentence can be established with straightforward multiplication as shown in the calculation process of self-attention in Fig. 8. Compared with RNN, there is no information loss or long-term dependency problem in this specific computation form of attention. However, dot-product self-attention does not take the positional information of text into account because in this attention layer the only operations that occur are the inner product between one vector and the others, as well as the linear superposition of word vectors. Fortunately, GRU can help encode positional information of sentence-level text into the density matrix for

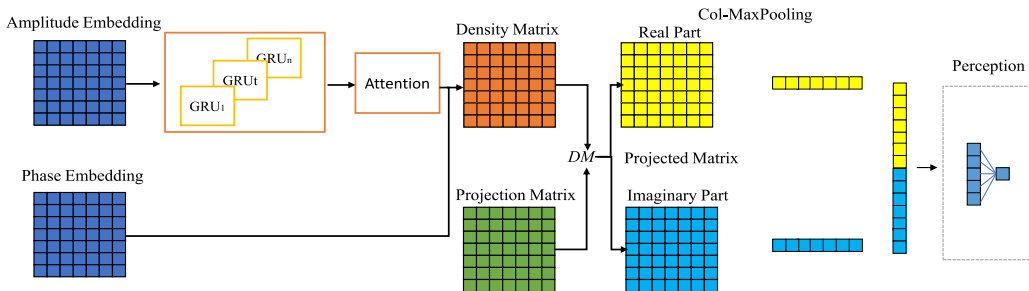


Fig. 9. The structure of ICWE-QNN. GRU and self-attention layers are targeted at extracting semantic features together. The projected matrix DM is the result matrix made up of double matrixes.

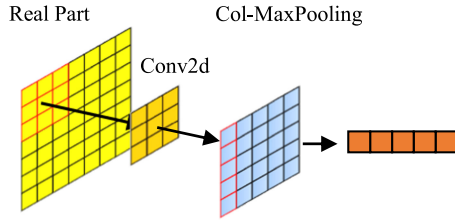


Fig. 10. Extraction for local textual features of projected matrix with CNN.

sentence representation. Therefore, we make the best of strength of the both deep learning techniques and exploit GRU-attention manner for extracting linguistic features of sentences together in the complex-valued neural network.

Two max-pooling layer for obtaining main features of each column can be conducted on the real part and imaginary part of projected matrix respectively. Thus the concatenate vector composed of real-imaginary part information of the projected matrix can be fed into a two-layer perceptron regarded as a sentence classifier for computing the classified label.

3.3 Convolutional Complex-Valued Neural Network Based on ICWE (CICWE-QNN)

In this section, we introduce the second proposed model Convolutional Complex-valued Neural Network based on ICWE (CICWE-QNN) with convolutional structure capturing local textual features of projected matrix.

3.3.1 Feature Extraction for Projected Matrix

In order to solve the problem of text-related information loss in CE-Mix model caused by neglecting the non-diagonal element of the projected matrix, we try to collect the ignored useful text features [33] of projected matrix as much as possible by involving convolutional structure in our model.

Convolutional Structure. Convolutional structure has been applied in a large number of research fields of Artificial Intelligence (AI) and deep algorithm models such as VGG [34] and ResNet [35]. Afterwards, various convolution-based modles have been universally exploited for feature extraction and achieved remarkable experimental results in AI field, especially in computer vision due to the unique and strong computation characteristics including parameter sharing and sparsity of connections shown in Eq. (7) which stands for the computation in convolutional layer. It demonstrates that the convolutional structure is essentially a special fully connected layer, where w and b still represent the related trainable weights. i.e.,

$$C_i = \sigma(w * x_i + b). \quad (7)$$

Different from the normal fully connected layer, convolution layer needs much fewer parameters in the case of obtaining the same amount of output units, which contributes to suppressing overfitting and upgrading training speed.

In our opinion, it could make the model more robust and achieve more remarkable performance on text classification by extracting more textual features. We adopt CNN with strong ability of feature extraction, where we use 3×3 convolutional kernel and pooling operation after projected matrix shown in Fig. 10 and propose the novel model in the Section 3.3.2.

3.3.2 Convolutional Complex-Valued Neural Network

Another model proposed in this paper called Convolutional Complex-valued Neural Network based on ICWE (CICWE-QNN) can further improve the performance of ICWE-QNN on classification task. 2-D convolutional structure in CICWE-QNN is exploited for capturing the local features of the projected matrix, where non-diagonal values of matrix are involved into calculation, leading to fully consideration of real-part and imaginary-part textual features of quantum mixed system. Compared with CE-Mix, our proposed model considers non-diagonal values of projected matrix and therefore decreases the loss of text information. We verify the superior performance of CICWE-QNN in Section 4.

A series of operations including convolution on the projected matrix are shown in Fig. 11, where two 2-D convolution kernels are employed side by side in the original model. The result matrix in fact consists of real-part and imaginary-part matrices. Finally, the same follow-up layers as ICWE-QNN including max-pooling and two-layer fully connected layers are implemented on double obtained feature, mapping from convolutional layer. 2-D CNN is applied rather than 1-D convolutional structure for textual information extraction, since the word information is encoded by a word-level density matrix which is an outer product of 1-D tensor in our work, leading to 2-dimensional distribution of word features on projected matrix. Therefore, it is more reasonable to adopt 2-D convolutional kernel as local structural part of CICWE-QNN in our work compared with commonly employing 1-D CNN for capturing semantic features of word vectors.

4 EXPERIMENT AND DISCUSSION

We carry out experiments on two proposed models including ICWE-QNN and CICWE-QNN designed in pytorch

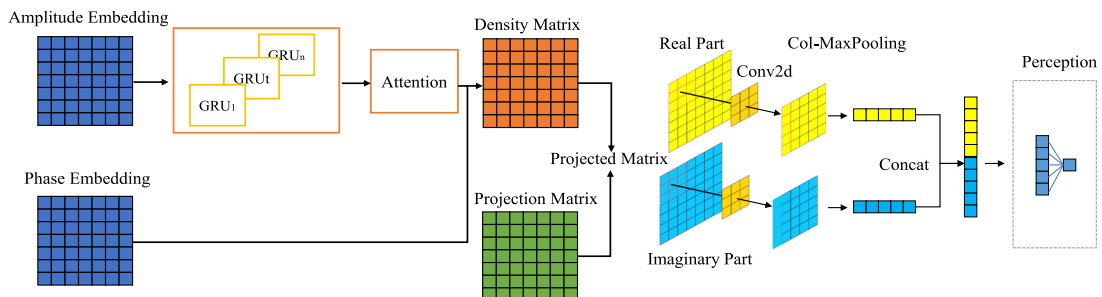


Fig. 11. The architecture of CICWE-QNN.

TABLE 2
Dataset Information After Split as Train,
Validation and Test Data

Dataset	Train	Validation	Test	Total	Labels
MR	8530	1065	1067	10662	<i>pos/neg</i>
SST	67349	872	1821	70042	<i>pos/neg</i>
SUBJ	8000	1000	1000	10000	<i>sub/obj</i>
CR	3024	364	384	3772	<i>pos/neg</i>
MPQA	8496	1035	1072	10603	<i>pos/neg</i>

Numbers in the table indicate the size of each set. All of the datasets except SST are divided into 8:1:1.

TABLE 3
Accuracy of Seven Models on Five Benchmarking Datasets

Model	MR	SST	SUBJ	CR	MPQA
CaptionRep BOW	61.9	-	77.4	69.3	70.8
DictRep BOW	76.7	-	90.7	78.7	87.2
Paragram-Phrase	-	79.7	-	-	-
CE-Sup	78.4	82.6	92.6	80.0	85.7
CE-Mix	79.8	83.3	92.8	81.1	86.6
ICWE-QNN	78.6	84.2	92.6	82.6	86.8
CICWE-QNN	78.3	85.0	93.2	83.3	87.2

The best score of each dataset is in bold.

framework 1.4 with RTX 2080 Ti, and verify the superior performance of proposed models in comparison with the complex neural network models proposed in Ref. [1]. Accuracy as the common metrics in classification tasks is analyzed for evaluating the comparative models. We directly choose the experimental results of CE-Sup, CE-Mix and several general supervised learning models shown in Ref. [1] for a comparison.

4.1 Datasets and Settings

Five binary classification (2-class) benchmarking datasets are carefully selected and preprocessed for validating the performance of proposed models. Concrete information about datasets we choose is shown in Table 2, where the division proportion of each dataset for training, validation and testing data is same as Ref. [1] for a fair comparison. There exist two categories of datasets used in the experimental phase, where Movie Review dataset (MR) [36], Stanford Sentiment Treebank dataset (SST), Customer Review dataset (CR) [37] and Opinion polarity dataset (MPQA) [38]

TABLE 4
F1 Score of Two Proposed Models on Five
Benchmarking Datasets

Model	MR	SST	SUBJ	CR	MPQA
ICWE-QNN	76.6	83.9	90.9	86.5	76.0
CICWE-QNN	76.1	83.6	92.0	86.7	78.2

The best score of each dataset is in bold.

are used to predict positive or negative sentences while Subjectivity dataset (SUBJ) [36] includes subjective or objective sentences. For model training, we adopt binary cross entropy as the loss function and Adam as the optimizer for back propagation. Pre-training Glove word vectors [39] whose dimension is selected as 100 are used for initial parameters of the embedding layer.

4.2 Results and Comparisons

We compare both the proposed models with the quantum theory-based models including CE-Sup and CE-Mix, between which there are different construction methods for a density matrix representing of the input sentence.

Our models ICWE-QNN and CICWE-QNN are both proposed on the basis of density matrix in the form of mixed state as CE-Mix outperforms CE-Sup in all of the datasets. From Table 3 about accuracy, we observe that for the case of first proposed model, ICWE-QNN surpasses CE-Mix in three datasets, which proves that sentence-level density matrix integrated with positional information provided by GRU can help improve the performance of CE-Mix. However, ICWE-QNN is inferior to CE-Mix in other two datasets, resulting from underutilizing the textual features of projected matrix. Thus we apply convolution layer full of strong extraction ability into ICWE-QNN for obtaining information maps, and the generated model CICWE-QNN obtains superior performance in four out of five datasets in comparison with CE-Mix and two traditional supervised learning models CaptionRep BOW and DictRep BOW, which verifies the effectiveness of convolutional layer for textual feature extraction. In a word, ICWE-QNN and CICWE-QNN validate that it is feasible to combine deep learning algorithms with quantum theory for completing text classification task. We guarantee the both models are designed and implemented in an identical progress and the experimental results are reliable and valid. In the end, we show the F1 score values of two proposed models in Table 4

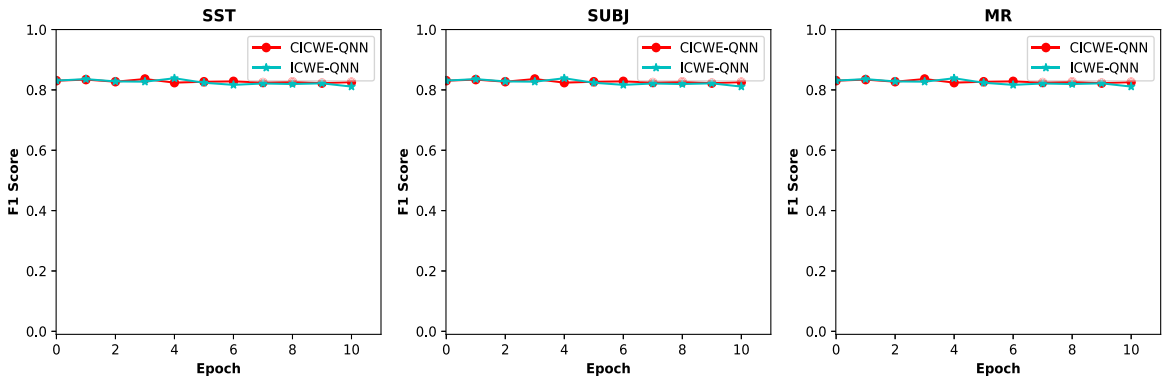


Fig. 12. F1 Score of two end-to-end classification models on three benchmarking datasets.

TABLE 5
Recall of two Proposed Models on Five Benchmarking Datasets

Model	MR	SST	SUBJ	CR	MPQA
ICWE-QNN	83.56	87.19	94.39	88.71	89.77
CICWE-QNN	82.92	89.53	94.19	88.65	88.41

The best score of each dataset is in bold.

and the F1 score trend of models on three benchmarking datasets in Fig. 12, from which we conclude that our models could obtain an end-to-end training and effectively complete the text classification task.

In addition, there is still one point about the improvement of ICWE which is also a direction we will continue to study in the future. ICWE is proposed as enhanced complex-valued embedding on the basis of feature extraction for amplitude word vectors. In our future work, we will focus on research on phase word vectors for higher-level features and generate novel embedding with common extraction for amplitude and phase word vectors.

5 CONCLUSION AND FUTURE WORK

ICWE method is proposed to design two end-to-end deep neural networks (ICWE-QNN and CICWE-QNN) for binary text classification in this paper. We make the best of the superiority of deep learning algorithms including GRU, self-attention and CNN for text feature extraction. Our proposed models prove that deep learning methods could be integrated into quantum-inspired complex neural networks and can solve the problem of text information loss in CE-Mix model caused by neglecting the important linguistic features of text. Experiments are conducted on five benchmarking classification datasets. For accuracy shown in Table 3, our proposed model CICWE-QNN has better performance than all the compared traditional models and also has better performance than CE-Mix for four datasets. Especially CICWE-QNN receives the highest promotion 16.4% on accuracy for MPQA dataset compared with traditional CaptionRep BOW model, and 2.2% for CR dataset compared with CE-Mix model. In the end, we give recall describing the performance of two proposed models shown in Table 5 and Fig. 13. In a word, it is successful to combine the complex-valued word embedding in an amplitude-phase manner with deep learning methods for capturing more semantic information.

In the future, the application of other deep learning methods in quantum theory-based neural networks can be investigated. Firstly, we would concentrate on the influence of multi-head attention mechanism on the process of constructing density matrix for sentence representation. A word, playing an important role in classification task, should have higher weight coefficient in the context. Secondly, bidirectional GRU can be also taken into account for each word obtaining the following information in the sentence. Finally, according to the competent and comprehensive survey on large-scale machine learning in Ref. [40], [41], we may consider designing a simplified model with the characteristics of quantum computing under the premise of ensuring the performance of the models on NLP tasks and reducing the computational cost of the model when dealing with large-scale datasets. We can also pay more

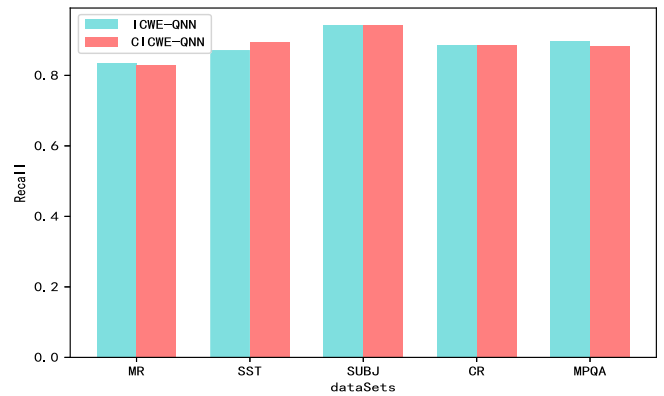


Fig. 13. Recall of two end-to-end classification models on five benchmarking datasets.

attention to the collision of quantum mechanics and deep learning in other NLP tasks such as question answer matching, machine translation and so on.

REFERENCES

- [1] Q. Li, S. Upreti, B. Wang, and D. Song, "Quantum-inspired complex word embedding," 2018, *arXiv:1805.11351*.
- [2] H. Peng *et al.*, "Hierarchical taxonomy-aware and attentional graph capsule RCNNs for large-scale multi-label text classification," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 6, pp. 2505–2519, Jun. 2021.
- [3] A. Moreo, A. Esuli, and F. Sebastiani, "Learning to weight for text classification," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 2, pp. 302–316, Feb. 2020.
- [4] S. R. Basha, J. K. Rani, and J. Yadav, "A novel summarization-based approach for feature reduction enhancing text classification accuracy," *Eng., Technol. Appl. Sci. Res.*, vol. 9, no. 6, pp. 5001–5005, 2019.
- [5] D. Liang and Y. Zhang, "AC-BLSTM: Asymmetric convolutional bidirectional LSTM networks for text classification," 2016, *arXiv:1611.01884*.
- [6] W. Yin, K. Kann, M. Yu, and H. Schütze, "Comparative study of CNN and RNN for natural language processing," 2017, *arXiv:1702.01923*.
- [7] B. Liu, Y. Zhou, and W. Sun, "Character-level text classification via convolutional neural network and gated recurrent unit," *Int. J. Mac. Learn. Cybern.*, vol. 11, no. 8, pp. 1939–1949, 2020.
- [8] X. Zhang, J. Zhao, and Y. Lecun, *Character-Level Convolutional Networks for Text Classification*. Cambridge, MA, USA: MIT Press, 2015.
- [9] M. Hughes, I. Li, S. Kotoulas, and T. Suzumura, "Medical text classification using convolutional neural networks," *Stud. Health Technol. Inform.*, vol. 235, pp. 246–50, 2017.
- [10] Y. Lan, Y. Hao, K. Xia, B. Qian, and C. Li, "Stacked residual recurrent neural networks with cross-layer attention for text classification," *IEEE Access*, vol. 8, pp. 70401–70410, 2020.
- [11] Q. Li, B. Wang, and M. Melucci, "CNM: An interpretable complex-valued network for matching," 2019, *arXiv:1904.05298v1*.
- [12] W. Zeng and B. Coecke, "Quantum algorithms for compositional natural language processing," 2016, *arXiv:1608.01406*.
- [13] L. C. Wan, C. H. Yu, S. J. Pan, G. Fei, Q. Y. Wen, and S. J. Qin, "Asymptotic quantum algorithm for the toeplitz systems," *Phys. Rev. A*, vol. 97, no. 6, 2018, Art. no. 062322.
- [14] P. Zhang, J. Niu, Z. Su, B. Wang, L. Ma, and D. Song, "End-to-end quantum-like language models with application to question answering," 2018, *arXiv:2008.09943v1*.
- [15] W. Wang, N. Yang, F. Wei, B. Chang, and M. Zhou, "Gated self-matching networks for reading comprehension and question answering," in *Proc. 55th Annu. Meeting Assoc. Comput. Linguistics*, 2017, pp. 189–198.
- [16] W. Yang *et al.*, "End-to-end open-domain question answering with bertserini," 2019, *arXiv:1902.01718*.
- [17] G. Liu and J. Guo, "Bidirectional LSTM with attention mechanism and convolutional layer for text classification," *Neurocomputing*, vol. 337, pp. 325–338, 2019.

- [18] F. Hill, K. Cho, A. Korhonen, and Y. Bengio, "Learning to understand phrases by embedding the dictionary," *Trans. Assoc. Comput. Linguistics*, vol. 4, pp. 17–30, 2016.
- [19] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," 2014, *arXiv:1412.3555*.
- [20] A. Vaswani *et al.*, "Attention is all you need," in *Proc. Adv. Neural Inform. Process. Syst.*, 2017, pp. 5998–6008.
- [21] K. S. Tai, R. Socher, and C. D. Manning, "Improved semantic representations from tree-structured long short-term memory networks," 2015, *arXiv:1503.00075*.
- [22] S. Lai, L. Xu, K. Liu, and J. Zhao, "Recurrent convolutional neural networks for text classification," in *Proc. 29th AAAI Conf. Artif. Intell.*, 2015, pp. 2267–2273.
- [23] A. M. Gleason, "Measures on the closed subspaces of a hilbert space," *J. Math. Mech.*, vol. 6, no. 6, pp. 885–893, 1957.
- [24] R. De Wolf, "Quantum computing: Lecture notes," 2019, *arXiv:1907.09415*.
- [25] M. Tan, C. Dos Santos, B. Xiang, and B. Zhou, "Improved representation learning for question answer matching," in *Proc. 54th Annu. Meeting Assoc. Comput. Linguistics*, 2016, pp. 464–473.
- [26] X. Zhu, S. Zhang, R. Hu, Y. Zhu, and J. Song, "Local and global structure preservation for robust unsupervised spectral feature selection," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 3, pp. 517–529, Mar. 2018.
- [27] Y. Zhang *et al.*, "CFN: A complex-valued fuzzy network for sarcasm detection in conversations," *IEEE Trans. Fuzzy Syst.*, early access, Apr. 12, 2021, doi: [10.1109/TFUZZ.2021.3072492](https://doi.org/10.1109/TFUZZ.2021.3072492).
- [28] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," 2018, *arXiv:1810.04805*.
- [29] R. A. Potamias, G. Siolas, and A.-G. Stafylopatis, "A transformer-based approach to irony and sarcasm detection," *Neural Comput. Appl.*, vol. 32, no. 23, pp. 17 309–17 320, 2020.
- [30] X. Mao, S. Chang, J. Shi, F. Li, and R. Shi, "Sentiment-aware word embedding for emotion classification," *Appl. Sci.*, vol. 9, no. 7, 2019, Art. no. 1334.
- [31] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [32] R. Rana, "Gated recurrent unit (GRU) for emotion classification from noisy speech," 2016, *arXiv:1612.07778*.
- [33] S. Zhang, Z. Qin, C. X. Ling, and S. Sheng, "Missing is useful": Missing values in cost-sensitive decision trees," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 12, pp. 1689–1693, Dec. 2005.
- [34] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
- [35] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
- [36] B. Pang and L. Lee, "Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales," 2005, *arXiv:cs/0506075*.
- [37] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2004, pp. 168–177.
- [38] J. Wiebe, T. Wilson, and C. Cardie, "Annotating expressions of opinions and emotions in language," *Lang. Resour. Eval.*, vol. 39, no. 2, pp. 165–210, 2005.
- [39] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *Proc. Conf. Empirical Methods Nat. Lang. Process.*, 2014, pp. 1532–1543.
- [40] W. Meng, W. Fu, S. Hao, H. Liu, and X. Wu, "Learning on big graph: Label inference and regularization with anchor hierarchy," *IEEE Trans. Data Eng.*, vol. 29, no. 5, pp. 1101–1114, May 2017.
- [41] M. Wang, W. Fu, X. He, S. Hao, and X. Wu, "A survey on large-scale machine learning," *IEEE Trans. Knowl. and Data Eng.*, early access, Aug. 11, 2020, doi: [10.1109/TKDE.2020.3015777](https://doi.org/10.1109/TKDE.2020.3015777).



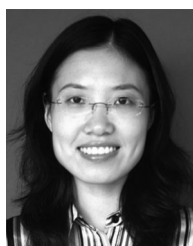
Jinjing Shi received the BS and PhD degrees from the School of Information Science and Engineering, Central South University, Changsha, China, in 2008 and 2013, respectively. She is currently an associate professor with the School of Computer Science and Engineering, Central South University. She was selected in the Shenghua lying talent program of Central South University and Special Foundation for Distinguished Young Scientists of Changsha in 2013 and 2019, respectively. Her research interests include quantum computation and quantum cryptography. She has presided over the National Natural Science Foundation Project of China and that of Hunan Province. There are 50 academic papers published in important international academic journals and conferences. She was the recipient of the second prize of natural science and the outstanding doctoral dissertation of Hunan Province in 2015 and the Best Paper Award in the international academic conference MSPT2011 and Outstanding Paper Award in IEEE ICACT2012.



Zhenhuan Li received the BS degree from Xiangtan University. He is currently working toward the master's degree with the School of Computer Science and Engineering, Central South University, Hunan, China. His research interests mainly include quantum computing, deep learning, and quantum neural networks.



Wei Lai received the BS degree from the East China University of Technology. He is currently working toward the master's degree with the School of Computer Science and Engineering, Central South University, Hunan, China. His research interests mainly include quantum computing, deep learning, and quantum neural networks.



Fangfang Li received the MS degree in geographic information system and the PhD degree in photogrammetry and remote sensing from Wuhan University, China, in 2006 and 2009, respectively. Her research interests include machine learning and textual mining.



Ronghua Shi He received the BS, MS, and PhD degrees in computer application technology from Central South University (CSU) in 1986, 1989, and 2007, respectively. He is currently the supervisor of Ph.D. students, and the team leader of the communication system and network security group with the School of Computer Science and Engineering, Central South University, Changsha 410083, China. He is also the executive director of Railways Specialty Committee, chairman of Hunan Internet of Things Committee, vice-chairman of the Hunan

Higher Education Computer Society Professional Committee, and the executive director of Provincial Communication Society. He has received the State Council Special Allowance. His professional field covers computer science and technology, information and communication engineering, etc., He has authored or coauthored more than 80 articles in domestic and foreign academic journals, which include about 40 SCI or EI articles. His research interests include network and information security, quantum cryptography, quantum secure communications, etc. His teaching curriculums contain Special Topics of information security technology, modern cryptography theory, and application, computer network communications, introductions to information science, etc. In recent years, he has hosted more than 10 research projects such as the National 863 Project, Natural Science Foundation Project, and the Ministry of Education Doctoral Fund Project, etc. He was the recipient of the two Provincial and Ministerial Appraisal projects, two projects for the Second Prize of Provincial Science and Technology Progress and one project for that of Third Prize, and the two articles for the Second Prize of Outstanding Papers of Natural Science in Hunan Province.



Shichao Zhang (Senior Member, IEEE) received the PhD degree in computer science from Deakin University, Australia. He is currently a China National-Title professor with the School of Computer Science and Technology, Central South University, China. His research interests include information quality and pattern discovery. He was/is an associate editor for the *ACM Transactions on Knowledge Discovery from Data*, *IEEE Transactions on Knowledge and Data Engineering*, *Knowledge and Information Systems*, and the *IEEE Intelligent Informatics Bulletin*. He is a senior member of the IEEE Computer Society and a member of the ACM.

► For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/csdl.



Yanyan Feng received the BS degree from Henan University and the MS degree from Central South University. she is currently working toward the PhD degree with the School of Computer Science and Engineering, Central South University, Changsha, China. Her research interests mainly include quantum communication, quantum circuit models and their applications, and quantum circuit learning supremacy.