Neural Network Model Reconstructed from Entangled Quantum States

1st Junwei Zhang
College of Intelligence and Computing,
Tianjin University,
Tianjin, China.
Email: junwei@tju.edu.cn

2nd Zhao Li (✉)
Zhejiang University, Zhejiang, China.
Hangzhou Yugu Technology Co., Ltd., China.
Link2Do Technology Ltd., China.
Email: zhao_li@zju.edu.cn

3rd Jianmao Xiao
Jiangxi Normal University,
Jiangxi, China.
Email: jm_xiao@jxnu.edu.cn

4th Ming Li
The Key Laboratory of Intelligent Education Technology and Application of Zhejiang Province, Zhejiang Normal University,
Zhejiang, China. Email: mingli@zjnu.edu.cn

2022 International Joint Conference on Neural Networks (IJCNN) | 978-1-7281-8671-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/IJCNN55064.2022.9892246

***Abstract*—Neural Networks (NNs) have received extensive attention and research due to their ability to extract and combine different features non-manually and to mine the internal relationships between features. In quantum mechanics, the entangled state simultaneously describes the classical and non-classical correlation between subsystems, so it can reveal important quantum phenomena such as non-locality, entanglement, etc., but its essence is the characterization of strong statistical relationships. Considering the superiority of the entangled state, this article attempts to use the entangled state to reconstruct the neurons of NNs to achieve that the network model can characterize the strong statistical relationship between the features, namely, the classical and non-classical correlation. Specifically, based on concurrence, a quantification method of entanglement, we propose a regularizer that can constrain a state vector to an entangled state, and apply it to the optimization process to ensure that the vector passed to the neuron is a legal entangled state. Finally, the entangled state is measured to obtain the output of the neuron. Experimental results show that our model is better than the baseline. Moreover, it performs well compared to the model inspired by quantum entanglement.**

***Index Terms*—Neural Networks, Entangled Quantum States, Quantum Entanglement, Strong Statistical Relationships**

I. Introduction

Neural Networks (NNs) are computational models constructed by simulating the human brain nervous system [1], [2], which has received extensive attention and research due to its ability to extract and combine different features non-manually and to mine the internal relationships between features [3]. The deep learning technology with NNs as the core component has achieved good results in various artificial intelligence fields, but it is still not on par with brain-like intelligence [4], [5]. For example, humans can build internal connections between features and make decisions based on a small number of samples; moreover, humans are good at mining abstract connections between features and performing associative analysis.

To improve the learning ability of deep learning models to face more difficult feature extraction tasks, research work based on classic paradigms imitates the human brain nervous system to build larger and more complex model structures. However, large and complex models require more computing power and more corpus, and resource constraints will eventually limit their development [6]. The non-classical paradigm, such as quantum mechanics theory, has attracted the attention of researchers due to its unique advantages. Rebentrost et al. [7] coded the network into the probability amplitudes of quantum states to store an exponentially large network in a polynomial number of qubits. Chen et al. [8] proposed a quantum probabilistic network model, which leverages quantum parallelism to track all possible network states to improve performance. In addition, Verdon et al. [9] proposed a quantum graph neural network and Cong et al. [10] proposed a quantum convolutional neural network.

This article attempts to reconstruct the neural network model from the core component of NNs, that is, neurons. The neurons of NNs imitate biological neurons, which receive input information, integrate the information to achieve the exchange of information, and transmit it out, but for biological neurons, the process of integrating nerve impulses is complicated and not well known [11]. In quantum mechanics, the entangled state simultaneously describes the classical and non-classical correlation between subsystems, so it can reveal important quantum phenomena [12] such as non-locality [13], entanglement [14], etc., but its essence is the characterization of strong statistical correlation, and the correlation is stronger than the classical statistical correlation [15].

Considering the superiority of the entangled state, this article attempts to use the entangled state to reconstruct the neurons of NNs to achieve that the network model can characterize the strong statistical relationship between the features, namely, the classical and non-classical correlation. Specifically, based on concurrence, a quantification method of entanglement, we propose a regularizer that can constrain a state vector to an entangled state, and apply it to the optimization process to ensure that the vector passed to the neuron is a legal

978-1-7281-8671-9/22/$31.00 ©2022 IEEE

entangled state. Finally, the entangled state is measured to obtain the output of the neuron. Based on the neurons, we constructed a network model, called Quantum Neural Network (QNN). Experimental results show that QNN is better than the baseline. Moreover, it performs well compared to the model inspired by quantum entanglement.

The contribution of this paper is to introduce non-classical correlation into NNs. To achieve this goal, we propose a regularizer based on concurrence to constrain the state vector as an entangled state, define the measurement process of the entangled state as a neuron, and construct a network structure. Due to space constraints, we only built the most basic network model, but it has the ability to build more complex models.

II. Preliminaries and Our Proposed Strategies

For readers without a background in quantum information processing, it is difficult to accept that quantum mechanics theory, a science that describes the laws of motion of microscopic particles, is used to solve problems in the field of information processing. In this section, we will answer the questions that readers may have, and propose strategies that will be applied to model construction and quantitative analysis.

A. Why can quantum mechanics be used?

Although quantum mechanics is generally regarded as microphysical theory, its connotation is about information rather than physics. The mathematical principles of quantum mechanics are called **quantum probability theory**. Since Hardy [16], the information nature of quantum mechanics has been increasingly clarified: it can be proved that quantum mechanics can be formally derived from 4-5 general information processing axioms which are conceptually natural and technically concise. Moreover, if a specific limitation is imposed on the set of information processing axioms derived from quantum mechanics, a special case of quantum mechanics, namely **classical probability theory**, can be derived. Therefore, the law of quantum mechanics should not only be regarded as the law of the microphysical world, but should be regarded as the general law of information processing [17].

Moreover, we should also clearly realize that quantum mechanics theory (also called quantum probability theory) is more general than classical probability theory. Quantum probability theory not only covers the classical probability theory, but also reveals unique features, such as negative probability [18] and the strong statistical correlation revealed [15] by QE that will be used in this article.

B. Why use quantum probability theory?

To answer the question, we consider two simple systems exhibiting some degree of similarity:

- a classical system consisting of a projectile initially at rest, which explodes and produces two fragments carrying opposite momenta; and
- a quantum system similar to the classical system, that is, the maximum entangled state or Bell state.

We assume that each system has N copies, and measure the subsystems of each system, a and b, and the composite system, ab, to obtain their expected values, $\langle a\rangle$, $\langle b\rangle$ and $\langle ab\rangle$.

We first consider the classical system. Suppose that the initial angular momentum of the projectile is Θ, when the projectile produces two fragments after the explosion, their angular momentum is Θ_1 and Θ_2 respectively, and $\Theta_2 = -\Theta_1$. Let $\bar{\alpha}$ and $\bar{\beta}$ be two arbitrarily selected unit vectors, and the angle between them is θ. When the experiment is repeated N times, the observable $sign(\bar{\alpha}\cdot\Theta_1)$ is performed on the first fragment, and the result is $a_i = \pm 1$, $1 \leq i \leq N$; the observable $sign(\bar{\beta}\cdot\Theta_2)$ is performed on the second fragment, and the result is $b_i = \pm 1$. The expected values of the two measurements are, respectively:

$$\langle a\rangle = \frac{1}{N}\sum_{i=1}^{N} a_i \quad and \quad \langle b\rangle = \frac{1}{N}\sum_{i=1}^{N} b_i. \tag{1}$$

Their values are typically of the order $1/\sqrt{N}$, thus, very close to zero. The expected value of the composite system is

$$\langle ab\rangle^{classical} = \frac{1}{N}\sum_{i=1}^{N} a_i b_i. \tag{2}$$

Simple mechanical considerations show that if the momenta, Θ_1 and Θ_2, are uniformly distributed, then

$$\langle ab\rangle^{classical} = \frac{\theta-(\pi-\theta)}{\pi} = -1 + \frac{2\theta}{\pi}. \tag{3}$$

The largest correlation occurs when $\theta = \pi$:

$$\max_{\theta}\langle ab\rangle^{classical} = -1 + \frac{2\theta}{\pi} \tag{4}$$

$$= -1 + 2 = +1. \tag{5}$$

Next, we examine the quantum system. Let $\bar{\alpha}$ and $\bar{\beta}$ be two arbitrarily chosen unit vectors, and the angle between them is θ. The observables for the two particles are $\bar{\alpha}\cdot\sigma_1$ and $\bar{\beta}\cdot\sigma_2$ respectively, where σ_1 and σ_2 are the Pauli spin matrices for the two particles. Because the quantum system is Bell state,

$$\sigma_2|\psi\rangle = -\sigma_1|\psi\rangle \tag{6}$$

where $|\psi\rangle$ represents a vector under the Dirac symbol. The same as the measurement result of the classical system, the result of the quantum system is also $a_i, b_i = \pm 1$. If we repeat the quantum experiment N times, then quantum mechanics predicts that the expected value of the measurement results is equal to zero:

$$\langle a\rangle = 0 \quad and \quad \langle b\rangle = 0. \tag{7}$$

The expected value of the composite system is

$$\langle ab\rangle^{quantum} = \langle\psi^{\dagger}|(\bar{\alpha}\cdot\sigma_1)(\bar{\beta}\cdot\sigma_2)|\psi\rangle \tag{8}$$

$$= -\cos(\theta). \tag{9}$$

We apply the identitys

$$(\bar{\alpha}\cdot\sigma)(\bar{\beta}\cdot\sigma) = \bar{\alpha}\cdot\bar{\beta} + i(\bar{\alpha}\times\bar{\beta})\cdot\sigma \tag{10}$$

to the quantum system in an entangled state, $\sigma_2|\psi\rangle = -\sigma_1|\psi\rangle$, and conclude that

$$\langle ab\rangle^{quantum\ state} = -\bar{\alpha}\cdot\bar{\beta} = -\cos(\theta), \tag{11}$$

with θ, $\bar{\alpha}$ and $\bar{\beta}$ arbitrarily chosen by the experimenter. The largest correlation occurs for $\theta = \pi$, then

$$\max_{\theta}\langle ab\rangle^{quantum\ state} = -\cos(\pi) = +1. \tag{12}$$

In the interval $\theta \in [0, \pi]$, the two correlations are equal with the values 0 or ± 1, when θ takes the values $\frac{\pi}{2}$ or 0 and π, respectively. In the rest of the interval, the classical correlation is a linear function, while the quantum correlation is a cosine function that runs below the straight line in the negative value range for $\theta \in (0, \frac{\pi}{2})$ and above it in the positive valus range for $\theta \in (\frac{\pi}{2}, \pi)$, repectively.

At this point, we can introduce an important measure of correlation, i.e., Correlation Function [19],

$$\mathcal{CF}_{ab} = \langle ab\rangle - \langle a\rangle\langle b\rangle, \tag{13}$$

to quantify the correlation revealed by quantum probability theory and by classical one, and get the following conclusions:

Corollary 1: For the physical composite systems ab composed of two subsystems with the same momentum but opposite directions, the classical correlation described by the classical probability theory, using the Correlation Function Eq. 13 as a measurement method, is smaller than the quantum correlation described by the quantum probability theory, i.e.,

$$|\mathcal{CF}_{ab}^{quantum}| \geq |\mathcal{CF}_{ab}^{classical}|. \tag{14}$$

The diagram of the inequality is shown in Fig. 1. The proof method is shown in the derivation process.

Since in the density matrix form of $|\psi\rangle$, i.e., $\rho = |\psi\rangle\langle\psi|$, the elements on the off-diagonal line describe the coherence between the subsystems, so we can define that ρ has full correlation (i.e., quantum correlation) and $diag(\rho)$ can be defined as having only classical correlation [20]. The more general two-qubit composite system has the following conclusions:

Corollary 1′: When ρ is a two-qubit composite system, then the non-classical correlation ($\mathcal{NCC}$) described by the entangled state can be defined by the correlation function ($\mathcal{CF}$), i.e., Eq. (13), as:

$$|\mathcal{CF}(\rho)| \geq |\mathcal{CF}(diag(\rho))|. \tag{15}$$

C. How to use quantum correlation?

Quantum correlation is a combination of classical and non-classical correlation. The correlation between two systems is often called entanglement or quantum entanglement. Quantum correlation is stronger than classical one, and it is an important quantum resource [15], [21].

In quantum mechanics, physical systems are described by quantum states, and quantum states can be divided into pure and mixed states. The pure state is expressed as a vector $|\psi\rangle$ in the complex Hilbert space $\mathcal{H}$, and its density matrix expression is $\rho = |\psi\rangle\langle\psi|$ where $\langle\psi| = (|\psi\rangle)^{\dagger}$ and $\dagger$ stands for conjugate transpose. The state vector in the pure state is normalized,

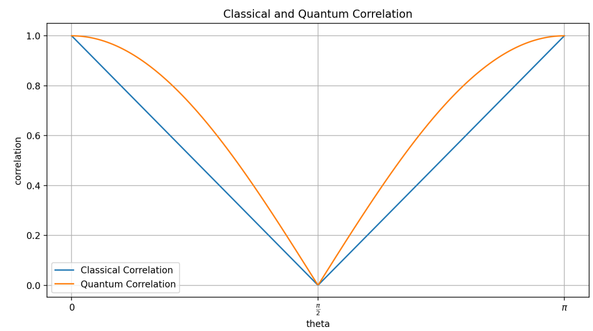

Fig. 1. The correlation described by quantum probability theory is stronger than that described by classical one.

i.e., $\langle\psi|\psi\rangle = 1$; on the contrary, the state vector in the mixed state is $\langle\psi|\psi\rangle < 1$. The mixed state can only be expressed in the form of a density matrix, that is, the mixed state is a probabilistic mixture of pure states,

$$\rho = \sum p_i\rho_i = \sum p_i|\psi_i\rangle\langle\psi_i| \tag{16}$$

where $0 \leq p_i \leq 1$ and $\sum p_i = 1$. It can be seen that the pure state is a special case of the mixed state.

Example 1: The state of a qubit is described by a two-dimensional complex vector $|\psi\rangle$, and $|0\rangle$ and $|1\rangle$ are the basis vectors of the complex space $\mathcal{H}$, then the qubit can be expressed as

$$|\psi\rangle = a|0\rangle + b|1\rangle \tag{17}$$

where $|0\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $|1\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, $a, b \in \mathbb{C}$ and $(a)^2 + (b)^2 = 1$.

The pure state of a composite system is a vector on the tensor space $\mathcal{H}\otimes\mathcal{H}\otimes\cdots\otimes\mathcal{H}$. Let $|h\rangle$, $h \in \{i, j, ..., k\}$, be the basis vector of $\mathcal{H}$, then there is

$$|\psi\rangle = \sum_{i,j,\ldots,k} a_{ij\ldots k}|ij\ldots k\rangle \tag{18}$$

where $|ij...k\rangle = |i\rangle\otimes|j\rangle\otimes\cdots\otimes|k\rangle$, $\sum a_{ij\ldots k}\cdot a^*_{ij\ldots k} = 1$ and $a_{ij\ldots k} \in \mathbb{C}$. The construction method of the mixed state of the composite system is the same as that of the single system.

Example 2: The state vector $|\psi\rangle$ of a two-qubit composite system is described by a vector on the $2\otimes 2$ tensor space $\mathcal{H}\otimes\mathcal{H}$, which is generally expressed as

$$|\psi\rangle = a_{11}|00\rangle + a_{12}|01\rangle + a_{21}|10\rangle + a_{22}|11\rangle \tag{19}$$

where $\sum(a_{ij})^2 = 1$, $a_{ij} \in \mathbb{C}$ and $i, j = \{1, 2\}$.

So far, we have given the form of single and composite systems, and also given the definition of their pure and mixed states. Because the strong statistical correlation, i.e., quantum correlation, revealed by quantum entanglement involved in this article occurs in composite systems, the following will focus on composite systems. Moreover, the study of quantum mechanics is still in its infancy, and many problems have not been thoroughly studied, such as quantum entanglement involved in this article. At this stage, the academic community only has a clear understanding of the two-qubit composite system, so our model will also be constructed based on the two-qubit composite system.

The entanglement degree of the separable state is zero, but for the non-separable state, namely, the entangled state, an appropriate quantity is needed to measure the degree of entanglement. The entanglement quantification methods include *the entanglement of formation* [22], *the negative degree* [23], *the relative entropy of entanglement* [24] and *the entanglement cost* [25], but because most entanglement quantification methods have inherent computational difficulties, they are not suitable for use in calculation models. Since concurrence [14] is easy to calculate in special cases and can be used as a criterion to constrain composite systems, this article will focus on using concurrence.

For general two-body pure state $|\psi\rangle$, concurrence is defined as

$$\mathcal{C}(|\psi\rangle) = \sqrt{2(1 - Tr(\rho_1^2))} \tag{20}$$

where $\rho_1 = Tr_2(|\psi\rangle\langle\psi|)$ and Tr represents the trace of the matrix, and concurrence of its mixed state ρ is defined as

$$\mathcal{C}(\rho) = \sum_i p_i \mathcal{C}(|\psi_i\rangle) \tag{21}$$

where $0 \leq p_i \leq 1$ and $\sum p_i = 1$. For the special case of two-qubit pure states

$$|\psi\rangle = a_{11}|00\rangle + a_{12}|01\rangle + a_{21}|10\rangle + a_{22}|11\rangle \tag{22}$$

where $\sum(a_{ij})^2 = 1$, $a_{ij} \in \mathbb{C}$ and $i, j = \{1, 2\}$, concurrence is defined as

$$\mathcal{C}(|\psi\rangle) = 2|a_{11}a_{22} - a_{12}a_{21}|. \tag{23}$$

Therefore, we can give a basic constraint on the probability amplitudes for the two-qubit entangled state:

Corollary 2: Let $|\psi\rangle$ be a two-qubit pure state, e.g.,

$$|\psi\rangle = a_{11}|00\rangle + a_{12}|01\rangle + a_{21}|10\rangle + a_{22}|11\rangle \tag{24}$$

where $\sum(a_{ij})^2 = 1$, $a_{ij} \in \mathbb{C}$ and $i, j = \{1, 2\}$, and when its probability amplitudes satisfie the constraint

$$a_{11}a_{22} - a_{12}a_{21} \neq 0 \tag{25}$$

then $|\psi\rangle$ is an entangled state.

The proof is given in the supplemental material.

III. Network Model with Quantum Correlation

In this section, we will reconstruct the neurons of NN based on the entangled state (here mainly refer to the neurons in the hidden layer) to achieve the purpose of integrating quantum correlation revealed by QE into the NN model. To make readers have a macro understanding, but also for narrative simplicity, we will first reconstruct the simplest three-layer NN model (that is, there is only one input layer, one hidden layer and one output layer) to present the method of reshaping the classical NN model. In real application scenarios, the structure of the model can be modified as needed.

The content of this section is arranged as follows: 1. describe how to construct a neuron from entangled states; 2. describe how to build the simplest three-layer NN model from the reconstructed neuron; 3. the parameter optimization method of the model.

A. Reconstructing neurons from entangled states

The neurons in the hidden layer in the classical NN model can be formally defined as

$$y = \mathcal{F}(\mathbf{x}) = \sigma(\mathbf{W}\mathbf{x} + b) \tag{26}$$

where $\mathbf{x} = [x_1, x_2, ..., x_n]^\intercal \in \mathbb{R}^{n\times 1}$ is the input of the neuron, $y \in \mathbb{R}^{n\times 1}$ is the output, σ is the activation function, $\mathbf{W} \in \mathbb{R}^{n\times n}$ is weight and $b \in \mathbb{R}^{n\times 1}$ is bias. $\mathbf{W}$ and b are parameters that need to be learned. We know that the function of the neuron in the classical NN model, $\mathcal{F}(\cdot)$, is to complete a non-linear transformation based on σ.

Imitating the classic neuron, we can achieve a function similar to the classic neuron through the measurement process of the entangled state. First of all, we need to obtain an two-qubit quantum system. Since this quantum system is a 4-dimensional vector, it is necessary to map the n-dimensional input vector $\mathbf{x}$ to a 4-dimensional vector $\mathbf{v} \in \mathbb{R}^{4\times 1}$ and make it a legal pure state $|\Phi\rangle$. Then we define the following operation

$$\mathbf{v} = map(\mathbf{x}) = \mathbf{M}\mathbf{x} \tag{27}$$

where $\mathbf{M} \in \mathbb{R}^{4\times n}$, and

$$|\Phi\rangle = f(\mathbf{v}) = \frac{1}{\mathcal{N}}\mathbf{v} \tag{28}$$

where $\mathcal{N} = \sqrt{\sum_{i=1}^4 \mathbf{v}_i^2}$. In the end, we get a method to generate pure state, i.e.,

$$|\Phi\rangle = f(\mathbf{M}\mathbf{x}) \tag{29}$$

where $\mathbf{M}$ is the parameter that need to be learned, and $\mathbf{x}$ is the input vector of the neuron.

At this point, we have obtained the measured quantum state $|\Phi\rangle$, but at this time it is not necessarily an entangled state, so we need to define a regularizer and add the regularizer to the objective function to constrain the quantum state to become an entangled state. Inspired by Corollary 2, the regularizer is defined as

$$\mathcal{R}(|\Phi\rangle) = 1 - |\langle\Phi|\tilde{\Phi}\rangle| \tag{30}$$

where $|\tilde{\Phi}\rangle = (\sigma_y \otimes \sigma_y)|\Phi\rangle$ and $\langle\Phi| = (|\Phi\rangle)^\dagger$. Here σ_y is the Pauli operator, $\sigma_y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}$. The derivation process is given in the supplemental material.

Finally, we only need to use a measurement operator to measure the entangled state to get the final output value of the neuron. According to the definition of quantum measurement [26], the measurement operator can be all possible ground states in space, so the measurement operator can be defined as

$$\Pi = |\mathbf{s}\rangle\langle\mathbf{s}| \tag{31}$$

where $|\mathbf{s}\rangle = f(\mathbf{s}) = |\mathbf{s}_1\rangle \otimes |\mathbf{s}_2\rangle = f(\mathbf{s}_1) \otimes f(\mathbf{s}_2)$, $\mathbf{s} = \mathbf{s}_1 \otimes \mathbf{s}_2$ and $\mathbf{s}_i$, $i = \{1, 2\}$, are a 2-dimensional vector that needs to be learned. From this we can get the output of the neuron through measurement, and the final form of the neuron is

$$y = Tr(\Pi|\Phi\rangle\langle\Phi|) = \langle\Phi|\Pi|\Phi\rangle \quad (32)$$
$$= (\langle\Phi|\mathbf{s}\rangle)^2 = (f(\mathbf{s}^\intercal)f(\mathbf{Mx}))^2 \quad (33)$$

Comparing Eq. (26) and Eq. (32), we can find that they also have two quantities that need to be learned, namely, $\mathbf{M}$ and $\mathbf{s}$, but their main difference is that the neurons described by the entangled state require more constraints.

B. A simple example

In the previous section, we have been able to construct a neuron described by an entangled state. Next, we will show how to use the neuron to build the simplest network model. The model is similar to the classic three-layer NN model, which can be divided into input layer, hidden layer and output layer.

1) Input Layer:: Like the classical NN model, the input layer receives an input vector $\mathbf{x} \in \mathbb{R}^{N\times 1}$,

$$\mathbf{x} = [x_1, x_2, ..., x_N]^\intercal \quad (34)$$

where N is the dimension of the vector and also represents the number of neurons in the input layer.

2) Hidden Layer:: Compared with the simple mapping of the classic model, our neurons in the hidden layer are a measurement process of entangled states, so for the description of the hidden layer, we need to divide into two steps, namely, prepare a set of entangled states and measure the set of entangled states.

For a given input vector $\mathbf{x}$, we apply $map(\cdot)$ and $f(\cdot)$ one by one to make it a set of pure states, i.e.,

$$\Phi = [|\Phi_1\rangle, |\Phi_2\rangle, ..., |\Phi_H\rangle] \quad (35)$$
$$= f(map(\mathbf{x})) \quad (36)$$
$$= f(\mathbf{Mx}) \quad (37)$$

where $\Phi \in \mathbb{R}^{4\times H\times 1}$, $\mathbf{M} \in \mathbb{R}^{4\times H\times N}$ and H is the number of neurons in the hidden layer.

Since $|\Phi_i\rangle$ obtained here is a pure state and not the required entangled state, we constrain it to become a legal entangled state by adding $\mathcal{R}(\cdot)$ to Φ. The specific operation method is given in the parameter optimization section Sec. III-C.

With the entangled state, we can measure it to get the output $\mathbf{z}$ of the hidden layer, i.e.,

$$\mathbf{z} = (\Phi^\intercal f(\mathbf{s}))^2 \quad (38)$$

where $\mathbf{z} \in \mathbb{R}^{1\times H}$ and $\mathbf{s} \in \mathbb{R}^{4\times 1}$.

3) Output Layer:: Similar to the classical NN model, the output layer of our model is also a fully connected layer with a normalized function $Softmax(\cdot)$. Its form is

$$\mathbf{y} = Softmax(\mathbf{Wz}^\intercal + \mathbf{b}) \quad (39)$$

where $\mathbf{y} \in \mathbb{R}^{O\times 1}$, $\mathbf{W} \in \mathbb{R}^{O\times H}$, $\mathbf{b} \in \mathbb{R}^{O\times 1}$ and O is the number of neurons in the output layer.

C. Parameter optimization

Our model uses the optimizer $Adam$ to optimize the parameters. The loss function $Loss(\cdot,\cdot)$ of the model uses the cross-entropy loss function. Since $\mathcal{R}(\cdot)$ Eq. 30 is added to the pure state in the hidden layer, the final objective function is

$$\min \frac{1}{B}\sum_{i=1}^{B} Loss(\mathbf{y}_i, \hat{\mathbf{y}}_i) + \lambda \sum_{i=1}^{H} \mathcal{R}(|\Phi_i\rangle) \quad (40)$$

where λ is the coefficient of the regularization term and B is the number of training samples in each batch.

IV. Experiments

A. Compared with the classic network model

We chose two commonly used datasets, namely, Fashion-MNIST [27] and Cifar 10 [28], to verify QNN. The statistics of the datasets are shown in Tab. I.

TABLE I
DATASET STATISTICS: THE NUMBER OF TRAINING SET, EVALUATION SET AND TEST SET IN FASHION MNIST (F-MNIST) AND CIFAR 10. AND THE NUMBER OF CLASSES.

Dataset	Training	Evaluation	Test	Classes
Cifar 10	40000	10000	10000	10
F-MNIST	48000	12000	10000	10

The experimental process is set as follows:

- according to the different datasets, build different classic network models, and perform parameter tuning on the datasets to obtain the experimental results of the Control Group (CG);
- replace the neurons in the hidden layer of the classic model with neurons reconstructed from the entangled state, and also perform parameter tuning on the corresponding dataset to obtain the experimental results of the Experimental Group (EG). Since the neuron reconstructed from the entangled state requires 4 times the number of parameters compared to the classical neuron, every 4 classical neuron is replaced with a neuron reconstructed from the entangled state.

CG: The model under Fashion MNIST is a five-layer structure, which is an input layer of 28×28 neurons, two fully connected hidden layers of 64 neurons, a fully connected hidden layer of 32 neurons, and a fully connected output layer of 10 neurons. The activation function of the hidden layer is $Relu$, and the output layer does not use the activation function. The model uses the optimizer $Adam$, and $learning\ rate$, $batch\ size$ and $epochs$ are 0.001, 32 and 200, respectively.

The model under Cifar 10 and the model under Fashion MNIST have the same basic structure, that is, the above five-layer structure. The difference is that the model under Cifar 10 adds three convolutional layers on top of the basic structure, which are one 32-kernel and two 64-kernel convolutional layers. The size of the convolution kernel is 3×3, and the activation function is $Relu$. Moreover, two 2×2 pooling layers

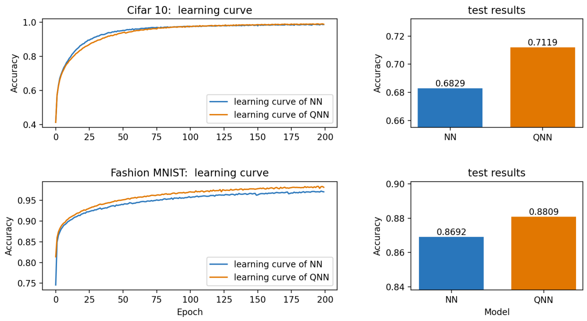

Fig. 2. The accuracy curve under the evaluation set and the test results under the test set on the datasets Cifar 10 and Fashion MNIST.

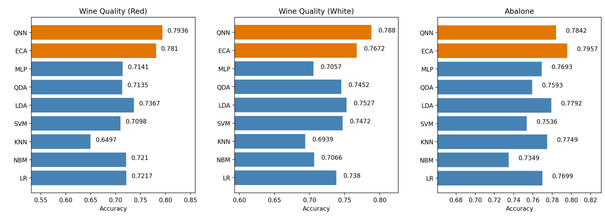

Fig. 3. The accuracy of the test set of each model under the datasets Wine Quality (Red), Wine Quality (White) and Abalone.

are added between the convolutional layers. The optimizer and hyper-parameters are the same as the model under Fashion MNIST.

EG: We use the neurons reconstructed from the entangled state to replace the classic neurons in the hidden layer of the basic structure at a ratio of 1:4 to construct the basic structure of EG. On the dataset Cifar 10, the convolutional layers of EG and CG are consistent. The optimizer and hyper-parameters are the same as those in CG. The coefficient λ of the regularization term is 0.01.

The experimental results of each model are shown in Fig. 2. From the experimental results, QNN is indeed better than the classical model.

TABLE II
DATASET STATISTICS: THE NUMBER OF SAMPLES AND ATTRIBUTES CONTAINED IN EACH DATASET IS SHOWN IN THE TABLE, AND THE NUMBER OF POSITIVE AND NEGATIVE SAMPLES IS SHOWN IN PARENTHESES.

Dataset	Samples	Attributes
Abalone	4177 (2096+2081)	8
Wine Quality (Red)	1599 (855 +744)	11
Wine Quality (White)	4898 (3258+1640)	11

B. Compared with the model inspired by QE

Zhang et al. [17] creatively introduced the measurement process of entangled states into machine learning, and constructed a classification algorithm inspired by QE, called ECA. ECA only uses the measurement process of the entangled state in the output layer, where the entangled state is given and only the measurement operator is learned. Our model extends the measurement process of the entangled state to all neurons in the hidden layer, and both the entangled state and the measurement operator are learned. In this section, we will compare with ECA to examine the pros and cons of QNN. The datasets come from UCI [29], and their statistical information is shown in Tab. II. In order to adapt to ECA (ECA is only applicable to two classification tasks), we adjust the datasets to two categories according to Ref. [17].

The model structure and parameters of ECA use the settings published in Ref. [17]. To make QNN and ECA as similar as possible in the model structure, we add a fully connected layer to QNN, that is, the first layer is the input layer, the second layer is the fully connected layer, the third layer is the hidden layer of neurons reconstructed from entangled states, and the last layer is the output layer. The number of neurons in each layer is N, 32, 10, and 2, respectively, where N represents the number of attributes. The activation function of the fully connected layer is $Relu$. The model uses the optimizer $Adam$, and $learning\ rate$, $batch\ size$ and $epochs$ are 0.01, 10 and 100, respectively. The coefficient λ of the regularization term is 0.001. Similar to Ref. [17], we also use the 5-fold cross-validation method to divide the training set and the test set, and obtain the experimental results.

The experimental results are shown in Fig. 3. Except for QNN, other data are taken from the results published in Ref. [17]. From the results, QNN has achieved good performance on Wine Quality (Red) and Wine Quality (White); QNN is better than the classic model but it is inferior to ECA on Abalone.

On the above three datasets, we also compared with the classical NN model under the same structure. Other hyper-parameters are the same as Exp. IV-A. The experimental results are shown in Fig. 4.

C. Quantitative analysis

The entangled state ab used in this article describes both classical and non-classical correlations, i.e., quantum correlation, so we need to clearly understand the contribution of Non-classical Correlation ($\mathcal{NCC}$) in the calculation process,

$$\mathcal{NCC} = |\mathcal{CF}_{ab}^{quantum}| - |\mathcal{CF}_{ab}^{classical}|. \quad (41)$$

In this section, we first analyze the number of entangled states input to the neuron based on concurrence, that is, to clarify the effect of the regularizer Eq. 30, and then quantitatively analyze the proportion of non-classical correlations based on the quantitative method Eq. 41.

By analyzing the data of Experiment IV-A, as shown in Fig. 5 (subgraphs on upper left and lower left), we can see that through the constraints of the regularizer, more than 99% of the pure states are trained as entangled states, i.e., $\mathcal{C} > 0$, which means that QNN is using entangled states to build the relationship between features, that is, the non-classical correlation is used. Regarding the question of why it is not 100%: because there may be uncorrelated phenomena between

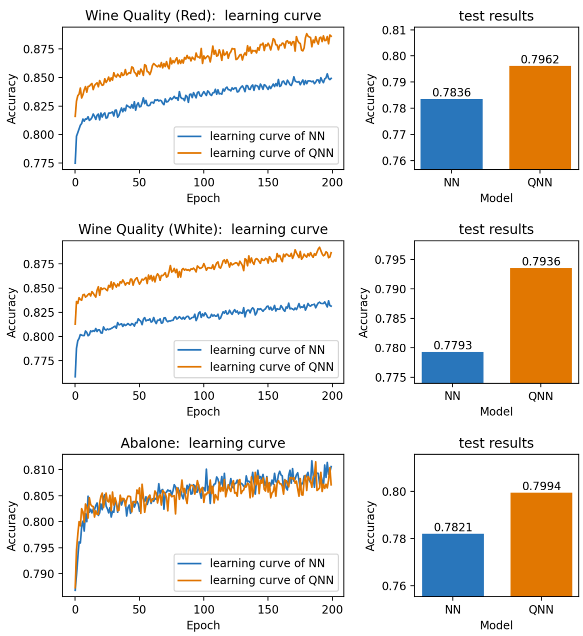

Fig. 4. The accuracy curve under the training set and the test results under the test set on the datasets Wine Quality (Red), Wine Qality (White) and Abalone. The classic network model and the network model reconstructed from the entangled state are abbreviated as NN and QNN respectively.

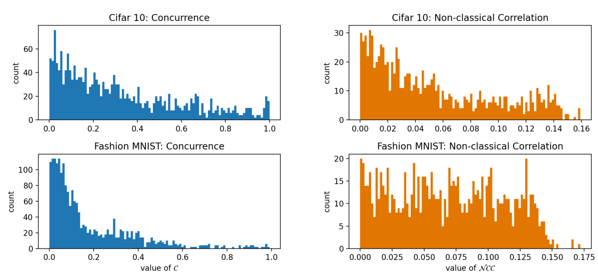

Fig. 5. Here we randomly select a neuron for analysis. The result of the illustration is to take 1000 test data and divide them into 100 groups for histogram statistics.

features; or because the computer may regard extremely small numbers as 0.

We conduct a quantitative analysis of non-classical correlation in the entangled state. As can be seen from Fig. 5 (subgraph on upper right), under the dataset Cifar 10, the contribution of non-classical correlation is concentrated at a lower level. The possible reason for this result is that the model adds three convolutional layers. The convolutional layers smooth the features effectively and the importance of the features is basically at an equal level, which ultimately leads to the contribution of non-classical correlation at a lower level. As can be seen from Fig. 5 (subgraph on lower right), under the dataset Fashion MNIST , the contribution of non-classical correlation is significant, and contributions at various levels exist, indicating that entangled states efficiently model the relationship between features through non-classical correlation.

V. Conclusion

Humans can establish inherent and abstract connections between features and make decisions based on a small amount of data, indicating that humans have strong feature extraction capabilities. Considering that the entangled quantum state can reveal the strong statistical correlation between subsystems, that is, the classical correlation and the non-classical correlation, this paper leverages the measurement process of the entangled state to reconstruct the neurons of NNs. Specifically, based on concurrence, a quantification method of entanglement, we propose a regularizer that can constrain a state vector to an entangled state, and apply it to the optimization process to ensure that the vector passed to the neuron is a legal entangled state. Finally, the entangled state is measured to obtain the output of the neuron. Experimental results show that our model is better than the baseline. Moreover, it performs well compared to the model inspired by quantum entanglement. The contribution of this paper is to inject non-classical correlations into neural networks.

Due to space constraints, we have not been able to build more models based on the reconstructed neuron to extensively verify the effectiveness and universality of the neuron. In future work, we will continue to research this area.

References

[1] S. Wang, "Artificial neural network," in *Interdisciplinary computing in java programming*. Springer, 2003, pp. 81–100.

[2] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, and F. E. Alsaadi, "A survey of deep neural network architectures and their applications," *Neurocomputing*, vol. 234, pp. 11–26, 2017.

[3] C. M. Bishop *et al.*, *Neural networks for pattern recognition*. Oxford university press, 1995.

[4] B. Sendhoff, E. Körner, O. Sporns, H. Ritter, and K. Doya, *Creating brain-like intelligence: from basic principles to complex intelligent systems*. Springer, 2009, vol. 5436.

[5] R. Kozma, C. Alippi, Y. Choe, and F. C. Morabito, *Artificial Intelligence in the Age of Neural networks and Brain computing*. Academic Press, 2018.

[6] R. Mitchell, J. Michalski, and T. Carbonell, *An artificial intelligence approach*. Springer, 2013.

[7] P. Rebentrost, T. R. Bromley, C. Weedbrook, and S. Lloyd, "Quantum hopfield neural network," *Physical Review A*, vol. 98, no. 4, p. 042308, 2018.

[8] J. Chen, L. Wang, and E. Charbon, "A quantum-implementable neural network model," *Quantum Information Processing*, vol. 16, no. 10, p. 245, 2017.

[9] G. Verdon, T. McCourt, E. Luzhnica, V. Singh, S. Leichenauer, and J. Hidary, "Quantum graph neural networks," *arXiv preprint arXiv:1909.12264*, 2019.

[10] I. Cong, S. Choi, and M. D. Lukin, "Quantum convolutional neural networks," *Nature Physics*, vol. 15, no. 12, pp. 1273–1278, 2019.

[11] A. Krenker, J. Bešter, and A. Kos, "Introduction to the artificial neural networks," *Artificial Neural Networks: Methodological Advances and Biomedical Applications. InTech*, pp. 1–18, 2011.

[12] D. C. Marinescu, *Classical and quantum information*. Academic Press, 2011.

[13] A. A. Grib and W. A. Rodrigues Jr, *Nonlocality in quantum physics*. Springer Science & Business Media, 2012.

[14] W. K. Wootters, "Entanglement of formation and concurrence." *Quantum Inf. Comput.*, vol. 1, no. 1, pp. 27–44, 2001.

[15] P. Wittek, *Quantum machine learning: what quantum computing means to data mining*. Academic Press, 2014.

[16] L. Hardy, "Quantum theory from five reasonable axioms," *arXiv preprint quant-ph/0101012*, 2001.

[17] J. Zhang, Y. Hou, Z. Li, L. Zhang, and X. Chen, "Strong statistical correlation revealed by quantum entanglement for supervised learning," ser. Frontiers in Artificial Intelligence and Applications, vol. 325. IOS Press, 2020, pp. 1650–1657.
[18] M. S. Bartlett, "Negative probability," in *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 41, no. 1. Cambridge University Press, 1945, pp. 71–73.
[19] V. Rojansky *et al.*, *Introductory quantum mechanics*. Prentice-Hall New York, 1938.
[20] N. J. Nilsson, *Principles of artificial intelligence*. Morgan Kaufmann, 2014.
[21] M. A. Nielsen and I. Chuang, *Quantum computation and quantum information*. American Association of Physics Teachers, 2002.
[22] W. K. Wootters, "Entanglement of formation of an arbitrary state of two qubits," *Physical Review Letters*, vol. 80, no. 10, p. 2245, 1998.
[23] A. F. Abouraddy, B. E. Saleh, A. V. Sergienko, and M. C. Teich, "Degree of entanglement for two qubits," *Physical Review A*, vol. 64, no. 5, p. 050101, 2001.
[24] L. Henderson and V. Vedral, "Information, relative entropy of entanglement, and irreversibility," *Physical review letters*, vol. 84, no. 10, p. 2263, 2000.
[25] P. M. Hayden, M. Horodecki, and B. M. Terhal, "The asymptotic entanglement cost of preparing a quantum state," *Journal of Physics A: Mathematical and General*, vol. 34, no. 35, p. 6891, 2001.
[26] N. J. Nilsson, "Principle of artificial intelligence," *Intelligent Systems IEEE*, vol. 29, no. 2, pp. 2 – 4, 1982.
[27] H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms," *arXiv preprint arXiv:1708.07747*, 2017.
[28] A. Krizhevsky, G. Hinton *et al.*, "Learning multiple layers of features from tiny images," 2009.
[29] D. Dua and C. Graff, "UCI machine learning repository," 2017. [Online]. Available: http://archive.ics.uci.edu/ml