Application of the Quantum Kernel Algorithm on the Particle Identification at the BESIII Experiment

Teng Li[1], Zhipeng Yao[1], Xingtao Huang[1], Jiaheng Zou[2], Tao Lin[2], Weidong Li[2]

[1]Institute of Frontier and Interdisciplinary Science, Shandong University, Qingdao, China
[2]Institute of High Energy Physics, CAS, Beijing, China

E-mail: tengli@sdu.edu.cn

Abstract. Particle identification is one of most fundamental tools in various particle physics experiments. For the BESIII experiment on the BEPCII, the realization of numerous physical goals heavily relies on advanced particle identification algorithms. In recent years, the emerging of quantum machine learning could potentially arm particle physics experiments with a powerful new toolbox. In this work, targeting at the muon/pion discrimination problem at BESIII, we have developed a quantum SVM classifier for the Noisy Intermediate-Scale Quantum (NISQ) device. By studying and optimizing various encoding circuits, the quantum SVM trained with the BESIII MC data shows comparable discrimination power than other traditional machine learning models. This has demonstrated the potential of using quantum machine learning techniques to form a new approach for particle identification in particle physics experiments. In this talk, we present the application of the quantum SVM for particle identification at BESIII, and demonstrate how to construct and optimize the quantum kernel. Furthermore, we present results obtained from the noisy simulator as well as the small-scale hardware to show the potential advantage of the quantum SVM algorithm.

Keywords: Particle Identification, Quantum Computing, Quantum Machine Learning, Quantum Support Vector Machine

1. Introduction

In collider experiments, particle identification (PID) is usually one of the most fundamental tools for physics analysis studies. In the past two decades, machine learning algorithms have gradually become one of the mainstream methods for PID. One of the most remarkable merits of machine learning algorithms is the capability to extract effective features among the extensive number of input variables, thus usually offering superior performance. In recent years, the emerging of quantum machine learning [1, 2] has armed particle physics with a powerful new toolbox. A number of previous studies [3, 4, 5] show that quantum machine learning has the potential to outperform traditional machine learning in a few aspects. For instance, the quantum kernel algorithm supplies a new set of kernels for SVM that are hard be be simulated classically by encoding the input data into high dimensional Hilbert spaces, which could outperform traditional Gaussian kernels in some cases. Previous study [6] showed that the quantum SVM equipped with a quantum random access memory (qRAM) is able to achieve a exponential speedup comparing to the training of the classical SVM.

Therefore, it's interesting to study quantum algorithms to evaluate the feasibility of applying quantum machine learning to particle physics experiments. In this work, targeting at the $\mu^{\pm}/\pi^{\pm}$ discrimination, one of the most difficult PID problems at BESIII, we have developed a classifier based on a quantum kernel algorithm, and tested the model using the IBM quantum simulator [7] and OriginQ Wuyuan hardware [8] under the NISQ assumption.

2. Quantum Kernel Algorithm

2.1. SVM with Classical Kernels

SVM is one of the most widely used and well studied supervised machine learning algorithms for classification problems. The dual formulation of SVM can be written as:

$$\begin{aligned} &maximize \quad L(\vec{\alpha}) = \sum_{i=1}^{N} y_i \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j K(\vec{x}_i, \vec{x}_j) \\ &subject\ to \quad \sum_{i=1}^{N} \alpha_i y_i = 0 \ \ and \ \ 0 \leq \alpha_i \leq C, \forall i = 1, 2, ..., N \end{aligned} \tag{1}$$

The term $K(\vec{x}_i, \vec{x}_j) = \Phi(\vec{x}_i)^T \Phi(\vec{x}_j)$ denotes the kernel matrix, where the $\Phi(\vec{x})$ defines a kernel function that projects original data points to a high dimensional feature space to implement the non-linearity. The form of the kernel function is not explicitly regulated by the SVM, which only takes the kernel matrix to optimize the problem.

2.2. Quantum Kernel

The entries in the kernel matrix are essentially the similarity between data points in the high dimensional feature space. In quantum computers, the kernel matrix can be calculated by measuring the inner product $|\langle \Phi(\vec{x}_i) | \Phi(\vec{x}_j) \rangle|^2$ of the quantum states representing the data points [9].

The procedure encoding the data vector $\vec{x}$ into a quantum state $|\Phi(\vec{x})\rangle$ is performed by the state preparation circuit, formally defined as:

$$|\Phi(\vec{x}_i)\rangle = \mathcal{U}_{\Phi(\vec{x}_i)} \mid 0^{\otimes n} \rangle = H^{\otimes n} U_{\Phi(\vec{x}_i)} H^{\otimes n} U_{\Phi(\vec{x}_i)} \mid 0^{\otimes n} \rangle \tag{2}$$

The $\mathcal{U}_{\Phi(\vec{x}_i)}$ denotes the encoding circuit, which is usually a multiple times repetition of the feature maps $U_{\Phi(\vec{x}_i)}$ preposed with parallel Hadamad gates $H^{\otimes n}$. On top of the state preparation circuit, the inner product of two quantum states can be estimated using a SWAP test [10], or an alternative circuit that uses only half the qubits based on adjunt feature maps [11], shown in figure 1.

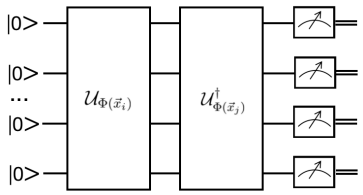


Figure 1: The quantum circuit that uses inverted encoding circuit to estimate the inner product, which can be estimated by measuring the frequency of $\mid 0^{\otimes n} \rangle$ of the output.

Each implementation of the quantum feature map uniquely defines a kernel function. Their competitiveness decides the discrimination power of the quantum SVM. In our study, we mainly used the Pauli feature maps proposed in [11], which are formally defined as:

$$U_{\Phi(\vec{x})} = exp(i \sum_{S \in [n]} \Phi_s(\vec{x}) \prod_{i \in S} P_i) \tag{3}$$

In equation 3, the $P_i \in \{I, X, Y, Z\}$ denotes the Pauli matrices, while the index S describes the connectivity between qubits. For simplicity, here we only use one- and two-qubit logical operations.

3. Muon and Pion Discrimination for BESIII

3.1. PID system in BESIII

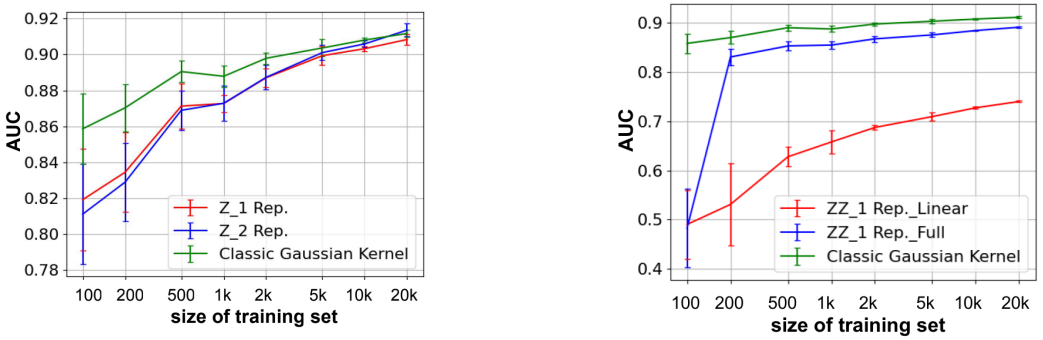
The BESIII detector on the BEPCII is designed for multiple physics objectives, such as the study of the electroweak and strong interactions in the τ-Charm energy region and the search for new physics etc. The BESIII detector consists of a beam pipe, a main drift chamber (MDC), a time of flight counter (TOF), a CsI(Tl) electromagnetic calorimeter (EMC), a superconducting solenoid magnet and a muon counter (MUC), from the inner layer to the outter layer. The BESIII detector can provide the following information for PID:

- The ionization energy loss of charged tracks per unit pathlength in the MDC (dE/dx).
- The time of flight measured by the TOF counter combined with trajecory and momentum measured by the MDC provides additional discrimination power for particles with different masses.
- Shape of the electromagnetic or hadronic showers in the EMC.
- Hit pattern in the MUC.

3.2. Data Sample and Input Features

Table 1: AUC scores obtained with various simulated feature maps.

Circuit	Rep.	Entanglement	Test set AUC	Training set AUC
X	2	none	0.90834±0.0030	0.91658±0.0021
	3		0.91238±0.0036	0.93055±0.0027
Z	1	none	0.90834±0.0030	0.91658±0.0020
	2		0.91238±0.0036	0.93055±0.0027
	3		0.89240±0.0036	0.90949±0.0009
XX	2	linear	0.73146±0.0037	0.84744±0.0017
		full	0.86332±0.0052	0.99887±0.0001
	3	linear	0.73198±0.0048	0.93766±0.0009
		full	0.72999±0.0047	0.99970±0.0002
ZZ	1	linear	0.73146±0.0037	0.84744±0.0025
		full	0.86332±0.0052	0.99887±0.0002
	2	linear	0.73198±0.0048	0.93767±0.0009
		full	0.72999±0.0047	0.99970±0.0002
	3	linear	0.67960±0.0040	0.88481±0.0004
		full	0.62707±0.0035	0.99964±0.0001
X, XX	2	linear	0.87201±0.0043	0.97902±0.0012
		full	0.88359±0.0021	0.99974±0.0001
	3	linear	0.84779±0.0010	0.99364±0.0002
		full	0.80892±0.0020	1.00000±0.0000



(a) (b)

Figure 2: The AUC values versus the size of the training sample for four quantum feature maps in comparison with the classical Gaussian kernel. Each point is averaged from 20 independent datasets.

The training and testing data sample are simulated using the BOSS software [12] of BESIII. To reduce statistical fluctuation, the data sample is divided into 20 datasets, each containing 20,000 training and 10,000 testing tracks. Since the encoding circuits are based on rotation of angles decided by the feature values, all the features are scaled to the $[0, 2]$ range before being encoded into the quantum circuits. The selected nine features include the momentum and the angle of the reconstructed track, the PID likelyhood given by the dE/dx and the TOF, the ratios of the energy deposit in the EMC, as well as the penetration depth in the MUC.

3.3. Simulation Results

Firstly, we performed a broad search over a large number of Pauli feature maps for the optimal configuration, using the IBM quantum simulator Qiskit 0.26.0. For each feature map we calculate the mean and standard deviation of the Area Under the Curve (AUC) values over 20 datasets. The results of the simulation are shown in table 1. The second column means the number of repetition times of the encoding circuits, so that the data is loaded into the quantum states more than once to enhance expressiveness. The entanglement methods include the linear method, for which only the neighbor qubits are entangled, and the full method, for which every two qubits are entangled.

It is clear that the performances of various feature maps vary greatly. A few feature maps based on pure X or Z rotations without entanglements show the best discrimination power on the test sets. Other circuits, in particular those with complicated entanglements show much stronger discrimination power on the training sets, but are prone to overfit, thus showing poor performance discriminating unseen data. This hints that these quantum kernels estimated in the very high dimensional Hilbert space is able to capture much more complicated structures in the data.

The influence of the training sample size is also investigated. Figure 2 shows the AUC values of four different quantum feature maps versus the size of the training sample, in comparison with the classical SVM with the Gaussian kernel. The performance increase with the training sample size, and the Z feature maps start to show close or better performance after the training sample is larger than 10,000.

To compare with the classical machine learning, three baseline models, the classical SVM with Gaussian kernel, the BDT and the Multilayer Perceptron neural network are trained and tested with the same dataset. The ROC curves of these models are shown in Figure 3, from which we can observe that, the simulated quantum models show close performance with the classical models.

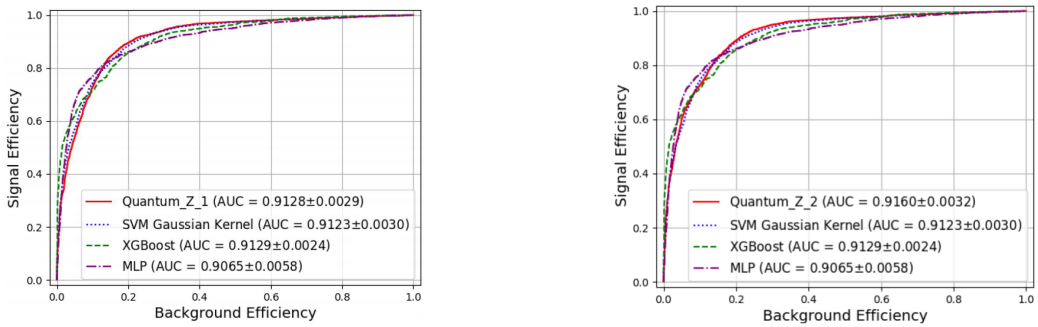


Figure 3: The ROC curves of the Z feature maps (1 repetition on the left and 2 repetitions on the right), in comparison with the classical SVM, BDT and MLP neural network.

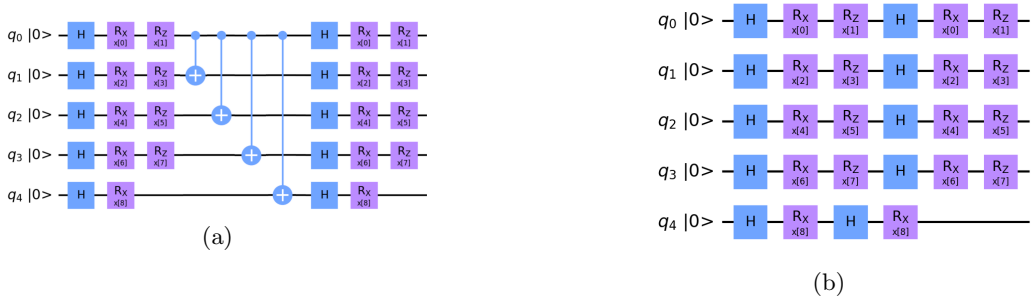


Figure 4: (a) The customized Feature Map A that encodes the nine features using both Pauli X and Z rotation gates, plus four controlled-NOT gates that entangle the first qubit with the other four qubits. (b) The customized Featue Map B without any entanglements.

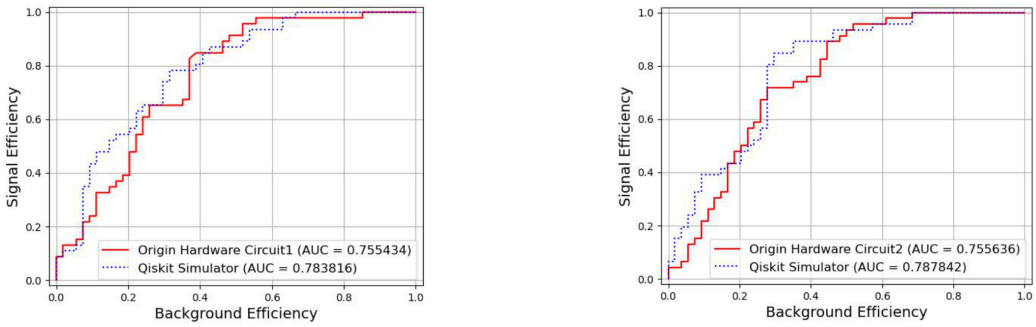
3.4. Results on Noisy Simulator and Quantum Hardware

Based on the studies with simulated quantum devices, it is now of great interests to see how the quantum SVM model perform on real quantum devices, with the disturbance of noises. To make use of the quantum hardware, the circuits need to be compressed due to hardware limitation. Therefore, we designed and optimized two circuits that encode two features into each qubit, as shown in Figure 4a and 4b . Under the Qiskit noiseless simulator, the AUC of feature map A reaches 0.90373±0.0024, while that of feature map B reaches 0.91029±0.0023 on the test tests, averaged over 20 datasets.

In order to evaluate the impact of the hardware noise on the model, we compared the results from the noiseless Qiskit simulator as well as from the OriginQ Wuyuan system, which is a 6-qubit quantum computer based on super-conducting technology. Due to the limited access time available for testing, we made five small datasets from the datasets used in section 4.2, each containing 100 training tracks and 100 test tracks with balanced number of $\mu^{\pm}$ and $\pi^{\pm}$. The comparison of the ROC curves and the AUC values obtained in the three tests are shown in Figure 5. The ROC comparison shows that the noise compromises the performance, but at a reasonable level.

4. Conclusion

In this work, targeted at the μ/π identification problem at the BESIII experiment, we studied the capabilities of the quantum SVM model. The simulation show that a few feature maps



(a) (b)

Figure 5: The ROC curves (on the test set) of the QSVM model on the OriginQ hardware, in comparison with the noiseless simulator.

using pure X or Z rotation gates show quite comparable performance with the classical machine learning models. Other feature maps equiped with complex entanglements show strong sign of over-fitting over the PID data, hinting that complicated feature maps are able to describe much more complicated structures in the data. We then run two compressed feature maps on the OriginQ Wuyuan hardware, in comparison with the noiseless simulator. The results show that the quantum hardware at small scale is able to achieve comparable results with simulators.

As a preliminary exploration on applying quantum machine learning to the data analysis in high energy physics experiments, this study shows the possibility of applying quantum computing on the processing of real physical data, and we foresee more applications with the development of the quantum hardware.

References

[1] P. Rebentrost, M. Mohseni, S. Lloyd, Phys. Rev. Lett. 113 (2014) 130503.
[2] M. Schuld, A. Bocharov, K. M. Svore, N. Wiebe, Phys. Rev. A 101 (2020) 032308.
[3] S. L. Wu, et al. (2021). arXiv:2104.05059.
[4] V. Belis, S. González-Castillo, C. Reissel, S. Vallecorsa, E. F. Combarro, G. Dissertori, F. Reiter, in: 25th International Conference on Computing in High-Energy and Nuclear Physics, 2021. arXiv:2104.07692.
[5] A. Blance, M. Spannowsky, Journal of High Energy Physics 2021 (2) (Feb 2021).
[6] P. Rebentrost, M. Mohseni, S. Lloyd, Phys. Rev. Lett. 113 (2014) 130503.
[7] A. Cross, The IBM Q experience and QISKit open-source quantum computing software, in: APS March Meeting Abstracts, Vol. 2018 of APS Meeting Abstracts, 2018, p. L58.003.
[8] OriginQ Wuyuan, `https://qcloud.originqc.com.cn/`, [Online; accessed 2-2-2022] (2022).
[9] M. Schuld, N. Killoran, Physical Review Letters 122 (4) (Feb 2019).
[10] H. Buhrman, R. Cleve, J. Watrous, R. de Wolf, Physical Review Letters 87 (16) (Sep 2001).
[11] V. Havlíček, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, J. M. Gambetta, Nature 567 (7747) (2019) 209–212.
[12] W. D. Li, et al., The offline software for the BESIII experiment, in: International Conference on Computing in High-Energy and Nuclear Physics, 2006.