

Quantum-Inspired Neural Network with Runge-Kutta Method

Zipeng Fan*, Jing Zhang*, Peng Zhang[†], Qianxi Lin, Hui Gao

College of Intelligence and Computing, Tianjin University, Tianjin, China
 {2018218026, pzheng}@tju.edu.cn

Abstract

In recent years, researchers have developed novel Quantum-Inspired Neural Network (QINN) frameworks for the Natural Language Processing (NLP) tasks, inspired by the theoretical investigations of quantum cognition. However, we have found that the training efficiency of QINNs is significantly lower than that of classical networks. We analyze the unitary transformation modules of existing QINNs based on the time displacement symmetry of quantum mechanics and discover that they are resembling a mathematical form similar to the first-order Euler method. The high truncation error associated with Euler method affects the training efficiency of QINNs. In order to enhance the training efficiency of QINNs, we generalize QINNs' unitary transformation modules to the Quantum-like high-order Runge-Kutta methods (QRKs). Moreover, we present the results of experiments on conversation emotion recognition and text classification tasks to validate the effectiveness of the proposed approach.

Introduction

In recent years, researchers have discovered quantum-like phenomena in language understanding (Bruza, Kitto, and McEvoy 2008), leading to the proposal of Quantum-Inspired Neural Networks (QINNs) (Mönning 2019; Shi et al. 2021; Li, Wang, and Melucci 2019; Gkoumas et al. 2021; Chen, Pan, and Dong 2021; Li et al. 2021b) based on quantum probability theory. Through theoretical analysis, researchers have demonstrated the rationality of integrating quantum theory with neural network frameworks (Zhang et al. 2018b, 2019). QINNs are primarily applied to Natural Language Processing (NLP) tasks. When modeling dynamic evolution processes of semantic information, such as dynamic emotional states in a conversation, QINNs utilize unitary transformation of quantum states to track this process (Li et al. 2021b). This approach is similar to the evolution of quantum states over time.

However, when the QINNs based on unitary transformation is applied to downstream prediction tasks, the training efficiency is significantly lower than that of classical

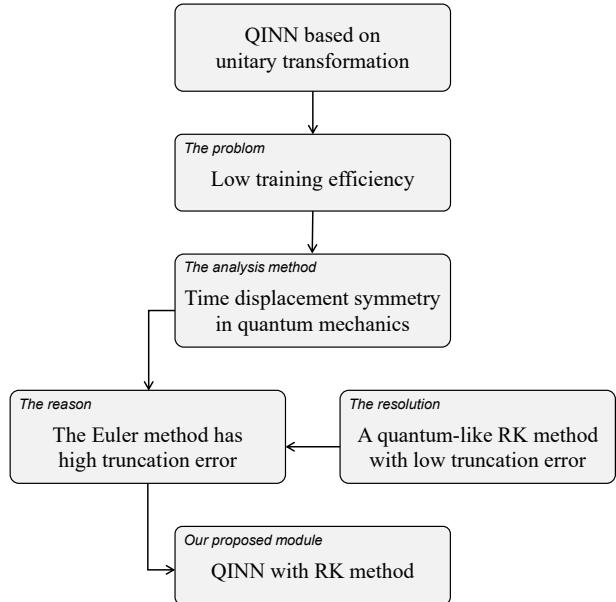


Figure 1: This is our research idea. The specific details are elaborated in Section Methodology.

networks. For example, when both the DialogueRNN (Majumder et al. 2019) and QINNs are used for prediction tasks, after the initial training periods (approximately two epochs), DialogueRNN's accuracy in predicting labels starts to improve, while QINN requires at least several times the training periods of DialogueRNN before the QINN's prediction accuracy improves. This signifies the evident inadequacy of QINN's training efficiency. In the experimental section, we provide detailed experimental results.

In order to explore the reasons behind this phenomenon, we conducted an analysis of the QINN's unitary transformation module. Our research idea is shown in Fig. 1. Based on theory of differential equations and the symmetry principle of time displacement in quantum mechanics (Jinyan 2000), we derived that the unitary transformation module in QINN is a numerical solving process similar to the Euler method. However, lower-order Runge-Kutta method (RK), including the Euler method, exhibit higher truncation errors compared

*These authors contributed equally.

[†]Corresponding author: Peng Zhang (pzheng@tju.edu.cn)
 Copyright © 2024, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

to higher-order RK methods, and the truncation error of the Euler method is not negligible (He et al. 2019). Based on the discovery, we designed relevant experiments, and the experimental results confirmed that the high truncation error has an impact on training efficiency. Therefore, we propose a modified unitary transformation module based on time displacement symmetry to alleviate the impact of truncation errors and improve training efficiency.

In our work, we improve the unitary transformation module of QINNs in order to address the issue of training efficiency. Our analysis indicates that the unitary transformation module can be replaced by the Quantum-like high-order Runge-Kutta methods (QRKs), which reduces the truncation errors introduced by the Euler method and significantly improves the training efficiency of QINN. In comparison to the Classical high-order Runge-Kutta methods (CRKs), QRKs are derived from the perspective of quantum mechanics symmetry principles, and the modeling approach is constrained by unitary property, resulting in a smaller range of truncation errors.

The main contributions of this paper are summarized as follows.

- We analyze the reason for the low training efficiency of QINNs from the perspective of the time displacement symmetry in quantum mechanics and differential equations.
- We propose the Quantum-like high-order Runge-Kutta methods (QRKs) based on time displacement symmetry to enhance the training efficiency and model performance of QINNs.
- We apply QRKs to the QINNs based on unitary transformation, namely Quantum Measurement inspired Neural Network (QMNN) and Quantum Language Model with Entanglement Embedding (QLM-EE), achieving superior performance on conversational emotion recognition and text classification datasets.

Preliminaries on Quantum Theory and Runge-Kutta Method

Quantum Theory

State Mathematically, an n -level quantum system can be described by an n -dimensional Hilbert space \mathcal{H}^n . Any quantum pure state can be described by a unit complex vector \mathbf{v} on \mathcal{H}^n . The pure state can be represented as a density matrix $\rho = \mathbf{v}\mathbf{v}^\dagger$, where \mathbf{v}^\dagger is conjugate transpose of \mathbf{v} . For the set of pure states $\{\mathbf{v}_i\}_{i=1}^{i=n}$ with weights $\{p_i\}_{i=1}^{i=n}$ that sum up to 1, this mixed state's density matrix ρ is computed by $\rho = \sum_{i=1}^n p_i \mathbf{v}_i \mathbf{v}_i^\dagger$.

Unitary Transformation The evolution of a closed quantum system is described by a unitary transformation (Nielsen and Chuang 2002). That is, the state \mathbf{v}_t of the system at time t is related to the state $\mathbf{v}_{t+\Delta t}$ of the system at time $t + \Delta t$ by a complex unitary matrix $\mathbf{U} \in \mathcal{H}$ which depends only on the times t and $t + \Delta t$ as $\mathbf{U}\mathbf{v}_t = \mathbf{v}_{t+\Delta t}$, and complex unitary matrix \mathbf{U} is satisfying $\mathbf{U}\mathbf{U}^\dagger = \mathbf{I}$. When the quantum system is represented as a density matrix ρ , the evolution makes the state change to $\mathbf{U}\rho_t\mathbf{U}^\dagger = \rho_{t+\Delta t}$.

Runge-Kutta Method

The Runge-Kutta (RK) methods are commonly used to solve Ordinary Differential Equation (ODE) in numerical analysis. The forward Euler method is a first-order RK method. High-order RK methods can achieve lower truncation errors than lower-order RK methods, including the forward Euler method. Therefore, the RK methods are ideal tools to construct network models from the dynamical systems view.

The RK methods are numerical methods originated from the Euler method. There are two types of RK methods: explicit and implicit ones. The family of RK methods is given by the following equations (Süli and Mayers 2003)

$$\begin{cases} y_{n+1} = y_n + \epsilon \sum_{i=1}^m \lambda_i K_i \\ K_1 = f(x_n, y_n) \\ K_i = f(x_n + a_{i-1}\epsilon, y_n + \epsilon \sum_{j=1}^{i-1} b_{ij} K_j) \end{cases} \quad (1)$$

where $i = 2, 3, \dots, r$, λ_i , a_i and b_{ij} are coefficients, ϵ is the time-step size that can be adaptive for different time steps, and $O(\epsilon^{m+1})$ is the truncation error of m -order RK method.

It is necessary to specify α in order to control the error of approximation in common numerical analysis. The varying time-step size can be adaptive to the regions with the different rates of change. The truncation error is lower when α is smaller. Adopting the RK methods brings higher prediction accuracy and better generalization capability into the neural network (Wang and Lin 1998).

Related Works of Quantum-Inspired Neural Network

After the advent of statistical language models, the Quantum Language Model (QLM) (Sordoni, Nie, and Bengio 2013) was introduced. It aimed to unite single words and compound terms within the same probability space, effectively preventing an exponential expansion of the term space. To enhance the practicality of quantum language models, the Neural Network based Quantum-like Language Model (NNQLM) (Zhang et al. 2018a) was later proposed. This model integrates quantum language model into an end-to-end Neural Network (NN) structure. Subsequently, a multitude of Quantum-Inspired Neural Network (QINN) models emerged for various Natural Language Processing (NLP) tasks. There are two QINN models include Quantum Measurement inspired Neural Network (QMNN) (Li et al. 2021b) and Quantum Language Model with Entanglement Embedding (QLM-EE) (Chen, Pan, and Dong 2021), which has a unitary transformation module.

QMNN is a quantum-like framework for the conversational emotion recognition task. In the task, the emotions of speakers are evolving throughout the conversation. Hence it is intuitive to employ quantum unitary transformation to track the dynamics of emotional states in a conversation. QMNN fuses the unimodal features (i.e., acoustic, visual and textual modalities) into a quantum mixed state ρ , and $\mathbf{U}\rho\mathbf{U}^\dagger$ represents the evolution of ρ over time, where \mathbf{U} is a complex-valued unitary matrix. QMNN implements a separate optimizer based on the Riemannian approach (Wisdom et al. 2016) to update unitary matrices.

QLM-EE employs a unitary transformation module to encode the correlations between the words as a quantum entangled state. In QLM-EE, each word w_i is embedded as a quantum pure state and described by a unit complex-valued vector \mathbf{w}_i corresponding to the superposition of sememes (Li, Wang, and Melucci 2019). The word sequences are given as the tensor product of word states, which takes as $\mathbf{s} = \mathbf{w}_1 \otimes \dots \otimes \mathbf{w}_L$. Subsequently, the transformation induced by the NN layer can be formally written as

$$\mathbf{s}' = \mathbf{W}\mathbf{s} \quad (2)$$

where \mathbf{W} is the weight matrix. Then the output vector must be normalized by

$$\mathbf{s}_E = \frac{\mathbf{s}'}{\|\mathbf{s}'\|_2} \quad (3)$$

According to the standard Gram-Schmidt procedure, Eq. 2 and Eq. 3 form a unitary transformation, and complete the entangled state representation of word sequences.

Methodology

In this section, we propose a Quantum-like Runge-Kutta Methods (QRKs). First, we introduce the relationship between unitary transformation and RK method, and then we introduce how to apply RK method to quantum-inspired neural network.

Unitary Transformation of QINNs Are an Euler Method

In QINNs, the unitary matrices are weight matrices subjected to the unitary constraint. From the perspective of quantum mechanics, utilizing unitary matrices solely as weight matrices in QINNs is insufficient. This approach does not take into account the significant role of unitary matrix in quantum mechanical symmetry. For a physical system, it undergoes evolution over time, referred to as time displacement. Time displacement belongs to the continuous transformation, which is a unitary transformation. Given that the research domain of QINNs primarily pertains to the process of modeling sequential information, we analyze the unitary transformation process in QINNs from the viewpoint of time displacement. This process must adhere to the time displacement symmetry as Theorem. 1.

Theorem 1 Time Displacement Symmetry (Jinyan 2000). According to the Schrodinger equation, $i\hbar \frac{\partial}{\partial t} \mathbf{v} = \mathbf{H}\mathbf{v}$. If the Hamiltonian \mathbf{H} does not depend explicitly on time t , then the evolution of the system state over time is independent of the choice of initial time, resulting in time displacement symmetry. The evolution of the state over time can be expressed as $\mathbf{v}_t = e^{-i\mathbf{H}t/\hbar} \mathbf{v}_0$, where $e^{-i\mathbf{H}t/\hbar}$ is the operator for time displacement t of the system. For any infinitesimal time displacement Δt and initial time t ,

$$\begin{aligned} \mathbf{v}_{t+\Delta t} &= \exp(-i\mathbf{H}\Delta t/\hbar) \mathbf{v}_t \\ &\approx (1 - i\mathbf{H}\Delta t/\hbar) \mathbf{v}_t \\ &= \mathbf{v}_t - i\mathbf{H}\alpha \mathbf{v}_t \end{aligned} \quad (4)$$

where α is a minimum value, which replaces $\Delta t/\hbar$.

Corollary 1 The equation describing the evolution of quantum state \mathbf{v}_t is equivalent to the first-order Euler numerical solution if $f(t, \mathbf{v}_t)$ is a continuous function.

Proof 1 From the perspective of linear algebra, \mathbf{H} can be considered as a square matrix that operates on the quantum state \mathbf{v}_t , where $\mathbf{H} = \mathbf{H}^\dagger$. This operation can be viewed as a linear transformation of quantum state \mathbf{v}_t . This linear transformation can be written as $f(t, \mathbf{v}_t) = -i\mathbf{H}\mathbf{v}_t$. The Eq. 4 can be written as a continuous function $\mathbf{v}_{t+\Delta t} = \mathbf{v}_t + \alpha f(\mathbf{v}_t)$. After we replace $f(\mathbf{v}_t)$ with K_1 , $\mathbf{v}_{t+\Delta t}$ can be expressed as

$$\mathbf{v}_{t+\Delta t} = \mathbf{v}_t + \alpha K_1 \quad (5)$$

According to the Eq. 1, the first-order Euler method is expressed as $y_{n+1} = y_n + \epsilon \lambda_1 K_1$. Mathematically, the evolution process of the quantum state \mathbf{v}_t is equivalent to the first-order Euler method.

For continuous function $f(\mathbf{v}_t)$ with K_1 , the Euler method shown as $\mathbf{v}_{t+\Delta t}$ has high truncation error because it is a first-order approximation to the true solution. This introduces the drawback of the first-order Euler method to the unitary transformation module, which is difficult to learn the long-range information in the continuous network, reducing accuracy of model outputs (Zhu, Chang, and Fu 2022; Li et al. 2021a). Therefore, due to the limitation of Euler method, it results in the low training efficiency of QINNs. In Experiments Section, we demonstrated this impact in the conversational emotion recognition task.

From Unitary Transformation to High-Order RK Method

By analyzing the physical meaning of unitary transformation, we found that the Euler method's truncation error limits the training efficiency of QINNs. While stacking unitary transformation modules may seem like a straightforward approach to quickly capture features and enhance the model's training efficiency, it is not always effective in NLP tasks. In fact, research has shown that when more layers are stacked, errors can propagate through the neural network and hinder the system's ability to benefit from an extremely deep model (Li et al. 2020; Dai et al. 2019).

In the differential equations, the Euler method is a low-order RK method. The higher-order RK methods have lower truncation errors than lower-order RK method (Zhang et al. 2021). Therefore, we need to consider how to use high-order RK methods to reasonably improve the existing QINNs.

According to the time displacement of a quantum state, we extend the unitary transformation of QINNs to the multi-time step time displacement. Taking two-time step as an example, according to Eq. 4, the quantum state $\mathbf{v}_{t+\Delta t}$ evolution process can be iteratively calculated as

$$\begin{cases} \mathbf{v}_{t+2\Delta t} = \mathbf{v}_t + \alpha K_1 + \alpha f(\mathbf{v}_{t+\Delta t}) \\ f(\mathbf{v}_{t+\Delta t}) = f(\mathbf{v}_t + \alpha K_1) \end{cases} \quad (6)$$

where $f(\mathbf{v}_{t+\Delta t})$ is the RK block K_2 in the second-order RK method. We generalize the single time step unitary transformation to the n -time step unitary transformation. We assume

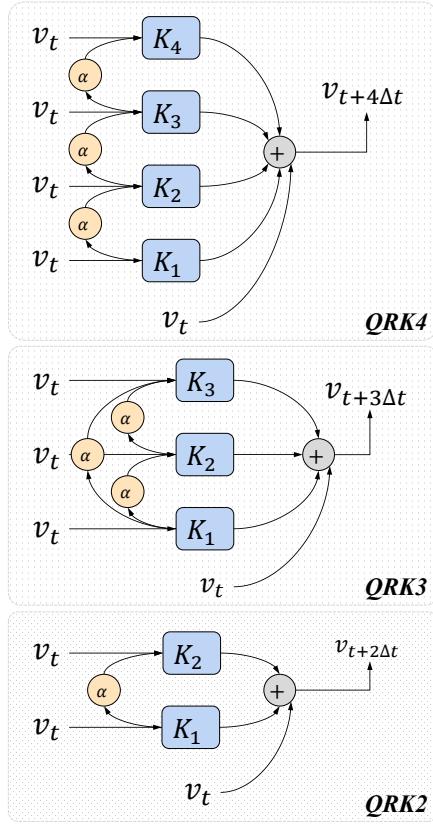


Figure 2: The structures of quantum-like second-order RK method, quantum-like third-order RK method and quantum-like fourth-order RK method.

that n -time step unitary transformation is performed on a state \mathbf{v}_t , and $\mathbf{v}_{t+n\Delta t}$ is expressed as

$$\begin{cases} \mathbf{v}_{t+n\Delta t} = \mathbf{v}_t + \sum_{i=1}^n \alpha K_i \\ K_1 = f(\mathbf{v}_t) \\ K_i = f(\mathbf{v}_t + \sum_{j=1}^{i-1} \alpha K_j) \end{cases} \quad (7)$$

where $i = 2, 3, \dots, n$, α is a minimum value, and $f()$ is a continuous function. In Eq. 1, the $\lambda, b \in \mathcal{R}$ are coefficients of classical high-order RK method. Since $\alpha \rightarrow 0^+$, $\frac{\alpha}{\lambda}$ and $\frac{\alpha}{b}$ approximate to α . Therefore, our investigation revealed that the n -time step unitary transformation can be interpreted as a Quantum-like n -order RK method, called QRK. When $n = 1$, it is the unitary transformation of QINNs, which is a first-order RK method, including the Euler method.

We establish a unifying framework for QINN's unitary transformation module by a formalism based on the high-order RK method. Specifically, we show that any multi-step time displacement can be expressed as a variant of the high-order RK method. Therefore, we can use the advantages of high-order RK method to improve the training efficiency of QINNs.

The Error Range of the Unitary Matrix and the Truncation Error of RK Method

Although we have established a correspondence between the unitary transformation and a high-order RK method, the direct substitution of the unitary transformation module with a classical high-order RK method is not feasible. The high-order RK method we derived relies on the multi-step time displacement of a quantum state, which must be constrained by its unitary property. In other words, our proposed method is a quantum-like method that leverages the unitary constraint to achieve a smaller truncation error. To establish this advantage, we conduct an analysis of the correlation between the quantum state and the high-order RK method.

For any quantum state $\mathbf{v}_{t+n\Delta t}$, we treat $\mathbf{v}_{t+n\Delta t}$ as a general continuous function from the perspective of linear algebra, and it can be expressed as a linear combination of a set of functions

$$\begin{cases} \mathbf{v}_{t+n\Delta t} = \mathbf{v}_t + n\Delta t \sum_{i=1}^m C_i K_i \\ K_i = f(t + a_i n\Delta t, \mathbf{v}_t + n\Delta t \sum_{j=1}^{i-1} b_{ij} K_j) \end{cases} \quad (8)$$

where m is the order of RK method. Then the truncation error is determined by Taylor expansion

$$\begin{aligned} T_{t+n\Delta t} &= \mathbf{v}_{t+n\Delta t} - \mathbf{v}_t - n\Delta t \sum_{i=1}^m C_i K_i \\ &= O((n\alpha\hbar)^{m+1}) \end{aligned} \quad (9)$$

where $\Delta t = \alpha\hbar$, and α also determines the truncation error of the RK method on the quantum state. In Eq. 4, $\mathbf{I} - i\mathbf{H}\alpha$ is a unitary matrix \mathbf{U} (Jinyan 2000) and satisfies the following condition:

$$\begin{aligned} \mathbf{U}^\dagger \mathbf{U} &= (\mathbf{I} - i\mathbf{H}^\dagger \alpha)(\mathbf{I} - i\mathbf{H}\alpha) \\ &= \mathbf{I} + i\alpha(\mathbf{H}^\dagger - \mathbf{H}) + O(\alpha^2) \end{aligned} \quad (10)$$

When $\alpha \rightarrow 0^+$, $\mathbf{U}^\dagger \mathbf{U} \approx \mathbf{I}$. So α is the error introduced when modeling the unitary matrix \mathbf{U} . In Eq. 1, the step-size of the high-order RK method satisfies $\epsilon > 0$.

We find that the truncation error of the high-order RK method and the truncation error of the quantum state-based high-order RK method satisfy $O(\epsilon^{m+1}) > O((n\alpha\hbar)^{m+1})$. Therefore, the unitary transformation is a quantum-like high-order RK method, the classical high-order RK method cannot be directly used to replace the unitary transformation module of QINNs. Due to the existence of the unitary constraint, the truncation error becomes a controllable variable, thereby further enhancing the effectiveness and the training efficiency of QINNs.

Note that it is crucial to handle the value of α with care. In the case where $\alpha = 0$, the unitary matrix \mathbf{U} degenerates into the identity matrix \mathbf{I} , leading to inefficient quantum state evolution. Moreover, the unitary transformation module might even lose its functionality. In the ablation experiments, we verified the crucial role of α in QINNs in both conversational emotion recognition task and text classification task.

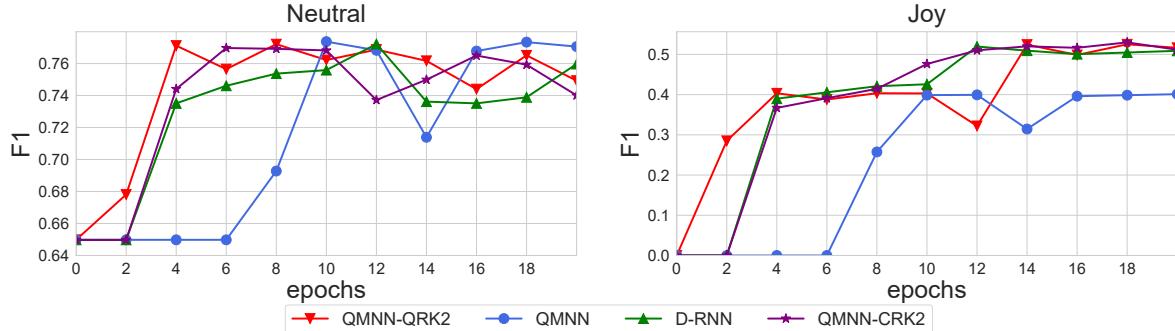


Figure 3: We display in the training efficiency of QMNN-QRK2, QMNN-CRK, QMNN and DialogueRNN (D-RNN) on Neutral F1 and Joy F1. The QMNN-QRK2 has the best training efficiency.

Quantum-Inspired Neural Network With Runge-Kutta Method

We describe the workflow of QINN with high-order QRK method to model the word sequence (sentence or phrase) in Fig. 2. In our work, we use the second-order QRK method (QRK2), third-order QRK method (QRK3) and fourth-order QRK method (QRK4).

In QINNs, they use unitary transformation module to extract and model language semantics as $\mathbf{v}' = \mathbf{U}\mathbf{v}$, where \mathbf{v} is the quantum state of word or phase. We have already proved the relationship of unitary transformation and RK method. Therefore, we use QRKs instead of unitary transformation module. Taking QRK4 as an example, there are four QRK blocks as shown in Eq. 11 and $f() = -i\mathbf{H}\alpha$.

$$\left\{ \begin{array}{l} \mathbf{v}_{t+\Delta t} = \mathbf{v}_t + \sum_{i=1}^4 \alpha K_i \\ K_1 = f(\phi(t)) \\ K_2 = f(\phi(t) + \alpha K_1) \\ K_3 = f(\phi(t) + \alpha K_2) \\ K_4 = f(\phi(t) + \alpha K_3) \end{array} \right. \quad (11)$$

where Δt is the sum of the four evolution times. QRK4 performs four feature extractions on \mathbf{v}_t in turn, so QRK4 can be regarded as a multi-layer quantum-inspired neural network module.

Experiments

Experimental Setup

Baselines. (1) To analyze QINN’s training efficiency, we choose **QMNN** (Li et al. 2021b) and **QLM-EE** (Chen, Pan, and Dong 2021) because they use the unitary transformation modules. To provide a comprehensive comparison, we also included four important QINN models, **NNQLM-I**, **NNQLM-II** (Zhang et al. 2018a), **C-NNQLM-I** and **C-NNQLM-II** (Zhang et al. 2022).

(2) To compare QINN with classical models that share similar frameworks, we have choose **Multimodal Transformer** (MuLT) (Tsai et al. 2019), **DialogueRNN** (Majumder et al. 2019), **Fasttext** (Joulin et al. 2016) and **Text-CNN** (Rakhlin 2016). The reason for choosing DialogueRNN is that QMNN is a quantum-like RNN, and the reason for choosing CNN is that NNQLM and C-NNQLM

are QINN based on the CNN framework, and QLM-EE is currently a representative model of this type of QINN model.

(3) To compare quantum-like high-order RK methods with classical high-order RK methods, we replace the unitary transformation modules with classical high-order RK methods as **QMNN-CRK** and **QLM-EE-CRK**.

Tasks, Datasets, and Metrics. We conducted experiments on conversational emotion recognition and text classification tasks.

(1) For conversational emotion recognition task, we chose MELD (Poria et al. 2018). MELD is a multi-party conversation dataset crawled from the Friends TV series. In training dataset, MELD contains 1039 dialogues and 9989 utterances. *F1* and *precision* are used for evaluation.

(2) For text classification task, we chose four text classification datasets: MR (Hu and Liu 2004), CR (Pang and Lee 2005), SUBJ (Wiebe, Wilson, and Cardie 2005), MPQA (Pang and Lee 2004). MR contains 11.9K training data, 20K vocabulary, and two classes. CR contains 4K training data, 6K vocabulary, and two classes. SUBJ contains 10K training data, 21K vocabulary, and two classes. MPQA contains 11K training data, 6K vocabulary, and two classes. *Accuracy* (ACC) is used for evaluation.

Implementations. (1) For conversational emotion recognition task, we trained the QINN with quantum-like high-order RK (**QMNN-QRK**) methods on one NVidia Tesla K80 GPU. QMNN-QRK hyperparameters are searched within embedding dimensions $d \in \{50, 100, 120, 160, 200\}$, the size of last hidden layer in $\{16, 24, 32, 48, 64\}$. Stochastic gradient descent (SGD) is used as the optimizer with a learning rate $lr \in \{0.001, 0.002, 0.005, 0.008\}$. The batch size varies in $\{24, 32, 48\}$. The dropout rate for the last hidden layer varies in $\{0.001, 0.005, 0.01, 0.05, 0.1, 0.2, 0.3, 0.5\}$. The α varies in $\{0, 0.5, 0.05, 0.0001, 0.00001\}$. We set the number of parties $K = 1$ for MELD.

(2) For text classification task, we trained the QINN with quantum-like high-order RK (**QLM-EE-QRK**) methods on one NVidia Tesla K80 GPU. We search the hyperparameters from a parameter pool, with batch size in $\{4, 8, 16, 32\}$, learning rate in $\{0.01, 0.1, 0.3, 0.5\}$, L2-regularization rate is 0.001, the number of filter is 100, the size of filter is 200.

Model	Neutral		Surprise		Sad		Joy		Angry	
	F1	P								
MulT	76.46	71.74	47.50	41.34	22.52	36.17	52.40	51.82	46.64	42.75
DialogueRNN	74.70	72.13	47.76	41.13	15.58	24.00	49.48	51.62	46.09	34.51
QMNN	77.00	71.23	49.76	45.81	16.50	24.30	52.08	53.48	43.17	42.86
QMNN-CRK2	76.50	71.27	48.96	44.47	21.71	26.76	51.65	52.77	38.65	43.06
QMNN-QRK2	77.61	73.13	49.51	51.00	22.77	31.41	53.52	55.10	42.78	49.05

Table 1: Performances of models on MELD in percentage. The best performance values among all models are in bold. P is Precision.

Model	MR	SUBJ	CR	MPQA
Text-CNN	76.9	90.7	80.1	84.5
NNQLM-II	74.8	83.6	77.8	83.2
C-NNQLM-II	76.2	90.9	83.3	86.9
C-NNQLM-II [†]	75.1	88.9	77.8	84.7
C-NNQLM-II ^{††}	77.4	89.1	79.9	85.3
Fasttext	72.6	86.7	75.2	82.8
NNQLM-I*	62.9	86.6	75.5	82.9
NNQLM-I**	73.1	89.4	78.2	84.9
C-NNQLM-I*	73.9	89.6	78.1	86.3
C-NNQLM-I**	76.9	90.1	79.4	85.5
QLM-EE	71.7	89.6	78.5	80.9
QLM-EE-CRK4	77.1	91.8	82.2	81.5
QLM-EE-QRK4	78.8	94.6	84.4	85.5

Table 2: The results of text classification task. * and ** are two important version of NNQLM-I and C-NNQLM-I, * adopts only diagonal elements of density matrices, ** processes all elements of density matrices. [†] and ^{††} are two important version of C-NNQLM-II, [†] is weight sharing scheme with words, ^{††} is weight sharing scheme with dimensions.

The number of words in a sequence is $N \in \{1, 2, 3\}$, the word embedding dimension is $D \in \{4, 6, 8, 16\}$, the number of measurement vectors is $M \in \{800, 1000, 1500\}$. We test single-layer and two-layer fully connected neural networks with $\{128, 256, 512\}$ neurons for entanglement embedding.

Main Results

QINN With QRK Outperforms QINN and Other Baselines. (1) As shown in Table. 1, we compared QMNN-QRK2 and QINN baselines (original QMNN and QMNN-CRK2) on MELD dataset, where RK2 represents the second-order RK method. We also compared QMNN-QRK2 with the classical baselines (DialogueRNN and MulT) on MELD dataset. QMNN-QRK2 achieves the highest F1 and precision in most emotional labels. Compared to QMNN with unitary transformation module, the QRK method can significantly improve the performance of QMNN. Furthermore, when we replaced the unitary transformation module of QMNN with the classical second-order RK method, the performance of QMNN-CRK2 was inferior to that of QMNN-QRK2. This result suggests that the QRK method with unitary constraint offers distinct advantages over the classical RK method.

(2) As shown in Table. 2, we compared QLM-EE-QRK4

Model (param)	MR	SUBJ	CR	MPQA
QLM-EE(298K)	71.7	89.6	78.5	80.9
QLM-EE-CRK2(249K)	70.2	86.5	79.0	79.4
QLM-EE-QRK2(249K)	78.5	93.2	82.2	85.3
QLM-EE-CRK3(249K)	77.1	91.8	80.1	81.5
QLM-EE-QRK3(249K)	77.8	93.8	83.8	85.0
QLM-EE-CRK4(249K)	77.1	91.8	82.2	81.5
QLM-EE-QRK4(249K)	78.8	94.6	84.4	85.5

Table 3: The results of QLM-EE-QRK’s sensitivity on the order. The param is the parameters of QLM-EE, QLM-EE-CRKs, and QLM-EE-QRKs.

and QINN baselines (original QLM-EE and QLM-EE-CRK4) on four datasets, where RK4 represents the fourth-order RK method. QLM-EE-QRK4 achieves the best results compared to QINN baselines on MR, SUBJ, and CR, but slightly underperforms C-NNQLM on MPQA. Compared to QLM-EE with unitary transformation module, the QRK method can significantly improve the performance of QLM-EE. Specifically, QLM-EE-QRK4 achieves 7.1, 5.0, 5.9, and 4.6 ACC score improvements on MR, SUBJ, CR, and MPQA, respectively. Due to unitary constraint, QLM-EE-QRK4 outperforms QLM-EE-CRK4.

The Quantum-Like High-Order RK Method Can Enhance the Training Efficiency of QINN. We analyze the training efficiency of QMNN-QRK2 and QMNN. As illustrated in Fig. 3, where the epoch number is plotted on the abscissa and the F1 score on the ordinate, we can observe significant differences. Let’s consider Neutral F1 in the MELD dataset as an example. We noticed that the Neutral F1 score predicted by QMNN remained unchanged until the sixth epoch, whereas the model performance of DialogueRNN started improving after the second epoch. This indicates a substantial disparity in training efficiency between QMNN and DialogueRNN. However, when replacing the unitary transformation module in QMNN with the QRK2 method, we noticed an improvement in model performance starting from the first epoch, with the F1 score gradually stabilizing after the fourth epoch. These outcomes suggest that the quantum-like RK method has indeed enhanced the training efficiency of QINN. This phenomenon is also consistent in the classification of other emotions, such as Joy F1.

QMNN-QRK2 value of α	Neutral		Surprise		Sad		Joy		Angry	
	F1	P								
0.5	75.64	70.11	47.86	44.05	16.00	23.81	50.75	47.00	35.93	43.00
0.05	76.07	70.65	47.92	46.78	17.39	21.90	51.97	49.65	38.43	45.75
0.0001	76.51	72.58	47.93	47.12	19.54	27.59	52.51	53.56	43.00	45.28
0.00001	77.61	73.13	49.51	51.00	22.77	31.41	53.52	55.10	42.78	49.05
0	75.52	70.90	46.28	42.43	2.80	13.33	47.33	43.61	28.21	39.32

Table 4: The results of QMNN-QRK’s sensitivity on the range of truncation error.

QLM-EE	α	MR	SUBJ	CR	MPQA
QRK2	0.5	69.9	88.4	73.7	78.6
	0.05	74.6	90.7	79.6	81.0
	0.0001	76.6	90.8	80.1	81.4
	0.00001	78.5	93.2	82.2	85.0
QRK3	0.5	76.7	90.8	80.4	81.7
	0.05	77.5	91.8	80.1	81.8
	0.0001	76.5	91.7	80.3	82.5
	0.00001	77.8	93.8	83.8	85.0
QRK4	0.5	76.7	90.8	80.1	82.3
	0.05	77.4	91.8	80.1	82.5
	0.0001	76.9	91.7	79.8	83.0
	0.00001	78.8	94.6	84.4	85.5

Table 5: The results of QLM-EE-QRK’s sensitivity on the range of truncation error.

Ablation Study

Sensitivity on the Order. (1) In previous experiments, we set the order of QLM-EE-QRK and QLM-EE-CRK to four. In this section, we perform the statistical analysis and ablation experiments on the text classification datasets.

(2) We set the order of QLM-EE-QRK and QLM-EE-CRK to 2, 3, and 4, that is, limiting the maximum number of layer to 2, 3, and 4, respectively. From Table. 3, the QLM-EE-QRK4 achieves the best results. We posit that the utilization of high-order RK methods can effectively reduce truncation error in text classification tasks. When the order of RK method is the same, QLM-EE-QRK outperforms QLM-EE-CRK, demonstrating the importance of unitary constraint. Meanwhile, we found that the number of parameters between QLM-EE-QRK and QLM-EE is equivalent, so the performance improvement of QLM-EE-QRK is not brought about by additional parameters.

Sensitivity on the Range of Truncation Error. (1) According to our derivation, the truncation error range of the QRK method is controlled by the minimum value α of the unitary matrix. In previous experiments, we set α of QLM-EE-QRK and QMNN-QRK to 0.00001. In this section, we perform the statistical analysis and ablation experiments on the text classification and the conversational emotion recognition datasets.

(2) We choose α of QLM-EE-QRK and QMNN-QRK in $\{0.5, 0.05, 0.0001, 0.00001, 0\}$. From Table. 4, setting α of QMNN-QRK to 0.00001 achieves the best results. Furthermore, we observe a gradual improvement in the model’s per-

formance as α value ranges from 0.5 to 0.00001. In addition, we set α to 0 to investigate the impact of omitting training during the unitary transformation. Our results indicate that the model’s performance is significantly poorer without unitary transformation, underscoring the crucial role of the unitary transformation in effectiveness.

(3) From Table. 5, setting α of QLM-EE-QRK to 0.00001 achieves the best results. We found that as α decreases, the gap in performance between QLM-EE-QRKs of different orders also decreases. However, the time complexity $O(nL)$ of the model increases as the order n of the model increases, where L is the sequence length. The result in Table. 5 demonstrates that the performance of lower-order models can be improved by decreasing α , thereby reducing the time complexity of the model.

Conclusion

In this paper, we have demonstrated the inherent consistency between unitary transformations and high-order RK methods. Based on this foundation, we propose a Quantum-like high-order RK (QRK) module for QINN. Through experiments, we have verified the effectiveness of our method in enhancing the training efficiency of QINN based on unitary evolution, and it also shows improvements in experimental results compared to other QINN approaches. However, this work serves as a foundational study on QINN training efficiency and does not extensively explore model performance enhancements. In the future, we will further explore the relationship between NeuralODE and QINNs, and enhance model performance by refining multi-modal modeling methods based on QINN.

Acknowledgments

This work is supported in part by the Natural Science Foundation of China (grant No.62276188 and No.61876129), TJU-Wenge joint laboratory funding, Tianjin Research Innovation Project for Postgraduate Students (grant No.2021YJSB167).

References

- Bruza, P.; Kitto, K.; and McEvoy, D. 2008. Entangling words and meaning. In *Quantum Interaction: Proceedings of the Second Quantum Interaction Symposium (QI-2008)*, 118–124. College Publications.
- Chen, Y.; Pan, Y.; and Dong, D. 2021. Quantum language model with entanglement embedding for question answering. *IEEE Transactions on Cybernetics*.

Dai, Z.; Yang, Z.; Yang, Y.; Carbonell, J.; Le, Q. V.; and Salakhutdinov, R. 2019. Transformer-xl: Attentive language models beyond a fixed-length context. *arXiv preprint arXiv:1901.02860*.

Gkoumas, D.; Li, Q.; Dehdashti, S.; Melucci, M.; Yu, Y.; and Song, D. 2021. Quantum cognitively motivated decision fusion for video sentiment analysis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 827–835.

He, X.; Mo, Z.; Wang, P.; Liu, Y.; Yang, M.; and Cheng, J. 2019. Ode-inspired network design for single image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1732–1741.

Hu, M.; and Liu, B. 2004. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, 168–177.

Jinyan, Z. 2000. Quantum Mechanics, Volume I.

Joulin, A.; Grave, E.; Bojanowski, P.; and Mikolov, T. 2016. Bag of tricks for efficient text classification. *arXiv preprint arXiv:1607.01759*.

Li, B.; Du, Q.; Zhou, T.; Zhou, S.; Zeng, X.; Xiao, T.; and Zhu, J. 2021a. ODE transformer: An ordinary differential equation-inspired model for neural machine translation. *arXiv preprint arXiv:2104.02308*.

Li, B.; Wang, Z.; Liu, H.; Jiang, Y.; Du, Q.; Xiao, T.; Wang, H.; and Zhu, J. 2020. Shallow-to-deep training for neural machine translation. *arXiv preprint arXiv:2010.03737*.

Li, Q.; Gkoumas, D.; Sordoni, A.; Nie, J.-Y.; and Melucci, M. 2021b. Quantum-inspired neural network for conversational emotion recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 13270–13278.

Li, Q.; Wang, B.; and Melucci, M. 2019. CNM: An Interpretable Complex-valued Network for Matching. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4139–4148.

Majumder, N.; Poria, S.; Hazarika, D.; Mihalcea, R.; Gelbukh, A.; and Cambria, E. 2019. Dialoguernn: An attentive rnn for emotion detection in conversations. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, 6818–6825.

Mönning, N. 2019. *Deep Complex-Valued Neural Networks for Natural Language Processing*. Ph.D. thesis, University of York.

Nielsen, M. A.; and Chuang, I. 2002. Quantum computation and quantum information.

Pang, B.; and Lee, L. 2004. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. *arXiv preprint cs/0409058*.

Pang, B.; and Lee, L. 2005. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. *arXiv preprint cs/0506075*.

Poria, S.; Hazarika, D.; Majumder, N.; Naik, G.; Cambria, E.; and Mihalcea, R. 2018. Meld: A multimodal multi-party dataset for emotion recognition in conversations. *arXiv preprint arXiv:1810.02508*.

Rakhlin, A. 2016. Convolutional Neural Networks for Sentence Classification. *GitHub*.

Shi, J.; Li, Z.; Lai, W.; Li, F.; Shi, R.; Feng, Y.; and Zhang, S. 2021. Two end-to-end quantum-inspired deep neural networks for text classification. *IEEE Transactions on Knowledge and Data Engineering*.

Sordoni, A.; Nie, J.-Y.; and Bengio, Y. 2013. Modeling term dependencies with quantum language models for ir. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, 653–662.

Süli, E.; and Mayers, D. F. 2003. *An introduction to numerical analysis*. Cambridge university press.

Tsai, Y.-H. H.; Bai, S.; Liang, P. P.; Kolter, J. Z.; Morency, L.-P.; and Salakhutdinov, R. 2019. Multimodal transformer for unaligned multimodal language sequences. In *Proceedings of the conference. Association for Computational Linguistics. Meeting*, volume 2019, 6558. NIH Public Access.

Wang, Y.-J.; and Lin, C.-T. 1998. Runge-Kutta neural network for identification of dynamical systems in high accuracy. *IEEE Transactions on Neural Networks*, 9(2): 294–307.

Wiebe, J.; Wilson, T.; and Cardie, C. 2005. Annotating expressions of opinions and emotions in language. *Language resources and evaluation*, 39(2): 165–210.

Wisdom, S.; Powers, T.; Hershey, J.; Le Roux, J.; and Atlas, L. 2016. Full-capacity unitary recurrent neural networks. *Advances in neural information processing systems*, 29.

Zhang, J.; Zhang, P.; Kong, B.; Wei, J.; and Jiang, X. 2021. Continuous self-attention models with neural ODE networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 14393–14401.

Zhang, L.; Zhang, P.; Ma, X.; Gu, S.; Su, Z.; and Song, D. 2019. A generalized language model in tensor space. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 7450–7458.

Zhang, P.; Hui, W.; Wang, B.; Zhao, D.; Song, D.; Lioma, C.; and Simonsen, J. G. 2022. Complex-valued Neural Network-based Quantum Language Models. *ACM Transactions on Information Systems (TOIS)*, 40(4): 1–31.

Zhang, P.; Niu, J.; Su, Z.; Wang, B.; Ma, L.; and Song, D. 2018a. End-to-end quantum-like language models with application to question answering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32.

Zhang, P.; Su, Z.; Zhang, L.; Wang, B.; and Song, D. 2018b. A quantum many-body wave function inspired language modeling approach. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, 1303–1312.

Zhu, M.; Chang, B.; and Fu, C. 2022. Convolutional neural networks combined with Runge–Kutta methods. *Neural Computing and Applications*, 1–15.