Quantum Correlation Revealed by Bell State for Classification Tasks

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

Ruifang He(✉)
College of Intelligence and Computing,
Tianjin University, Tianjin, China
Email: rfhe@tju.edu.cn

Zhao Li(✉)
Search Division,
Alibaba Group, Hangzhou, China
Email: lizhao.lz@alibaba-inc.com

Ji Zhang(✉)
School of Sciences,
The University of Southern
Queensland, Australia
Ji.Zhang@usq.edu.au

Biao Wang
Zhejiang Lab,
Hangzhou,
Zhejiang, China
wangbiao@zhejianglab.com

Zhaolin Li
Tianjin Xiniu Huaan
Technology Co., Ltd,
Tianjin, China
15900208166@163.com

Tianyuan Niu
College of Intelligence and
Computing, Tianjin University,
Tianjin, China
niutianyuan@qq.com

2021 International Joint Conference on Neural Networks (IJCNN) | 978-1-6654-3900-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/IJCNN52387.2021.9534416

***Abstract*—In machine learning, classification algorithms often use statistical methods to build the correspondence between features (or attributes) and categories (or labels), that is, the statistical correlation between features and categories. In quantum theory, a large number of experimental results show that quantum correlation is far stronger than what can be explained by local hidden theory (i.e., classical or non-quantum theory), that is, quantum mechanics theory reveals a statistical correlation stronger than that described by classical theory. Based on this, this paper will use the strong statistical correlation revealed by Bell state to build a classification algorithm to verify the validity and superiority of the formal framework of quantum mechanics in specific classification tasks. Specifically, we use quantum joint probabilities derived from the measurement process of Bell state to model the quantum statistical correlation between features and categories. The paper first theoretically proves that the formal framework used has the ability to violate Bell inequality; moreover, a classification algorithm is implemented and verified on classic machine learning datasets. Experimental results show that the algorithm is significantly better than most mainstream machine learning algorithms.**

***Index Terms*—Quantum Machine Learning, Quantum Algorithm, Bell State, Quantum Correlation.**

I. Introduction

In machine learning, classification algorithms are an important research area, not only because of the importance of classification algorithms in engineering and applications, but also because classification algorithms are an effective test field for testing new frameworks or models. Classification algorithms often use (classical) statistical methods to build the correspondence between the features (attributes) and categories (labels) of instances (samples), that is, to establish the statistical correlation between features and categories [1]–[3].

In quantum theory, physicists have done a lot of experiments to verify the correctness of quantum theory [4]–[6]. The experimental results show that the local entity theory [7] advocated by Albert Einstein does not match the experimental results, but the predictions of quantum mechanics are consistent. It is pointed out that Quantum Correlation (QC) is far stronger than what can be explained by local hidden theory (i.e., classical or non-quantum theory), that is, quantum mechanics theory reveals a statistical correlation stronger than that described by classical theory [8]. This strong statistical correlation revealed by quantum mechanics theory is called entanglement [9].

Entanglement, or called Quantum Entanglement (QE), is an important quantum resource for quantum computing and quantum information processing, and thus has become a very hot research topic in quantum informatics [10]–[12]. QE is a correlation that is strikingly different from the correlation in classical information theory. Moreover, entanglement has been successfully applied to many fields as an important physical resource, such as quantum teleportation [13]–[15], quantum cryptography [16]–[18], quantum algorithms [19], etc.

Although quantum mechanics is generally regarded as a microphysical theory, its connotation is about information rather than physics. Since Hardy [20], the informational nature of quantum mechanics has gradually become more and more rigorous. Therefore, the laws of quantum mechanics should not only be regarded as the laws of the microphysical world, but should be regarded as the general rules of information processing [21]–[23]. Based on this, in this paper, we will use the formal framework of quantum mechanics to build a classification algorithm to learn the statistical correlation between the features and categories of instances. What needs to be emphasized here is that the statistical correlation is a strong statistical correlation (i.e., quantum correlation), but it does not mean that there is a quantum phenomenon (or quantum effect) between features and categories but rather a closer relationship between features and categories. This relationship cannot or cannot be fully (or effectively) established by classical methods or theories, but it cannot be equated with a related relationship in the microphysical world.

In quantum mechanics, QC occurs during the measurement process of entangled states (systems), so the core of the classification algorithm is to reproduce the measurement process of entangled states. Since the Bell state [24] is the entangled

978-0-7381-3366-9/21/$31.00 ©2021 IEEE

Authorized licensed use limited to: TIANJIN UNIVERSITY. Downloaded on November 09,2022 at 17:10:21 UTC from IEEE Xplore. Restrictions apply.

state with the largest entanglement degree under two qubits, we use the Bell state as the measured entangled state in order to be able to establish a stronger connection between the subsystems. Quantum mechanics incorporates complex numbers into its formal system, and complex numbers inevitably lead to the generation of phases. Phase is an important factor in describing QC, so we choose a phase matrix (operator) to describe the observations of the subsystem.

In this paper, we use a fully connected network layer to learn the parameters of observations of the subsystem, and use weighted sums to integrate the measured probability values of each entangled state. In short, it can be understood that the hidden layer nodes of the Multi-Layer Perceptron (MLP) are replaced with a measurement process. We validate the classification algorithm on classic machine learning datasets, and the experimental results show that the classification algorithm is superior to most mainstream classification algorithms.

The contribution of this paper is to implement a classification algorithm using the formal framework of quantum mechanics, theoretically prove that the formal framework used has the ability to violate Bell inequality, that is, it can describe strong statistical correlation (i.e., quantum correlation), and experimentally verify that the formal framework is effectiveness and superiority in specific classification tasks.

The paper is organized as follows: the related work on QC and QE is given in Sec. II; we analyze the formal framework used and verify it by Bell inequality in Sec. III; we use the formal framework to implement a classification algorithm to verify the effectiveness of the formal framework in specific tasks in Sec. IV; we experimentally verify the classification algorithm in Sec. V; the last section gives a summary and outlook.

II. Related Work

There have been many excellent achievements in the work of quantum machine learning, but the work of combining QC (or QE) and machine learning is not much known by the author. In this part, we mainly introduce the work related to QC and machine learning.

QC is an important physical resource, like time, energy, momentum, etc., and it can extract and transform. QC is gradually being applied to quantum informatics to exert its unique advantages. Among them, the entanglement characteristics of quantum many-body are studied and applied most deeply and extensively. Carleo et al. [25] proposed a variational representation of quantum states based on artificial neural networks with a variable number of hidden neurons, which effectively captures the complexity of one- and two-dimensional entangled many-body systems. Although the restricted boltzmann machine used is simple, this approach achieves high accuracy in describing prototypical interacting spins models in one- and two-dimensions. Deng et al. [26] explored the data structures that encode the physical features in the network states by studying the entanglement properties with a focus on the restricted boltzmann machine architecture. They prove that the entanglement entropy of all short-range restricted boltzmann machine states satisfies an area law for arbitrary dimensions and bipartition geometry. Van Nieuwenburg et al. [27] proposed a neural-network approach to finding phase transitions based on the performance of a neural network after it is trained with data that are deliberately labeled incorrectly. See also Refs. [28]–[30]

In the theoretical study of QC and QE: Vedral et al. [31] presented conditions every measure of entanglement has to satisfy, and constructed a whole class of entanglement measures. Moreover, Vidal et al. [32] presented a measure of entanglement that can be computed effectively for any mixed state of an arbitrary bipartite system. They showed that it does not increase under local manipulations of the system, and used it to obtain a bound on the teleportation capacity and on the distillable entanglement of mixed states. Kitaev et al. [33] formulated a universal characterization of the many-particle quantum entanglement in the ground state of a topologically ordered two-dimensional medium with a mass gap. See also Refs. [34]–[38]

In short, the combination of machine learning and QC (or QE) is still an emerging field in academia and industry, and more research resources are needed.

III. Theoretical Analysis and Verification by Bell Inequality

Compared with classical theory, quantum mechanics theory has its own unique advantages. For example, it can break through the upper bound that classical theory can reach. Entanglement is such a feature described by quantum mechanics [8]. From this we can get a simple understanding that it is more valuable to use a quantum mechanics framework that violates the classical theoretical framework to construct a computing model. The theoretical tool for verifying whether the quantum mechanical framework can violate the classical theoretical framework is the Bell inequality [24].

In the following sections we will theoretically analyze the quantum mechanics framework used in this paper and verify it using Bell inequality. Since entanglement occurs during the measurement process of entangled states, we will elaborate from the definition and measurement of entangled states in order.

A. Define Entangled States

Definition 1: Suppose a composite system is composed of two subsystems A and B, and the Hilbert spaces of these two subsystems are $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively. The Hilbert space of the composite system $\mathcal{H}_{AB}$ is

$$\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B \tag{1}$$

where $\otimes$ represents the tensor product. Let the quantum states of A and B be $|\psi\rangle_A$[1] and $|\psi\rangle_B$, respectively. If the quantum state of a composite system $|\psi\rangle_{AB}$ cannot be written as a tensor product

$$|\psi\rangle_{AB} \neq |\psi\rangle_A \otimes |\psi\rangle_B, \tag{2}$$

[1] $|\cdot\rangle$ is a Dirac symbol often used in quantum mechanics, which represents a vector, called Ket; $\langle\cdot|$ is a dual vector of $|\cdot\rangle$, called Bra.

Authorized licensed use limited to: TIANJIN UNIVERSITY. Downloaded on November 09,2022 at 17:10:21 UTC from IEEE Xplore. Restrictions apply.

then the composite system is called an entanglement system of A and B, and the two subsystems are entangled with each other.

In the same way, we can get the definition of entangled state of multiple subsystems, namely, the composite system cannot perform tensor decomposition. Below we give some quantum entangled states that will be used in this paper or often mentioned in other treatise:

Example 1: The largest entangled state used to describe a two-qubit system is often referred to as the Bell state. It is named after the Irish physicist, the famous Bell Inequality proponent John Stewart Bell [24]. Its specific form is:

$$|\Phi^{\pm}\rangle = \frac{1}{\sqrt{2}}(|0\rangle_A \otimes |0\rangle_B \pm |1\rangle_A \otimes |1\rangle_B) \tag{3}$$

$$|\Psi^{\pm}\rangle = \frac{1}{\sqrt{2}}(|0\rangle_A \otimes |1\rangle_B \pm |1\rangle_A \otimes |0\rangle_B). \tag{4}$$

In order to prove the generality of the quantum mechanical framework we will use, we define an entangled state with an arbitrary degree of entanglement in the case of two qubits. Let the entangled state be

$$|\Psi\rangle = \alpha|00\rangle + \beta|11\rangle, \tag{5}$$

where $\alpha = e^{i\eta}\sin(\xi)$ and $\beta = e^{-i\eta}\cos(\xi)$, and $|\alpha|^2 + |\beta|^2 = 1$. i represents the imaginary number symbol, i.e., $i^2 = -1$. $\eta \in \mathbb{R}$ and $\xi \in \mathbb{R}$ represent arbitrary real-valued parameters. The density matrix form of this entangled state is defined as

$$\rho = |\Psi\rangle\langle\Psi|. \tag{6}$$

B. Measuring Entangled States

In the field of quantum computing and quantum information, projection measurement is widely used. Projection measurement is described by a observation, Hermite operator M, on the state space of the observed system. This observation has a spectral decomposition,

$$M = \sum_m m P_m, \tag{7}$$

where P_m is a projection onto the eigenspace M of the eigenvalue m. A possible measurement result is the eigenvalue m corresponding to the measurement operator P_m. When the state Ψ is measured, the probability that the result m is obtained is

$$\mathcal{P}(m) = \langle\Psi|P_m|\Psi\rangle. \tag{8}$$

After the measurement result m is given, the state of the quantum system after the measurement, $|\Psi'\rangle$, is immediately modified to

$$|\Psi'\rangle = \frac{P_m|\Psi\rangle}{\sqrt{\mathcal{P}(m)}}. \tag{9}$$

In this paper, the observations of the subsystems of the entangled system are defined as

$$M_r = \sigma_x \cos(\phi_r) + \sigma_y \sin(\phi_r) \tag{10}$$

$$= \begin{bmatrix} 0 & e^{-i\phi_r} \\ e^{i\phi_r} & 0 \end{bmatrix} \tag{11}$$

where r represents the corresponding subsystem, say A and B, then $r \in \{A, B\}$. $\phi_r \in \mathbb{R}$ represents arbitrary real-valued parameter. σ_x and σ_y represent the Pauli matrix. The Pauli matrix is four commonly used matrices, i.e.,

$$\sigma_x \equiv X \equiv \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \qquad \sigma_y \equiv Y \equiv \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \tag{12}$$

$$\sigma_z \equiv Z \equiv \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \qquad I \equiv \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{13}$$

The main motivation for using M_r here is that it only describes phase information of the subsystems. Phase is an inevitable result of the introduction of complex numbers in quantum mechanics, and this operation has led to many unique phenomena of quantum mechanics, such as entanglement. Therefore, this article will only focus on the information processing capabilities of phase.

By applying the observation of the subsystem to the entangled state, we can obtain the observable probability distribution of the entangled system,

$$\mathcal{P}(A, B) = Tr[(M_A \otimes M_B)\rho] \tag{14}$$

$$= \langle\Psi|(M_A \otimes M_B)|\Psi\rangle \tag{15}$$

$$= \sin(\xi)\cos(\xi)\cos(\phi_A + \phi_B + 2\eta) \tag{16}$$

which is the joint probability of the subsystems A and B. Tr represents a matrix function, which is the trace of a matrix.

C. Verified by Bell Inequality

The Bell inequality is a powerful mathematical inequality proposed by Bell in 1964 to verify the correctness of quantum mechanical theory [24]. Many experimental results based on the Bell inequality are consistent with the predictions of quantum mechanics theory, indicating that quantum correlations, such as quantum entanglement, are far stronger than the explanation provided by the local hidden variable theory.

The Bell inequality is also used as an important tool to verify whether a formalized framework has the ability to describe non-classical (quantum) correlations. Violating the Bell inequality can prove that the formalized framework has the ability to describe quantum correlation. The Wigner inequality [39] is an important Bell inequality, and its mathematical form is

$$Pr(X, Y) \leq Pr(X, Z) + Pr(Z, Y), \tag{17}$$

where Pr represents a symbol of probability and X, Y and Z represent observations. To prove convenience, we shift the inequality, that is,

$$Pr(X, Y) - Pr(X, Z) - Pr(Z, Y) \leq 0. \tag{18}$$

For the problem of verifying whether the inequality can be broken, just give a counter example. For Wigner inequality, let the entangled state $|\Psi\rangle$ is the maximum entangled state, $|\Psi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$, i.e., $\xi = \frac{\pi}{4}$, $\eta = 0$, and $\phi_X = \phi_Y = 0$ and $\phi_Z = \frac{\pi}{2}$, then $\mathcal{P}(X, Y) = 1$, $\mathcal{P}(X, Z) = 0$ and $\mathcal{P}(Z, Y) = 0$, i.e.,

$$\mathcal{P}(X, Y) - \mathcal{P}(X, Z) - \mathcal{P}(Z, Y) = 1 \nleq 0. \tag{19}$$

Authorized licensed use limited to: TIANJIN UNIVERSITY. Downloaded on November 09,2022 at 17:10:21 UTC from IEEE Xplore. Restrictions apply.

From the proof that violates the Wigner inequality, we can confirm that the formal framework has the ability to reproduce quantum correlation, that is, quantum entanglement. In the third section, we will build a classification model based on the formal framework to verify the validity and superiority of the formal framework on classic machine learning datasets.

D. Analysis

From the verification of the Bell inequality, we can get a consensus that this formal framework can model the strong statistical correlation of entangled states. But here we also need to be clear: whether the quantum effect appeared in the process of using the formal framework and how to verify it is still a difficulty in the physical world, we can only give a rough judgment from the perspective of experimental effects.

It can be seen from Eq. (16) that when using the quantum joint probability derived from quantum entanglement, we do not necessarily have to make all parameters adjustable. In order to reduce the difficulty of learning, we use an entangled state with a fixed maximum entanglement degree, that is, the Bell state. That is to say the parameters $\xi = \frac{\pi}{4}$ and $\eta = 0$.

IV. Classification Algorithm with QC

In order to verify the effectiveness and superiority of the formal framework of quantum mechanics in specific tasks, we use this framework to implement a classification algorithm to facilitate verification on classic machine learning datasets. This algorithm is called a classification algorithm inspired by QC (or QE), and is abbreviated as QCCA.

The structure of this section is as follows: we first describe the construction method of the entanglement relationship between each feature and the label, that is, a measurement process of the two qubit entangled state, then describe how to integrate the entanglement relationship between all features and the label, that is, integrated into a complete classification algorithm; finally, the parameter learning process of the entire algorithm is given.

A. Measurement Process of Entangled State

In this paper, we use the Bell state to describe the strong statistical correlation between each feature and the label, that is, one subsystem of the Bell state describes the feature and the other describes the label. It can be seen that if there are N features, N Bell states are needed. Here we choose the Bell state as the entangled state because: 1. The Bell state is the entangled state with the maximum entanglement in the case of the two qubits; 2. From the analysis of the theoretical part, it can be seen that the use of entangled states with fixed parameters can reduce the difficulty of learning, but it will not have much impact on the results. As shown in Eq. (3), the entangled state between the feature F and the label L is denoted as:

$$|\Phi\rangle = \frac{1}{\sqrt{2}}(|0\rangle_F \otimes |0\rangle_L + |1\rangle_F \otimes |1\rangle_L) \tag{20}$$

$$= \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle). \tag{21}$$

In quantum mechanics, the observations of a system are described by Hermite operators. As stated in the theory section, phase is the essential cause of QE, so we use the phase operator function to represent the observation of the system. Moreover, since the label can be represented by a specific state, its parameter is set to a fixed value, i.e., $\phi = 0$. As shown in Eq. (10), the observation of the k^{-th} feature is defined as:

$$M_F^k(\phi^k) = \sigma_x \cos(\phi^k) + \sigma_y \sin(\phi^k) \tag{22}$$

$$= \begin{bmatrix} 0 & e^{-i\phi^k} \\ e^{i\phi^k} & 0 \end{bmatrix}, \tag{23}$$

and the observation of the label is defined as:

$$M_L = \sigma_x \cos(\phi = 0) + \sigma_y \sin(\phi = 0) \tag{24}$$

$$= \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \tag{25}$$

From the definition of projection measurement, the measurement result is the eigenvalue of the observation, and the measurement operator is the eigenstate of the observation. So the measurement operators of the k^{-th} feature are

$$M_F^k(\phi^k) = \Pi_F^{k+}(\phi^k) - \Pi_F^{k-}(\phi^k) \tag{26}$$

$$= \frac{1}{2}\begin{bmatrix} 1 & e^{-i\phi^k} \\ e^{i\phi^k} & 1 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} 1 & -e^{-i\phi^k} \\ -e^{i\phi^k} & 1 \end{bmatrix} \tag{27}$$

and the measurement operators of the label are

$$M_L = \Pi_L^+ - \Pi_L^- \tag{28}$$

$$= \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} - \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix} \tag{29}$$

From this we can see that for a 2-dimensional qubit, there are at most two eigenvalues, that is, it can only be used for the two-class (binary) classification problem. Of course, we can also implement the multi-class classification problem by adding auxiliary qubits, but this is not the main focus of this paper.

Now we can construct the entangled state measurement operator by combining the measurement operators of the subsystems, that is, the measurement operators of the feature and the measurement operators of the label. The measurement operator of the positive example (positive class) is defined as:

$$\Pi^{k+}(\phi^k) = \Pi_F^{k+}(\phi^k) \otimes \Pi_L^+, \tag{30}$$

and the measurement operator of the negative example (negative class) is defined as:

$$\Pi^{k-}(\phi^k) = \Pi_F^{k+}(\phi^k) \otimes \Pi_L^-. \tag{31}$$

Of course, we can also use $\Pi^{k+}(\phi^k) = \Pi_F^{k-}(\phi^k) \otimes \Pi_L^+$ to represent the measurement operator of the positive example, and $\Pi^{k-}(\phi^k) = \Pi_F^{k-}(\phi^k) \otimes \Pi_L^-$ to represent the measurement operator of the negative example. Their effects are the same.

Authorized licensed use limited to: TIANJIN UNIVERSITY. Downloaded on November 09,2022 at 17:10:21 UTC from IEEE Xplore. Restrictions apply.

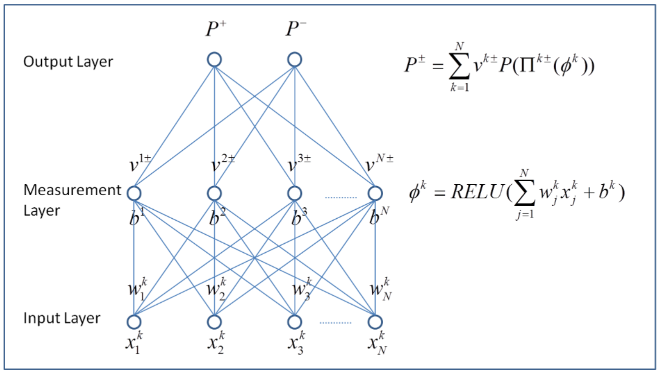

Fig. 1. Schematic diagram of QCCA: Each node in the measurement layer represents a measured process of an entangled state.

So the probability obtained after measuring the entangled state, i.e., $|\Phi\rangle$, can be defined as

$$P(\Pi^{k+}(\phi^k)) = \langle\Phi|\Pi^{k+}(\phi^k)|\Phi\rangle \tag{32}$$
$$= \langle\Phi|(\Pi_F^{k+}(\phi^k) \otimes \Pi_L^{+})|\Phi\rangle \tag{33}$$

and

$$P(\Pi^{k-}(\phi^k)) = \langle\Phi|\Pi^{k-}(\phi^k)|\Phi\rangle \tag{34}$$
$$= \langle\Phi|(\Pi_F^{k+}(\phi^k) \otimes \Pi_L^{-})|\Phi\rangle. \tag{35}$$

To this end, we can get the probability of the measurement result of the entangled state composed of each feature and the label, that is, the probability value of the positive example and the probability of the negative example. In the following section, we will describe how to integrate the probabilities into the final output probability, and also describe how to represent and learn the parameters of the measurement operators of the features.

B. Design QCCA through Measurement Process

In order to allow readers to have a global understanding of QCCA, we first give a schematic diagram of the structure of QCCA, as shown in Fig. 1. From the diagram we can clearly see that we use the fully connected network layer to learn the parameters of the measurement operator of the feature, that is,

$$\phi^k = ReLU\left(\sum_{j=1}^{N} w_j^k x_j^k + b^k\right), \tag{36}$$

where $w^k \in \mathbb{R}^N$ is the weight vector, $x^k \in \mathbb{R}^N$ is the features of the instance, $b^k \in \mathbb{R}$ is the bias term, N is the number of the features and $ReLU$ (Rectified Linear Unit) [40] is an activation function. Moreover, we use the weighted sum to integrate the measured probability values of the entangled state together as the final output value of QCCA, that is,

$$P^{\pm} = \sum_{k=1}^{N} v^{k\pm} P(\Pi^{k\pm}(\phi^k)), \tag{37}$$

where $v^{k\pm} \in \mathbb{R}^N$ is the weight vector.

TABLE I
DATASET STATISTICS: 'INSTANCES' REFERS TO THE NUMBER OF SAMPLES CONTAINED IN THE DATASET, AND THE NUMBER OF POSITIVE AND NEGATIVE SAMPLES IS INDICATED IN PARENTHESES; 'ATTRIBUTES' REFERS TO THE NUMBER OF ATTRIBUTES OF THE SAMPLE, WHICH DOES NOT CONTAIN THE LABEL OF THE SAMPLE; 'SPECIES' REFERS TO THE NUMBER OF SAMPLE CATEGORY, SUCH AS POSITIVE AND NEGATIVE IN THE TWO-CLASS CLASSIFICATION TASK.

Dataset	Instances	Attributes	Species
Abalone	4177 (2096+2081)	8	2
Wine Quality (Red)	1599 (855 + 744)	11	2
Wine Quality (White)	4898 (3258+1640)	11	2

So far, the model structure of QCCA has been described. In the following sections, we will describe how to learn the parameters in the model.

C. Parameter Learning

In machine learning, the loss function is usually used as the objective function of the model. This paper uses the most basic and most commonly used cross entropy loss function as the loss function of QCCA, and its specific form is

$$H(p,q) = -\sum_{j=1}^{N} p(x_j) log(q(x_j)). \tag{38}$$

Moreover, we use the optimizer Adam [41] to update the parameters of the model. Adam optimizer can be said to be the most widely used optimizer with fast convergence speed and stable convergence process.

In the end, we give the pseudo code of the model training process to facilitate the reader to better understand the model training process, which is Alg. 1.

Algorithm 1 Training Process of QCCA

Input: Training set $\mathcal{T}$
$\mathcal{T} = \{(x^k, y^k)|x^k \in R^N, y^k \in \{[0,1]), [1,0]\}\}_{k=1}^N$.
Output: Trained QCCA
1: Initialise $w^k \in R^N, b^k \in R$ and $v^k \in R^N$, $k \in \{1 \cdots N\}$
2: **repeat**
3: **for** each (x^k, y^k) in $\mathcal{T}$ **do**
4: $\phi^k = ReLU\left(\sum_{j=1}^N w_j^k x_j^k + b^k\right)$;
5: $P^{k+} = \sum_{k=1}^N v^{k+} P(\Pi^{k+}(\phi^k))$;
6: $P^{k-} = \sum_{k=1}^N v^{k-} P(\Pi^{k-}(\phi^k))$;
7: $P^k = [P^{k+}, P^{k-}]$;
8: Calculate the cross entropy of P^k and y^k, i.e. $H(P^k, y^k)$;
9: Use the optimizer to minimize $H(P^k, y^k)$ and update w, b and v;
10: **end for**
11: **until** Epochs

V. EXPERIMENTS

A. Datasets and Evaluation Metrics

Since our experiments are simulated on a classical computer, and limited by the computing power of the classical computer,

Authorized licensed use limited to: TIANJIN UNIVERSITY. Downloaded on November 09,2022 at 17:10:21 UTC from IEEE Xplore. Restrictions apply.

the program can only be verified on a lightweight dataset. Moreover, since the observation of qubits only has at most two eigenvalues, using a qubit to represent a label can only be used to represent the two-class (or binary) classification problem. For the multi-class classification problem, it is possible to increase the number of qubits for the label, but this paper focuses on verifying the validity and superiority of the formal framework of quantum mechanics.

The experiments were conducted on three most frequently used machine learning datasets from UCI [42], that is, Abalone[2], Wine Quality[3] (Red) and Wine Quality[3] (White). The statistics of each dataset are given in Tab. I.

Abalone is a dataset that predicts the age of abalone. We divide the age less than 10 into one class, and the greater than or equal to 10 into another class.

Wine Quality is a dataset that scores on wine quality. Each wine has a score of 0 to 10. We divide the scores less than or equal to 5 into one class, and the scores greater than 5 into another class. The reason for dividing the dataset in this way is that the number of samples in two classes can be made as close as possible.

All experiments use the 5-fold cross-validation method to divide the training set and the test set. The experimental evaluation metrics, that is, the *F1-Score*, the *Accuracy* and the *AUC (Area Under Curve)*, are taken as the average of the 5 results.

B. Compare with Classical Classification Algorithms

1) Baselines: We conduct a comprehensive comparison across a wide range of mainstream machine learning classification algorithms, including Naive Bayesian Model (NBM), Logistic Regressive (LR), Decision Tree (DT), K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA) and Multi-Layer Perceptron (MLP).

2) Parameter Settings: QCCA has four hyper-parameters, which are *learning rate*, *mini-batch*, *training epochs* and *initial weight*. The same parameter settings are used on all datasets, which are: *learning rate* is 0.001, *mini-batch* is 1, *training epochs* is 500 and *initial weight* is 0.01, except for the different *learning rate* under Abalone dataset, which is: *learning rate* is 0.0001. We compare the experimental results to determine the hyper-parameters of the model. Moreover, in order to evaluate the model as objectively as possible, we choose a fixed *initial weight*, which is set to 0.01; we also uniformly use the minimum batch (i.e., *mini-batch*) setting, which is set to 1. The hyper-parameters in the baselines are set to: In MLP, *activation* is *relu* , *solver* is *adam* and *alpha* is 0.0001; In LDA, *solver* is *svd*; In SVM, *C* is 1.0 and *kernel* is *rbf*; In KNN, *n-neighbors* is 5; In LR, *penalty* is *L2*. Other parameters not listed use the default value of the framework Scikit-Learn[4].

3) Experiment Results: The experiment results are shown in Tab. II. The best-performed values of each evaluation metric in each dataset are in bold; moreover, we also give the improvement ratio compared to the standard MLP. The experiment results show that QCCA has obvious advantages over the mainstream machine learning classification algorithms mentioned in the paper. This result can basically prove the effectiveness and superiority of QCCA.

TABLE II
EXPERIMENT RESULTS ON DATASETS ABALONE, WINE QUALITY (RED) AND WINE QUALITY (WHITE). THE BEST-PERFORMED VALUES OF EACH EVALUATION METRIC ARE IN BOLD. AT THE SAME TIME, WE ALSO GIVE THE INCREASE (OR DECREASE) RATIO OF QCCA OVER STANDARD MLP.

Dataset	Algorithm	F1-Score	Accuracy	AUC
Abalone	LR	0.7708	0.7711	0.7711
	DT	0.7091	0.7139	0.7138
	NBM	0.7384	0.7340	0.7340
	KNN	0.7699	0.7699	0.7699
	SVM	0.7669	0.7536	0.7538
	LDA	0.7742	0.7771	0.7770
	QDA	0.7427	0.7593	0.7591
	MLP	0.7684	0.7530	0.7530
	QCCA	**0.7903**	**0.7895**	**0.7895**
	Over MLP	2.85% ↑	4.84% ↑	4.84% ↑
Wine Quality (Red)	LR	0.7460	0.7395	0.7395
	DT	0.7547	0.7385	0.7373
	NBM	0.7396	0.7248	0.7242
	KNN	0.6757	0.6566	0.6556
	SVM	0.7298	0.7098	0.7080
	LDA	0.7504	0.7379	0.7381
	QDA	**0.7561**	0.7266	0.7217
	MLP	0.7481	0.7285	0.7266
	QCCA	0.7478	**0.7448**	**0.7476**
	Over MLP	0.04% ↓	2.23% ↑	2.89% ↑
Wine Quality (White)	LR	0.8230	0.7472	0.6799
	DT	0.8233	0.7445	0.6900
	NBM	0.7795	0.7051	0.6666
	KNN	0.7871	0.7031	0.6428
	SVM	0.8266	0.7486	0.6737
	LDA	0.8225	0.7482	0.6840
	QDA	0.8166	0.7445	0.6900
	MLP	0.8197	0.7435	0.6767
	QCCA	**0.8272**	**0.7617**	**0.7167**
	Over MLP	0.91% ↑	2.44% ↑	5.91% ↑

It also indirectly illustrates the superiority of the formal framework of quantum mechanics in classical classification tasks. We have reason to believe that the strong (non-classical or quantum) statistical correlation revealed by QE plays an important role in this process.

In the following we will analyze the superiority of the model from the specific training process of QCCA.

C. Compare with Standard MLP

QCCA uses a fully connected network layer to learn the parameters of the measurement operator, and also uses the mainstream machine learning optimizer Adam to optimize the parameters in the model. Specifically, we are equivalent to modifying the output layer of the standard MLP to a quantum measurement layer (that is, a measured process of entangled states). Based on this, we have reason to compare with the standard MLP to verify whether the model replaced by the

[2] http://archive.ics.uci.edu/ml/datasets/Abalone
[3] http://archive.ics.uci.edu/ml/datasets/Wine+Quality
[4] https://scikit-learn.org/stable/index.html

Authorized licensed use limited to: TIANJIN UNIVERSITY. Downloaded on November 09,2022 at 17:10:21 UTC from IEEE Xplore. Restrictions apply.

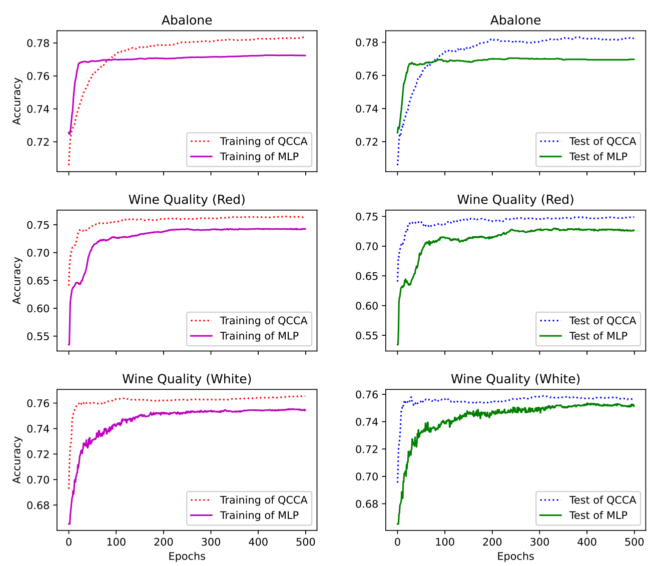

Fig. 2. Experiment results on datasets Abalone, Wine Quality (Red) and Wine Quality (White). The left column is the accuracy curve under the training set, and the right column is the accuracy curve under the test set.

measurement layer (i.e. QCCA) has a significant improvement over the original model (i.e. MLP).

1) Baselines: The model structure of QCCA is the same as described in Sec. III. The MLP is a two-layer network. The number of input nodes is N, where N is the number of attributes of the samples; the number of hidden layer nodes is also N and uses *ReLU* as its excitation function; the output layer has only one node and uses *Sigmoid* as its excitation function. Moreover, like QCCA, the mainstream optimizer, Adam, is used to learn the parameters of the model, and the basic cross-entropy loss function is used as the optimization objective function.

2) Parameter Settings: Both QCCA and MLP have four hyper-parameters, which are *learning rate*, *mini-batch*, *training epochs* and *initial weight*. The same settings are used on all datasets, that is, *learning rate* is 0.001, *mini-batch* is 1, *training epochs* is 500 and *initial weight* is 0.01.

Due to the unsatisfactory learning effect, we adjusted the *learning rate* of QCCA under Abalone dataset to 0.0001, and the *learning rate* of the MLP under Wine Quality (Red) dataset to 0.0001. Note that since we want to ensure the objectiveness and fairness of the experiments as much as possible, we have given a uniform *initial weight* and set the *mini-batch* to 1.

3) Experiment Results: Fig. 2 shows the accuracy curves of QCCA and MLP under different datasets, Abalone, Wine Quality (Red) and Wine Quality (White). For clarity, we present the results of the training set and the test set separately, that is, the left column is the accuracy curve under the training set, and the right column is the accuracy curve under the test set. Note that this test set is the verification set divided by the cross-validation method. In this section, we use it as a measurement set.

It can be clearly seen from the graphical results that QCCA is significantly better than standard MLP in accuracy and can maintain an increase of more than one percentage point. Since the *learning rates* of the models under the two datasets, Abalone and Wine Quality (Red), are different, the convergence speed cannot be compared to each other. However, the *learning rate* and other hyper-parameters of the models under Wine Quality (White) dataset are the same, and it can be seen that the convergence speed of QCCA is significantly better than standard MLP.

It should be noted that the classic model has performed well on these three datasets, and the space that can be improved is not large. QCCA can guarantee an increase of more than one percentage point, which can indicate the superiority of QCCA. It further shows the validity and superiority of the formal framework of quantum mechanics in classical classification tasks.

VI. Conclusions and Future Work

In this paper, we use the formal framework of quantum mechanics to implement a classification algorithm, and analyze the formal framework theoretically and verify the algorithm experimentally. Experimental results show that the formal framework has certain advantages in classical classification tasks. It also indirectly shows that the strong statistical correlation (i.e., QE) can more effectively describe the closer relationship between features and categories in classic classification tasks. Based on Hardy's theoretical work, this paper treats the formal framework of quantum mechanics as a general rule of information processing. We do not assume the existence of quantum phenomena or quantum effects in classical tasks, nor do we make analogies between objects in classical phenomena and objects in non-classical phenomena. From this perspective, the work of this article is groundbreaking and has a guiding significance.

Since we are simulating a quantum algorithm on a classic computer and are limited by the lack of computing power of the classic computer, the algorithm can only be verified on the lightweight datasets. Moreover, because the quantum entanglement is built under the complex field, the current computing framework does not provide an efficient computational support. The above problems will become the focus of my research in the future.

Acknowledgment

This work is funded in part of the Zhejiang Lab (111007-PI2001) and the Zhejiang Provincial Natural Science Foundation (LZ21F030001), the Alibaba Innovation Research Foundation 2017 and the European Unions Horizon 2020 research and innovation programme under the Marie Skodowska-Curie grant agreement No. 721321. Part of the work was performed when Junwei Zhang visited the Alibaba Inc. in 2018.

References

[1] M. James, *Classification algorithms*. Wiley-Interscience, 1932.
[2] J. Gama and P. Brazdil, "Characterization of classification algorithms," in *Portuguese Conference on Artificial Intelligence*. Springer, 1995, pp. 189–200.

Authorized licensed use limited to: TIANJIN UNIVERSITY. Downloaded on November 09,2022 at 17:10:21 UTC from IEEE Xplore. Restrictions apply.

[3] R. Kumar and R. Verma, “Classification algorithms for data mining: A survey,” *International Journal of Innovations in Engineering and Technology (IJIET)*, vol. 1, no. 2, pp. 7–14, 2012.
[4] M. Jammer *et al.*, *The conceptual development of quantum mechanics*. McGraw-Hill New York, 1966.
[5] A. Bohm and M. Loewe, *Quantum mechanics: foundations and applications*. Springer, 1986, vol. 3.
[6] R. Jackiw, *Intermediate quantum mechanics*. CRC Press, 2018.
[7] A. Einstein, B. Podolsky, and N. Rosen, “Can quantum-mechanical description of physical reality be considered complete?” *Physical review*, vol. 47, no. 10, p. 777, 1935.
[8] P. Wittek, *Quantum machine learning: what quantum computing means to data mining*. Academic Press, 2014.
[9] E. Schrödinger, “Discussion of probability relations between separated systems,” in *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 31, no. 4. Cambridge University Press, 1935, pp. 555–563.
[10] M. A. Nielsen and I. Chuang, “Quantum computation and quantum information,” 2002.
[11] C. H. Bennett and D. P. DiVincenzo, “Quantum information and computation,” *Nature*, vol. 404, no. 6775, p. 247, 2000.
[12] H. K. Lo, T. Spiller, and S. Popescu, *Introduction to quantum computation and information*. World Scientific, 1998.
[13] M. Riebe, H. Häffner, C. Roos, W. Hänsel, J. Benhelm, G. Lancaster, T. Körber, C. Becher, F. Schmidt-Kaler, D. James *et al.*, “Deterministic quantum teleportation with atoms,” *Nature*, vol. 429, no. 6993, p. 734, 2004.
[14] A. Karlsson and M. Bourennane, “Quantum teleportation using three-particle entanglement,” *Physical Review A*, vol. 58, no. 6, p. 4394, 1998.
[15] D. Bouwmeester, J. Pan, K. Mattle, M. Eibl, H. Weinfurter, and A. Zeilinger, “Experimental quantum teleportation,” *Nature*, vol. 390, no. 6660, p. 575, 1997.
[16] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, “Quantum cryptography,” *Reviews of modern physics*, vol. 74, no. 1, p. 145, 2002.
[17] C. H. Bennett, F. Bessette, G. Brassard, L. Salvail, and J. Smolin, “Experimental quantum cryptography,” *Journal of cryptology*, vol. 5, no. 1, pp. 3–28, 1992.
[18] A. K. Ekert, “Quantum cryptography based on bell¡¯s theorem,” *Physical review letters*, vol. 67, no. 6, p. 661, 1991.
[19] R. Cleve, A. Ekert, C. Macchiavello, and M. Mosca, “Quantum algorithms revisited,” *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1969, pp. 339–354, 1998.
[20] L. Hardy, “Quantum theory from five reasonable axioms,” *arXiv preprint quant-ph/0101012*, 2001.
[21] M. P. Mueller and L. Masanes, “Information-theoretic postulates for quantum theory,” in *Quantum Theory: Informational Foundations and Foils*. Springer, 2016, pp. 139–170.
[22] G. Chiribella and R. W. Spekkens, *Quantum theory: Informational foundations and foils*. Springer, 2016.
[23] L. Hardy, “Reconstructing quantum theory,” in *Quantum theory: informational foundations and foils*. Springer, 2016, pp. 223–248.
[24] J. S. Bell, “On the einstein podolsky rosen paradox,” *Physics Physique Fizika*, vol. 1, no. 3, p. 195, 1964.
[25] G. Carleo and M. Troyer, “Solving the quantum many-body problem with artificial neural networks,” *Science*, vol. 355, no. 6325, pp. 602–606, 2017.
[26] D. Deng, X. Li, and S. D. Sarma, “Quantum entanglement in neural network states,” *Physical Review X*, vol. 7, no. 2, p. 021021, 2017.
[27] E. P. Van Nieuwenburg, Y. Liu, and S. D. Huber, “Learning phase transitions by confusion,” *Nature Physics*, vol. 13, no. 5, p. 435, 2017.
[28] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, “Neural-network quantum state tomography,” *Nature Physics*, vol. 14, no. 5, p. 447, 2018.
[29] W. J. Huggins, P. Patil, B. Mitchell, K. B. Whaley, and M. Stoudenmire, “Towards quantum machine learning with tensor networks,” *Quantum Science and technology*, 2018.
[30] D. Liu, S. Ran, P. Wittek, C. Peng, R. B. García, G. Su, and M. Lewenstein, “Machine learning by two-dimensional hierarchical tensor networks: A quantum information theoretic perspective on deep architectures,” *arXiv preprint arXiv:1710.04833*, 2017.
[31] V. Vedral, M. B. Plenio, M. A. Rippin, and P. L. Knight, “Quantifying entanglement,” *Physical Review Letters*, vol. 78, no. 12, p. 2275, 1997.
[32] G. Vidal and R. F. Werner, “Computable measure of entanglement,” *Physical Review A*, vol. 65, no. 3, p. 032314, 2002.
[33] A. Kitaev and J. Preskill, “Topological entanglement entropy,” *Physical review letters*, vol. 96, no. 11, p. 110404, 2006.
[34] S. R. Batten and R. Robson, “Interpenetrating nets: ordered, periodic entanglement,” *Angewandte Chemie International Edition*, vol. 37, no. 11, pp. 1460–1494, 1998.
[35] W. K. Wootters, “Entanglement of formation of an arbitrary state of two qubits,” *Physical Review Letters*, vol. 80, no. 10, p. 2245, 1998.
[36] L. Lami, “Non-classical correlations in quantum mechanics and beyond,” *arXiv preprint arXiv:1803.02902*, 2018.
[37] J. Zhang and Z. Li, “Quantum contextuality for training neural networks,” *Chinese Journal of Electronics*, vol. 29, no. 6, pp. 1178–1184, 2020.
[38] J. Zhang, Y. Hou, Z. Li, L. Zhang, and X. Chen, “Strong statistical correlation revealed by quantum entanglement for supervised learning,” in *ECAI 2020 - 24th European Conference on Artificial Intelligence, 29 August-8 September 2020, Santiago de Compostela, Spain, August 29 - September 8, 2020 - Including 10th Conference on Prestigious Applications of Artificial Intelligence (PAIS 2020)*, ser. Frontiers in Artificial Intelligence and Applications, vol. 325. IOS Press, 2020, pp. 1650–1657.
[39] E. P. Wigner, “On hidden variables and quantum mechanical probabilities,” *American Journal of Physics*, vol. 38, no. 8, pp. 1005–1009, 1970.
[40] V. Nair and G. E. Hinton, “Rectified linear units improve restricted boltzmann machines,” in *Proceedings of the 27th international conference on machine learning (ICML-10)*, 2010, pp. 807–814.
[41] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” 2014.
[42] D. Dua and C. Graff, “UCI machine learning repository,” 2017. [Online]. Available: http://archive.ics.uci.edu/ml

Authorized licensed use limited to: TIANJIN UNIVERSITY. Downloaded on November 09,2022 at 17:10:21 UTC from IEEE Xplore. Restrictions apply.