



Quantum Entanglement Inspired Correlation Learning for Classification

Junwei Zhang^{1(✉)}, Zhao Li^{2,4(✉)}, Juan Wang¹, Yinghui Wang¹, Shichang Hu³,
Jie Xiao⁴, and Zhaolin Li⁵

¹ College of Intelligence and Computing, Tianjin University, Tianjin, China
junwei@tju.edu.cn, wangyinghui@tju.edu.cn

² Zhejiang University, Zhejiang, China
zhao.li@zju.edu.cn

³ Alibaba Group, Zhejiang, China
shichang.hsc@alibaba-inc.com

⁴ Hangzhou Yugu Technology Co., Ltd., Hangzhou, China

⁵ Tianjin Xiniu Huaan Technology Co., Ltd., Tianjin, China

Abstract. Correlation is an important information resource, which is often used as a fundamental quantity for modeling tasks in machine learning. Since correlation between quantum entangled systems often surpasses that between classical systems, quantum information processing methods show superiority that classical methods do not possess. In this paper, we study the virtue of entangled systems and propose a novel classification algorithm called Quantum Entanglement inspired the Classification Algorithm (QECA). Particularly, we use the joint probability derived from entangled systems to model correlation between features and categories, that is, Quantum Correlation (QC), and leverage it to develop a novel QC-induced Multi-layer Perceptron framework for classification tasks. Experimental results on four datasets from diverse domains show that QECA is significantly better than the baseline methods, which demonstrates that QC revealed by entangled systems can improve the classification performance of traditional algorithms.

Keywords: Quantum correlation · Quantum-inspired algorithms · Classification algorithm

1 Introduction

In machine learning, correlation is considered an important information resource and is often used as a fundamental quantity in the modeling process of learning tasks. Correlation is any statistical association, although it usually refers to the degree to which a pair of variables is linearly related [6].

In recent years, quantum information technology have been developed by leaps and bounds [4, 9]. Quantum information processing has the advantages that classical information processing cannot match, and can complete information processing tasks that cannot be achieved by classical methods [16, 18, 19],

such as quantum teleportation, quantum communication, etc. Quantum Correlation (QC) in quantum composite systems [1, 17], namely, non-classical statistical correlation, has become more and more important, because it is the core quantum resource and it is stronger than classical statistical correlation [14]. In fact, the reason why quantum information processing has the superiority that classical information processing does not possess is because there is a correlation between quantum systems that is beyond classical correlation [12].

Since quantum theory is not widely used in classical machine learning tasks, here we give answers to several questions that readers may be concerned about. Although quantum theory is generally regarded as a micro-physical theory, its connotation is about information rather than physics. Since Hardy [7], 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 micro-physical world, but should be regarded as the general rules of information processing [3, 8].

In this paper, we study the virtue of quantum entangled systems in the classification tasks and propose a novel classification algorithm called **Quantum Entanglement inspired the Classification Algorithm (QECA)** to learn the statistical correlation between the features and categories. Particularly, base on the Multi-layer Perceptron (MLP), we develop a novel QC-induced classification framework. The framework uses a fully connected layer to learn the parameters of observations of the subsystems, and then uses the weighted sums to integrate the measured probability values of each entangled state. In short, it can be understood that the hidden layer neurons (nodes) of the MLP are replaced with a measurement process of entangled states. This replacement makes QECA has the ability to learn the QC between features and categories during training process. We validate the effectiveness of proposed QECA on four machine learning datasets, and the experimental results show that QECA not only significantly outperforms the baseline method MLP, but also achieves the best performance than the other comparing methods in most cases.

The contribution of this paper is to apply QC revealed by quantum entanglement into traditional classification tasks of machine learning and leverage it to develop a novel QC-induced classification algorithm. Moreover, this paper theoretically analyzes that the framework used has the ability to violate Bell inequality, which proves that the framework has the ability to reproduce QC. Finally, this paper experimentally verifies that QC learned by the framework is effective for classification tasks and combining QC into traditional classification frameworks can further boost the classification performance.

2 Theoretical Analysis and Verification by Bell Inequality

In quantum theory, when several particles interact, the properties of each particle will be integrated into the properties of the overall system, and the properties

of each particle can only describe the properties of the overall system. This phenomenon is called Quantum Entanglement (QE). QE could also be defined as one multi-body quantum system in which tensor decomposition is not possible [13]. First, let us give the basic definition of entanglement for bipartite systems (namely, 2-qubit).

Definition 1. Let \mathcal{H}_1 and \mathcal{H}_2 be two Hilbert spaces and $|\psi\rangle \in \mathcal{H}_1 \otimes \mathcal{H}_2$ ¹. Then $|\psi\rangle$ is said to be disentangled, or separable or a product state if $|\psi\rangle = |\psi_1\rangle \otimes |\psi_2\rangle$, for some $|\psi_1\rangle \in \mathcal{H}_1$ and $|\psi_2\rangle \in \mathcal{H}_2$. Otherwise, $|\psi\rangle$ is said to be entangled.

We begin with an arbitrary bipartite entangled state in the bases $\sigma_3|\pm\rangle = \pm|\pm\rangle$ ² that

$$|\psi\rangle = \alpha|+-\rangle + \beta|-+\rangle \quad (1)$$

where α and β are the normalization condition with $|\alpha|^2 + |\beta|^2 = 1$ but $\alpha, \beta \neq 0$. Without losing generality, α and β can be parameterized as $\alpha = e^{i\eta} \sin(\xi)$, $\beta = e^{-i\eta} \cos(\xi)$, where i is the imaginary number with $i^2 = -1$ and η, ξ are two real parameters but $\sin(\xi), \cos(\xi) \neq 0$. The density matrix of the entangled state, $\rho = |\psi\rangle\langle\psi|$, can be separated to the local and non-local parts [15], $\rho = \rho_{lc} + \rho_{nlc}$. The local part

$$\rho_{lc} = \sin^2(\xi)|+-\rangle\langle+-| + \cos^2(\xi)|-+\rangle\langle-+|, \quad (2)$$

describes the classic statistical correlation between subsystems (or properties), which belongs to the classical statistics. The non-local part

$$\rho_{nlc} = \sin(\xi) \cos(\xi) (e^{i2\eta}|+-\rangle\langle-+| + e^{-i2\eta}|-+\rangle\langle+-|) \quad (3)$$

describes the phenomenon of interference between subsystems (or properties), which belongs to the quantum statistics.

2.1 The Measurement on Density Matrix

The observable of the subsystem r of the bipartite entangled system, say a and b , is defined as:

$$M_r = \begin{bmatrix} \cos(\theta_r) & e^{-i\phi_r} \sin(\theta_r) \\ e^{i\phi_r} \sin(\theta_r) & -\cos(\theta_r) \end{bmatrix} \quad (4)$$

where θ and ϕ are two arbitrary real parameters and $r \in \{a, b\}$. The observable has a spectral decomposition, $M_r = \sum_m m P_r^m$, where P_r^m is the projector onto

¹ The widely used Dirac notations are used in this paper, in which a unit vector \mathbf{v} and its transpose \mathbf{v}^T are denoted as a ket $|v\rangle$ and a bra $\langle v|$, respectively. \otimes denotes the tensor product.

² $\{|+\rangle, |-\rangle\}$ denotes an arbitrary orthonormal basis of the 1-qubit Hilbert space \mathbb{C}^2 . $\sigma_3 = \sigma_z$ denotes Pauli matrix, and Pauli matrix refers to four common matrices, which are 2×2 matrix, each with its own mark, namely $\sigma_x \equiv \sigma_1 \equiv X$, $\sigma_y \equiv \sigma_2 \equiv Y$, $\sigma_z \equiv \sigma_3 \equiv Z$ and $\sigma_0 \equiv I$.

the eigenspace of M_r with eigenvalue m . The possible outcomes of the measurement correspond to the eigenvalues, m , of the observable. Upon measuring the state $|\varphi\rangle$, the probability of getting result m is given by

$$p(m_r) = \text{Tr}[P_r^m(|\varphi\rangle\langle\varphi|)] = \langle\varphi|P_r^m|\varphi\rangle \quad (5)$$

where Tr denotes the trace of the matrix.

Projective measurements have many nice properties. In particular, it is very easy to calculate average values for projective measurements. By definition, the average value of the measurement is

$$E(M) = \sum_m mp(m) = \sum_m m\langle\varphi|P_m|\varphi\rangle = \langle\varphi|M|\varphi\rangle. \quad (6)$$

The average value of the observable M is often written $\langle M \rangle \equiv \langle\varphi|M|\varphi\rangle$.

Therefore, the joint probability derived from QE is obtained as:

$$p(+_a, +_b) = \text{Tr}[(P_a^+ \otimes P_b^+)\rho]. \quad (7)$$

It can be also divided into the local (classical probability) and non-local (quantum probability) parts

$$p_{\text{all}}(+_a, +_b) = \text{Tr}[(P_a^+ \otimes P_b^+)(\rho_{lc} + \rho_{nlc})] \quad (8)$$

$$= \text{Tr}[(P_a^+ \otimes P_b^+)\rho_{lc}] + \text{Tr}[(P_a^+ \otimes P_b^+)\rho_{nlc}] \quad (9)$$

$$= p_{lc}(+_a, +_b) + p_{nlc}(+_a, +_b). \quad (10)$$

Accordingly, the probability of other combinations, i.e., $p_{lc}(\pm_a, \pm_b)$, $p_{lc}(\mp_a, \pm_b)$, $p_{nlc}(\pm_a, \pm_b)$ and $p_{nlc}(\mp_a, \pm_b)$, can also be obtained. Moreover, the average values of a and b in the classical and quantum cases are

$$\langle ab \rangle_{lc} = -\cos(\theta_a)\cos(\theta_b) \quad (11)$$

and

$$\langle ab \rangle_{nlc} = \sin(\theta_a)\sin(\theta_b)\sin(2\xi)\cos(\phi_a - \phi_b + 2\eta), \quad (12)$$

respectively.

2.2 Verification by Bell Inequality

The theoretical tool for verifying QE is the Bell inequality [2]. Violating (or Destroying) Bell inequality is a sufficient condition for the existence of QE. The Bell inequality has many well-known promotion forms, the first and simple of which is the Clauser-Horne-Shimony-Holt (CHSH) inequality [11]. The form of the CHSH inequality is simpler and more symmetrical than many other Bell inequalities that are later proposed. The specific form of the CHSH inequality is

$$|E(Q, S) + E(R, S) + E(R, T) - E(Q, T)| \leq 2 \quad (13)$$

where E denotes the average value and Q, R, S and T denote observable.

The main conclusions and their proofs are given below:

Conclusion 1. *The local part of the joint probability derived from QE satisfies the CHSH inequality, which belongs to the local hidden variables theory.*

Proof. Let $E(a, b) = \langle ab \rangle_{lc}$, i.e. Equation (11), the simple formula transformation and the absolute value inequality can prove that the CHSH inequality holds, i.e.

$$|\langle QS \rangle_{lc} + \langle RS \rangle_{lc} + \langle RT \rangle_{lc} - \langle QT \rangle_{lc}| \leq 2. \quad (14)$$

It indicates that p_{lc} is a classical probability.

Conclusion 2. *The non-local part (quantum interference term) of the joint probability derived from QE does not satisfy the CHSH inequality, which belongs to the quantum mechanics theory.*

Proof. Let $E(a, b) = \langle ab \rangle_{nlc}$, i.e. Equation (12), we use a counterexample to prove that the non-local part can violate the CHSH inequality. For example, when $\theta_Q = \theta_R = \theta_S = \theta_T = \frac{\pi}{2}$, $\phi_Q = \frac{\pi}{3}$, $\phi_R = \phi_S = \frac{\pi}{6}$, $\phi_T = 0$, $\xi = \frac{\pi}{4}$ and $\eta = 0$, then

$$|\langle QS \rangle_{nlc} + \langle RS \rangle_{nlc} + \langle RT \rangle_{nlc} - \langle QT \rangle_{nlc}| \approx 2.232 \not\leq 2. \quad (15)$$

It indicates that p_{nlc} is a quantum probability.

Conclusion 3. *The joint probability derived from QE does not satisfy the CHSH inequality, which belongs to the quantum mechanics theory.*

Proof. Let $E(a, b) = \langle ab \rangle_{all} = \langle ab \rangle_{lc} + \langle ab \rangle_{nlc}$, the CHSH inequality can also be violated. For example, when $\theta_Q = 0$, $\theta_R = \frac{\pi}{2}$, $\theta_S = \frac{5\pi}{4}$, $\theta_T = \frac{7\pi}{4}$, $\phi_Q = \phi_R = \phi_S = \phi_T = 0$, $\xi = \frac{\pi}{4}$ and $\eta = 0$, then

$$|\langle QS \rangle_{all} + \langle RS \rangle_{all} + \langle RT \rangle_{all} - \langle QT \rangle_{all}| = 2\sqrt{2} \not\leq 2. \quad (16)$$

It indicates that $p_{all} = p_{lc} + p_{nlc}$ is a quantum probability.

2.3 Analysis

Almost all books on quantum mechanics have discussions about the double-slit experiment, that is, electrons passing through two open slits. See also Ref. [13]. Let A_k denote an event of passing through the slit with label k , here $k = 1, 2$. Interpretation of the results of this experiment has led to the following formula for the probability:

$$p(A_1 \cup A_2) = p(A_1) + p(A_2) + 2\sqrt{p(A_1)p(A_2)} \cos(\theta) \quad (17)$$

where p is a symbol of probability and θ is a certain parameter. Generally, $2\sqrt{p(A_1)p(A_2)} \cos(\theta)$ is interpreted as the self-interference inherent to the wave nature of an electron. It will be convenient to give another form to Eq. (17). Set

$C = A_1 \cup A_2$ where $A_1 \cap A_2 = \emptyset$ and rewrite Eq. (17) as a nonclassical (quantum) total probability formula:

$$p(C) = p(C|A_1)p(A_1) + p(C|A_2)p(A_2) \tag{18}$$

$$+ 2\sqrt{p(C|A_1)p(A_1)p(C|A_2)p(A_2)} \cos(\theta). \tag{19}$$

where, as usual, $p(C|A_i) = p(CA_k)/p(A_k)$ and $p(A_k) > 0, k = 1, 2$.

Based on the quantum total probability formula, a natural judgment can be drawn that the quantum effect can be described as a quantum interference term for classical probability. In this paper, we decompose the quantum joint probability derived from QE into the classical probability and the quantum interference term, that is, we present the specific form of QC (or called strong statistical correlation) revealed by QE and the way it works, and use the CHSH inequality to verify its correctness. In the following, we will experimentally verify the role of this interference term in classical tasks.

3 Implement Classification Algorithm by the Framework

From the analysis of the previous theoretical section, we can get the following cognition: The essential reason that QC revealed by QE can be stronger than the classical correlation is that the quantum interference term described by the phase information is added. In this section, we will construct a classification algorithm based on the mathematical formalization of QE to verify the validity of QC revealed by QE in classification tasks.

This section is organized as follows: First, we will describe how to calculate the quantum joint probability between features and categories. Second, we describe how to use a fully connected layer to learn the parameters in the subsystems of an entangled system, that is, how to construct QECA. Formally, it can be understood as replacing the output layer of the MLP with the measurement operation of the entangled state. Finally, the learning method of parameters in the model is given.

3.1 Calculate Joint Probability Between Features and Categories

Entanglement arises from the measurement process of entangled systems (states), that is, obtaining the quantum joint probability not only requires entangled systems, but also requires to define the observables of the entangled systems.

We choose the quantum system with the maximum entanglement under two qubits as the entangled system, e.g., Bell states, and its form is

$$|\Psi\rangle = \frac{1}{\sqrt{2}} (|0\rangle \otimes |0\rangle + |1\rangle \otimes |1\rangle) = \frac{1}{\sqrt{2}} (|00\rangle + |11\rangle). \tag{20}$$

The reason for choosing the entangled system of two qubits is that we want to describe one qubit as the attribute (feature) and the other as the label (category). It can be seen that if there are N attributes in each instance (sample), N Bell

states are needed. Since there are only two eigenvalues in each set of orthogonal bases of a qubit, it is suitable for binary classification tasks. Of course, one can achieve multi-class classification tasks by adding the number of qubits of the label, but this is not the focus of this paper. Moreover, from the form of the quantum interference term, the probability amplitude and phase information of the entangled system can be fully reflected by the polar and azimuth angles of the measurement operator. In order to reduce the number of parameters of QECA, we choose the maximum entangled state, i.e., Bell states, to represent the entangled system.

We define the observable of the subsystem of the entangled system as Eq. (4), and the spectral decomposition of the observable is

$$M_r(\theta_r, \phi_r) = P_r^+(\theta_r, \phi_r) - P_r^-(\theta_r, \phi_r) = |+_r\rangle\langle+_r| - |-_r\rangle\langle-_r| \quad (21)$$

with

$$|+_r\rangle = \cos\left(\frac{\theta_r}{2}\right)|0\rangle + e^{i\phi_r}\sin\left(\frac{\theta_r}{2}\right)|1\rangle \quad (22)$$

$$|-_r\rangle = \sin\left(\frac{\theta_r}{2}\right)|0\rangle - e^{i\phi_r}\cos\left(\frac{\theta_r}{2}\right)|1\rangle \quad (23)$$

where the polar and azimuth angles, θ_r and ϕ_r , are the arbitrary real parameters. For the measurement operator of the attribute, P_{att} , we use ϕ_r to represent the parameter value of the attribute, and θ_r to represent the degree of freedom of the attribute, e.g., weight. For the measurement operator of the label, P_{lab} , we use the determined measurement operator to represent the label, e.g., $\theta_{lab} = \frac{\pi}{2}$ and $\phi_{lab} = 0$,

$$P_{lab}^+ = \frac{1}{2}(\sigma_1 + \sigma_0), \quad P_{lab}^- = \frac{1}{2}(\sigma_1 - \sigma_0). \quad (24)$$

In fact, any set of eigenstates can be chosen to represent the label, only to satisfy the orthogonality. The reason why we select a set of orthogonal bases to represent the label is that the positive and negative examples (samples) of the two-class classification task are (often) binary opposition.

Now we can formally define the measurement operator of the entangled system, i.e., Eq. (20). Assuming that each instance (sample) has N attributes and one label, the positive and the negative measurement operators for the entangled system consists of the n -th attribute and the label are

$$\mathcal{P}_n^\pm(\theta_n, \phi_n) = P_n^+(\theta_n, \phi_n) \otimes P_{lab}^\pm. \quad (25)$$

P_n^+ can also be replaced by P_n^- , the effect is the same. Applying \mathcal{P}_n^+ and \mathcal{P}_n^- separately to each entangled system, i.e., Eq. (20), the probability values of both positive and negative examples will be obtained,

$$p_n^\pm(\theta_n, \phi_n) = Tr[\mathcal{P}_n^\pm(\theta_n, \phi_n)(|\Psi\rangle\langle\Psi|)] = \langle\Psi|\mathcal{P}_n^\pm(\theta_n, \phi_n)|\Psi\rangle. \quad (26)$$

Based on this formal framework, we can calculate the quantum joint probability between the label and any attribute, and then construct QECA.

3.2 Constructing QECA by Quantum Joint Probability

In QECA, we use a fully connected layer to learn the parameters of the observations of the entanglement system. Formally, it can be understood as replacing the output layer of the MLP with a measurement operation of the entangled system. To make it easier for readers to understand the structure of QECA, we use the illustrated method to give the architecture of QECA, see Fig. 1.

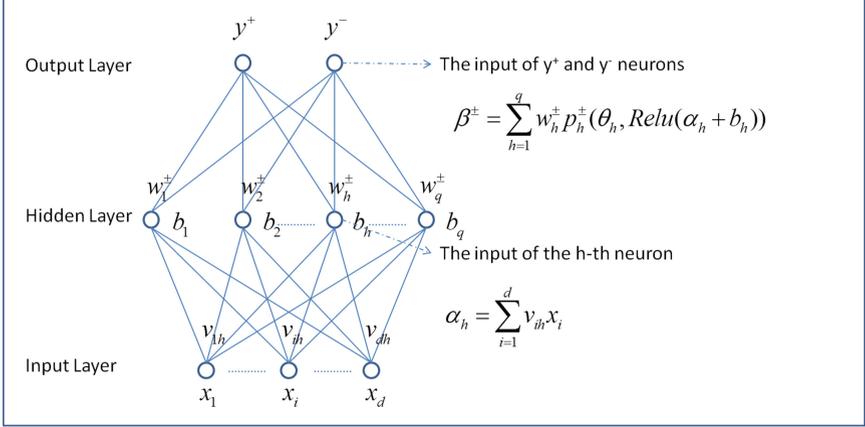


Fig. 1. Schematic diagram of QECA.

We perform weighting summation on the attributes of each instance, $x \in \mathbb{R}^d$, to get the input of the hidden layer neurons, which is $\alpha_h = \sum_{i=1}^d v_{ih} x_i$ where $v_h \in \mathbb{R}^d$ represents the weight. α_h plus the bias $b_h \in \mathbb{R}$, and then apply the activation function $ReLU$ (Rectified Linear Unit) [10] to get the parameters of the measurement operator of the entangled system, which is

$$\phi_h = ReLU(\alpha_h + b_h). \tag{27}$$

Together with the defined degrees of freedom, $\theta_h \in \mathbb{R}$, the measurement operator of the entangled system can be obtained, which is $\mathcal{P}_h^\pm(\theta_h, \phi_h)$, i.e., Eq. (25). By applying this measurement operator to the entangled state, i.e., Eq. (20), the joint probability value of the entangled state, i.e., Eq. (26), can be obtained.

Finally, perform weighting summation on $p_h^\pm(\theta_h, \phi_h)$ to get the final output value

$$y^\pm = \beta^\pm = \sum_{h=1}^q w_h^\pm p_h^\pm(\theta_h, \phi_h) \tag{28}$$

where $w^\pm \in \mathbb{R}^q$ represents the weight. β^\pm represent the input value of the output layer neurons, as shown in Fig. 1.

3.3 Parameter Learning

Machine learning uses the loss function to improve the performance of the model. This process of improvement is called optimization. QECA uses the classical cross-entropy loss function to act on its loss function. Since Adaptive Moment Estimation (Adam) defines a clear range of learning rates per iteration, making the parameters change smoothly, we use Adam as the optimizer for QECA.

4 Experiments

4.1 Datasets and Evaluation Metrics

The experiments were conducted on the four most frequently used machine learning datasets from UCI [5]. Due to the simulation of complex quantum operations on classical computers, limited by the computing power of classical computers, we can only verify our algorithms with lightweight datasets.

Abalone³ is a dataset that predicts the age of abalone through physical measurements. Since QECA is verified under a two-class task, it is divided into an adult group (covering $age \geq 10$) and adolescent group (covering $age < 10$). The purpose of this division is to make the amount of data in the two groups as close as possible.

Car Evaluation⁴ is a dataset that categorizes the car by a few simple indicators. We reclassified the original four categories into two, unacceptable (covering *unacc*) and acceptable (covering *acc*, *good* and *vgood*).

Wine Quality⁵ is a dataset that scores on wine quality. We divide the scores less than or equal to 5 into one class, and the others into another.

Breast Cancer⁶ is a dataset that is diagnosed by the patient’s physiological indicators, which is a two-class dataset.

All experiments use the 5-fold Cross-Validation method to divide the training set and test set. The experimental evaluation metrics, *F1-score*, *ACC* (Accuracy) and *AUC* (Area Under Curve), are taken as the average of 5 results.

4.2 Compared with Classical Classification Algorithms

Baselines: QECA is built on the basis of the standard MLP. Compared with the MLP of the same structure and setting, it can truly reflect the superiority of QECA. Both QECA and MLP uses an architecture of a single hidden layer, in order to compare them in a fair manner (or less interference). The number of neurons in the input layer is equal to the number of attributes; the number of neurons in the hidden layer is twice the number of input layers; because it is a binary classification task, the number of neurons in the output layer is two.

³ <http://archive.ics.uci.edu/ml/datasets/Abalone>.

⁴ <http://archive.ics.uci.edu/ml/datasets/Car+Evaluation>.

⁵ <http://archive.ics.uci.edu/ml/datasets/Wine+Quality>.

⁶ [http://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+\(original\)](http://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+(original)).

Both use the cross-entropy loss function to evaluate the model and the optimizer *Adam* to optimize the parameters.

We also conduct a comprehensive comparison across a wide range of machine learning algorithms, including Logistic Regressive (LR), Decision Tree (DT), Naive Bayesian Model (NBM), K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Gradient Boosting Decision Tree (GBDT) and Ada Boosting Decision Tree (ABDT).

Parameter Settings: QECA has three hyper-parameters, which are *learning rate*, *mini-batch* and *training epoch*, respectively, and uses the same settings on all datasets: the *learning rate* is 0.0001, the *mini-batch* is 1 and the *training epoch* is 500. Their *weights* are initialized to a truncated positive distribution, and the *biases* to 0.01. The permutation and combination method is used to select the hyper-parameters.

The hyper-parameters in the baselines are set to: in LR, *penalty* is *L2*; in DT, *min-samples-split* is 10; in SVM, *C* is 1.0 and *kernel* is *rbf*; in KNN, *n-neighbors* is 10; in LDA, *solver* is *svd* and *store-covariance* is *True*; in QDA, *store-covariance* is *True*; in MLP, *activation* is *relu* and *solver* is *adam*; in GBDT, *n-estimators* is 20; in ABDT, *n-estimators* is 20. Other hyper-parameters not listed use the default value of the framework scikit-learn⁷.

Experiment Results: Inspired by the quantum double-slit experiment, we also use the quantum interference term to characterize the strong statistical correlation revealed by QE and design an algorithm to verify the role of the quantum interference term in the classification task. Table 1 presents the experiment results under Abalone, Breast Cancer, Wine Quality (Red) and Car Evaluation respectively, where bold values are the best performances out of all algorithms. From the experimental results, the most metrics of QECA on four datasets are significantly better than the majority of machine learning algorithms. The basic conclusion that QECA has excellent classification ability can be drawn. This proves the effectiveness of QECA from a holistic perspective.

Moreover, the comparison with the MLP can explain that QECA is improved on the basis of MLP, and it shows that the quantum interference term plays an important role on QECA, that is, the quantum interference term described by the quantum phase has learned a strong statistical correlation between attributes and labels. Below we will analyze the entire learning process to determine whether the learning (or classification) ability is stable rather than accidental.

4.3 Comparison with the Training Process of Standard MLP

In order to analyze QECA's learning ability in more detail, we compared the training process of QECA with the baseline method MLP. We use the validation

⁷ <https://scikit-learn.org/stable/index.html>.

Table 1. Experiment results: the best-performed values for each dataset are in bold.

Dataset	Abalone			WQ (Red)			Car evaluation			Breast cancer		
	F1-score	ACC	AUC	F1-score	ACC	AUC	F1-score	ACC	AUC	F1-score	ACC	AUC
LR	0.7704	0.7708	0.7708	0.7559	0.7410	0.7404	0.7582	0.8611	0.8224	0.9494	0.9648	0.9594
DT	0.7153	0.7170	0.7170	0.7340	0.7191	0.7187	0.9626	0.9774	0.9756	0.8957	0.9298	0.9151
NBM	0.7387	0.7345	0.7345	0.7451	0.7298	0.7292	0.8519	0.9062	0.9043	0.9450	0.9604	0.9628
KNN	0.7791	0.7842	0.7842	0.6682	0.6447	0.6428	0.9422	0.9670	0.9477	0.9493	0.9648	0.9593
SVM	0.7678	0.7541	0.7543	0.7299	0.7110	0.7095	0.9565	0.9733	0.9732	0.9444	0.9589	0.9655
LDA	0.7443	0.7603	0.7601	0.7566	0.7273	0.7224	0.7698	0.8657	0.8323	0.9422	0.9605	0.9513
QDA	0.7390	0.7560	0.7558	0.7532	0.7204	0.7146	0.8500	0.9161	0.8816	0.9344	0.9516	0.9570
MLP	0.7896	0.7864	0.7865	0.7361	0.7373	0.7412	0.9384	0.9629	0.9564	0.8980	0.9280	0.9214
GBDT	0.7880	0.7859	0.7860	0.7609	0.7467	0.7460	0.9316	0.9571	0.9616	0.9279	0.9471	0.9477
ABDT	0.7884	0.7842	0.7843	0.7536	0.7335	0.7312	0.9155	0.9490	0.9409	0.9301	0.9517	0.9445
QECA	0.8018	0.8027	0.8027	0.7633	0.7536	0.7544	0.9759	0.9855	0.9841	0.9630	0.9736	0.9758
over MLP	1.54%↑	2.07%↑	2.05%↑	3.69%↑	2.21%↑	1.78%↑	3.99%↑	2.34%↑	2.89%↑	7.23%↑	4.91%↑	5.90%↑

set divided by the 5-fold Cross-Validation method as the test set to obtain the accuracy curve during the training process. The hyper-parameter settings of both QECA and MLP are exactly the same as those in Experiment Sect. 4.2. We selected three representative datasets for experiments: Wine Quality (Red) and Wine Quality (White) have the same data structure, but the amount of data in Wine Quality (Red) is balanced and Wine Quality (White) is not balanced; Moreover, in order to illustrate the effect of the number of attributes on the training effect, we use Abalone to compare with Wine Quality.

The experiment results are shown in Fig. 2. From the accuracy curve of the training process under the three datasets, compared with the MLP, QECA has significant improvement and its contribution is obvious. The experimental results of this section can prove that the quantum interference term plays an important role in QECA. It also further shows that QC revealed by QE can play an important role in the classic classification task.

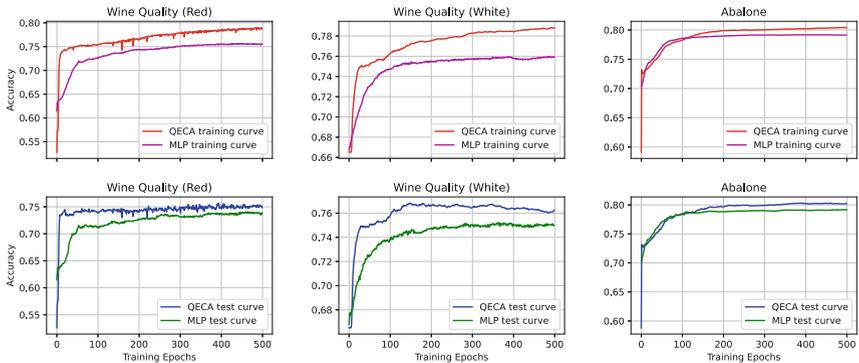


Fig. 2. Experiment results: the left column is the accuracy curve on the training set, and the right column on the test set.

5 Conclusion and Future Work

In this paper, we propose a novel classification framework, called Quantum Entanglement inspired Classification Algorithm (QECA), to learn a strong statistical correlation (i.e., QC) between features and categories and leverage it to improve the classification performance by integrating QC into MLP. QECA achieved excellent results on the four machine learning datasets compared with the baseline method MLP, which only uses the statistical correlation described by classical theory. More importantly, QECA also outperforms the other competitive methods in most metric. These results prove the effectiveness of QC in classification tasks.

References

1. Bai, Y.K., Yang, D., Wang, Z.: Multipartite quantum correlation and entanglement in four-qubit pure states. *Phys. Rev. A* **76**(2), 022336 (2007)
2. Bell, J.S.: On the einstein podolsky rosen paradox. *Phys. Physique Fizika* **1**(3), 195 (1964)
3. Chiribella, G., Spekkens, R.W.: *Quantum Theory: Informational Foundations and Foils*. Springer, Dordrecht (2016). <https://doi.org/10.1007/978-94-017-7303-4>
4. Dowling, J.P., Milburn, G.J.: Quantum technology: the second quantum revolution. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **361**(1809), 1655–1674 (2003)
5. Dua, D., Graff, C.: UCI machine learning repository (2017). <http://archive.ics.uci.edu/ml>
6. Hall, M.A.: Correlation-based feature selection for machine learning (1999)
7. Hardy, L.: Quantum theory from five reasonable axioms. arXiv preprint quant-ph/0101012 (2001)
8. Kochen, S.B.: A reconstruction of quantum mechanics. In: Bertlmann, R., Zeilinger, A. (eds.) *Quantum [Un]Speakables II*. TFC, pp. 201–235. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-38987-5_12
9. Hayashi, M.: *Quantum Information Theory*. Graduate Texts in Physics, Springer, Heidelberg (2017). <https://doi.org/10.1007/978-3-662-49725-8>
10. Hinton, G.E.: Rectified linear units improve restricted Boltzmann machines Vinod Nair (2010)
11. Khrennikov, A.: CHSH inequality: quantum probabilities as classical conditional probabilities. *Found. Phys.* **45**(7), 1–15 (2014)
12. Kok, P., Lovett, B.W.: *Introduction to Optical Quantum Information Processing*. Cambridge University Press, Cambridge (2010)
13. Nielsen, M.A., Chuang, I.: *Quantum computation and quantum information* (2002)
14. Wittek, P.: *Quantum Machine Learning: What Quantum Computing Means to Data Mining*. Academic Press, Cambridge (2014)
15. Zhang, H., Wang, J., Song, Z., Liang, J.Q., Wei, L.F.: Spin-parity effect in violation of bell's inequalities for entangled states of parallel polarization. *Mod. Phys. Lett. B* **31**(04), 1750032 (2017)
16. Zhang, J., et al.: Quantum correlation revealed by bell state for classification tasks. In: 2021 International Joint Conference on Neural Networks (IJCNN), pp. 1–8 (2021)
17. Zhang, J., Hou, Y., Li, Z., Zhang, L., Chen, X.: Strong statistical correlation revealed by quantum entanglement for supervised learning. In: ECAI (2020)
18. Zhang, J., Li, Z.: Quantum contextuality for training neural networks. *Chin. J. Electron.* **29**, 1178–1184 (2020)
19. Zhang, J., Li, Z., He, R., Zhang, J., Wang, B., Li, Z., Niu, T.: Interactive quantum classifier inspired by quantum open system theory. In: 2021 International Joint Conference on Neural Networks (IJCNN), pp. 1–7 (2021)