Quantum adversarial metric learning model based on triplet loss function

Yan-Yan Hou [1,2] · **Jian Li** [3] · **Xiu-Bo Chen** [4,5] · **Chong-Qiang Ye** [1] ·

Abstract Metric learning plays an essential role in image analysis and classification, and it has attracted more and more attention. In this paper, we propose a quantum adversarial metric learning (QAML) model based on the triplet loss function, where samples are embedded into the high-dimensional Hilbert space and the optimal metric is obtained by minimizing the triplet loss function. The QAML model employs entanglement and interference to build superposition states for triplet samples so that only one parameterized quantum circuit is needed to calculate sample distances, which reduces the demand for quantum resources. Considering the QAML model is fragile to adversarial attacks, an adversarial sample generation strategy is designed based on the quantum gradient ascent method, effectively improving the robustness against the functional adversarial attack. Simulation results show that the QAML model can effectively distinguish samples of MNIST and Iris datasets and has higher ϵ-robustness accuracy over the general quantum metric learning. The QAML model is a fundamental research problem of machine learning. As a subroutine of classification and clustering tasks, the QAML model opens an avenue for exploring quantum advantages in machine learning.

✉Jian Li

Lijian@bupt.edu.cn

[1]School of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing 100876, China.

[2]College of Information Science and Engineering, ZaoZhuang University, ZaoZhuang Shandong 277160, China.

[3]School of Cyberspace Security Security, Beijing University of Posts Telecommunications, Beijing 100876, China.

[4]Information Security Center, State Key Laboratory of Networking and Switching Technology, Beijing University of Post and Telecommunications, Beijing 100876, China.

[5]GuiZhou University, Guizhou Provincial Key Laboratory of Public Big Data, Guizhou Guiyang, 550025, China.

1 Introduction

Machine learning has developed rapidly in recent years and is widely used in artificial intelligence and big data fields. Quantum computing can efficiently process data in exponentially sizeable Hilbert space and is expected to achieve dramatic speedups in solving some classical computational problems. Quantum machine learning, as the interplay between machine learning and quantum physics, brings unprecedented promise to both disciplines. On the one hand, machine learning methods have been extended to quantum world and applied to the data analysis in quantum physics [1]. On the other hand, quantum machine learning exploits quantum properties, such as entanglement and superposition, to revolutionize classical machine learning algorithms and achieves computational advantages over classical algorithms [2]. Metric Learning is the core problem of some machine learning tasks [3], such as k-nearest neighbor, support vector machines, radial basis function networks, and k-means clustering. Its core work is to construct an appropriate distance metric that maximizes the similarities of samples of the same class and minimizes the similarities of samples from different classes. Linear and nonlinear methods can be used to implement metric learning. In general, linear models have a limited number of parameters and are unsuitable for characterizing high-order features of samples. Recently, neural networks have been adopted to establish nonlinear metric learning models, and promising results have been achieved in face recognition and feature matching.

Classical metric learning models usually extract low-dimensional representations of samples, which will lose some details of samples. Quantum states are in high-dimensional Hilbert spaces, and their dimensions grow exponentially with the number of qubits. This quantum enables quantum models to learn high-dimensional representations of samples without explicitly invoking a kernel function. A parameterized quantum circuit is used to map samples in high-dimensional Hilbert space. The optimal metric model is obtained by optimizing the loss function based on Hilbert-Schmidt distances. With the increase of the the dimension, this speed-up advantage will become more and more pronounced, and it is expected to achieve exponential growth in computing speeds. In recent years, researchers began to study how to adopt quantum methods to implement metric learning. Lloyd [4] firstly proposed a quantum metric learning model based on hybrid quantum-classical algorithms. A parameterized quantum circuit is used to map samples in high-dimensional Hilbert space. The optimal metric model is obtained by optimizing the loss function based on Hilbert-Schmidt distances. This model achieves better effects in classification tasks. Nhat [5] introduced quantum explicit and implicit metric learning approaches from the perspective of whether the target space is known or not. The research establishes the relationship between quantum metric learning and other quantum supervised learning models. The above two algorithms mainly focus on classification tasks. Metric learning is a fundamental problem in machine learning, which can be applied not only to classification but also to clustering, face recognition, and other issues. In our research, we are devoted to constructing a quantum metric learning model that can serve various machine learning tasks.

Angular distance is a vital metric that quantifies the included angle between normalized samples [6]. Angular distance focuses on the difference in the direction of samples and is more robust to the variation of local feature [7], [8]. Considering

the similarities between angular distances of classical data and inner products of quantum states, we design a quantum adversarial metric learning (QAML) model based on inner product distances, which is more suitable for image-related tasks. Unlike other quantum metric learning models, the QAML model maps samples from different classes into quantum superposition states and utilizes simple interface circuits to compute metric distances for multiple sample pairs in parallel. Furthermore, quantum systems in high-dimensional Hilbert space have counterintuitive geometrical properties [9]. The QAML model using only natural samples is vulnerable to adversarial attacks, under which some samples are closer to the false class, so the model is easy to make wrong judgements [10]. To solve this issue, we construct adversarial samples based on natural samples. The model's robustness is improved by the alternative train of natural and adversarial samples. Our work has two main contributions:(i) We explore a quantum method to compute the triplet loss function, which utilizes quantum superposition states to calculate sample distances in parallel and reduce the demand for quantum resources. (ii) We design an adversarial samples generation strategy based on the quantum gradient ascent, and the robustness of the QAML model is significantly improved by alternatively training generated adversarial samples and natural samples. Simulation results show that the QAML model separates samples by a larger margin and has better robustness for functional adversarial attacks than general quantum metric learning models.

The paper is organized as follows. Section 2 gives the basic method of the QAML model. Section 3 verifies the performances of the QAML model. Finally, we get a conclusion and discuss the future research directions.

2 Quantum adversarial metric learning

2.1 Preliminary theory

Triplet loss function is a widely used strategy for metric learning [11], commonly used in image retrieval and face recognition. A triplet set (x_i^a, x_i^p, x_i^n) consists of three samples from two classes, where anchor sample x_i^a and positive sample x_i^p belong to the same class, and negative sample x_i^n comes from another class. The goal of metric learning based on triplet loss function is to find the optimal embedded representation space, in which positive sample pairs (x_i^a, x_i^p) are pulled together and negative sample pairs (x_i^a, x_i^n) are pushed away. Fig.1 shows sample space change in the metric learning process. As we can see, samples from different classes become linearly separable through metric learning. Fig.2 shows the schematic of the metric learning model based on triplet loss function. Firstly, the model prepares multiple triplet sets, and one triplet set (x_i^a, x_i^p, x_i^n) is sent to convolutional neural networks (CNN), where three CNN with the same structure and parameters are needed. Each CNN acts on one sample of the triplet set to extract its features. The triplet loss function is obtained by computing metric distances for multiple sample pairs of triplet sets. In the learning process, the optimal parameters of CNN are obtained by minimizing the triplet loss function. Let one batch samples include N_1 triplet sets. The triplet loss function is

$$L = \frac{1}{N_1} \sum_{i=1}^{N_1} [D(g(x_i^a), g(x_i^p)) - D(g(x_i^a), g(x_i^n)) + \mu]_+, \tag{1}$$
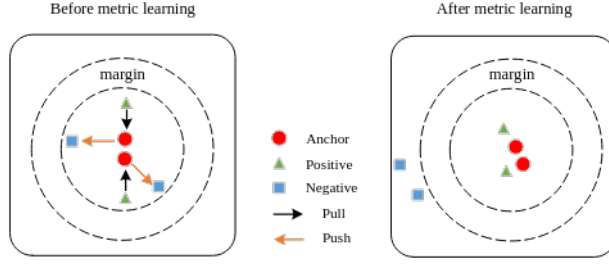
Fig. 1 Sample space change in metric learning process. Before metric learning, the distances between negative sample pairs are smaller, and samples from different classes are difficult to separate by linear functions. After metric learning, the distances between negative sample pairs become larger, and a large margin separates samples from different classes. Linear functions can easily separate positive and negative samples.
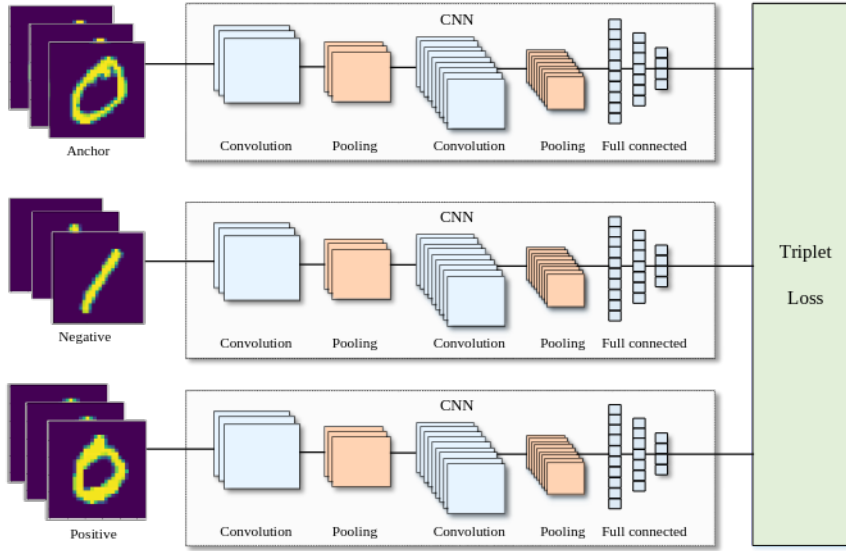

Fig. 2 The schematic of the metric learning model based on triplet loss function. A triplet set includes an anchor sample, a positive sample, and a negative sample. The input consists of a batch of triplet sets, and only one triplet set serves as input in each iteration. Three CNN with the same structure and parameters are used to map the triplet set into the embedded representation space. CNN, consisting of multiple convolutions, pooling, and fully connected layers, is responsible for extracting the features of samples. The triplet loss function is further constructed based on the extracted features.

where $g(\cdot)$ represents the function mapping input samples to the embedded representation space, $D(\cdot, \cdot)$ denotes the distance between a sample pair in the embedded representation space, and $[\,\cdot\,,\,\cdot\,]_+ = max(0,\,\cdot\,)$ represents the hinge loss function. The goal of metric learning is to learn a metric that makes the distances between negative sample pairs greater than the distance between the corresponding posi-

tive sample pairs and satisfies the specified margin $\mu \in \mathbb{R}^+$ [6]. In the triplet loss function, $D(g(x_i^a), g(x_i^p))$ penalizes the positive sample pair (x_i^a, x_i^p) that is too far apart, and $D(g(x_i^a), g(x_i^n))$ penalizes the negative sample pair (x_i^a, x_i^p) whose distance is less than the margin μ.

Metric learning can adopt various distance metric methods. Angular distance metric is robust to image illumination and contrast variation [7], which is an efficient way for metric learning tasks. In this method, samples need to be normalized to unit vectors in advance. The distance between a positive sample pair is

$$D(g(x_i^a), g(x_i^p)) = 1 - \frac{|g(x_i^a) \cdot g(x_i^p)|}{||g(x_i^a)||_2 ||g(x_i^p)||_2}, \tag{2}$$

where $|\ |$ and $||\ ||_2$ represent l_1-norm and l_2-norm, respectively, and $\cdot$ denotes the inner product operation for two vectors. The distance between negative sample pairs can be calculated in the same way.

2.2 Framework of quantum metric learning model

For most machine learning tasks, it is often challenging to adopt simple linear functions to distinguish samples of different classes. According to kernel theory [12], samples in high-dimensional feature space have better distinguishability. Classical machine learning algorithms usually adopt kernel methods to map samples to high-dimensional feature space, where the mapped samples can be separated by simple linear functions. Quantum states with n-qubits are in 2^n-dimensional Hilbert space, where quantum systems characterize the nonlinear features of data and efficiently process data through a series of linear unitary operations.

In the QAML model, samples should be firstly mapped into quantum systems by qubit encoding. The Hilbert space after encoding usually does not correspond to the optimal space for separating samples of different classes. To search for the optimal Hilbert space, the QAML model performs parameterized quantum circuits $W(\theta)$ on the encoded states [13]. As different variable parameters θ correspond to different mapping spaces, we can search the optimal space by modifying parameters $\theta = (\theta_1^1, ..., \theta_i^j)$. As long as $W(\theta)$ has strong expressivity, we can find the optimal Hilbert space by optimizing the loss function of metric learning [14,15]. $W(\theta)$ with different structures and layers have different expressivity. The more layers $W(\theta)$ has, the stronger the expressivity, and the easier it is to find the optimal metric space.

The classical metric learning model based on triplet loss function requires three identical CNN to map triplet sets (x_i^a, x_i^p, x_i^n) into the novel Hilbert space. To reduce the demand for quantum resources, we construct a quantum superposition state to represent one triplet set so that a triplet set only needs one $W(\theta)$ to transform it into Hilbert space. The core work of the building loss function is to compute inner products between sample pairs, but $W(\theta)$ and subsequent conjugate operation $W^\dagger(\theta)$ counteract each other's effects. To solve this issue, we add a repeated encoding operation after $W(\theta)$. It is worth mentioning that the repeated encoding operation is also conducive to the construction of high-dimensional features of samples.

The QAML model is mathematically represented as the minimization of the loss function with respect to the parameters θ. The triplet loss function consists of

metric distances for positive and negative sample pairs, so the kernel work of the QAML model is constructing the metric distances for sample pairs in the transformed Hilbert space. The mapping samples $h(x_i^a)/||h(x_i^a)||_2$ and $h(x_i^p)/||h(x_i^p)||_2$ of Equ.2 are replaced by the quantum states of x_i^a and x_i^p, then the second term of Equ.2 is converted to the inner product between quantum states of the positive sample pair (x_i^a, x_i^p), which can be got by the method of the Hadamard classifier [12]. The triplet loss function can be viewed as the weighted sum of the inner product of sample pairs (x_i^a, x_i^p) and the inner product of sample pairs (x_i^a, x_i^n). With the help of ancilla registers, the triplet set can be prepared in superposition states form. According to the entanglement property of superposition states, the triplet loss function can be implemented with one parameterized quantum circuit. Then, the triplet loss function value is transmitted to a classical optimizer, and parameters are optimized until the optimal metric is obtained. The QAML model constructs adversarial samples according to the gradient of natural samples and trains alternatively natural and adversarial samples to improve the model's robustness against adversarial attacks. The schematic of the QAML model is shown in Fig.3.

References

1. I. Cong, S. Choi, M.D. Lukin, Nature Physics **15**(12), 1273 (2019)
2. M. Benedetti, E. Lloyd, S. Sack, M. Fiorentini, Quantum Science and Technology **4**(4), 043001 (2019)
3. S. Chen, C. Gong, J. Yang, X. Li, Y. Wei, J. Li, arXiv preprint arXiv:1802.03170 (2018)
4. S. Lloyd, M. Schuld, A. Ijaz, J. Izaac, N. Killoran, arXiv preprint arXiv:2001.03622 (2020)
5. N.A. Nghiem, S.Y.C. Chen, T.C. Wei, arXiv preprint arXiv:2010.13186 (2020)
6. C. Mao, Z. Zhong, J. Yang, C. Vondrick, B. Ray, Advances in Neural Information Processing Systems **32** (2019)
7. J. Wang, F. Zhou, S. Wen, X. Liu, Y. Lin, in *Proceedings of the IEEE international conference on computer vision* (2017), pp. 2593–2601
8. Y. Duan, W. Zheng, X. Lin, J. Lu, J. Zhou, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2018), pp. 2780–2789
9. N. Liu, P. Wittek, Physical Review A **101**(6), 062331 (2020)
10. A. Madry, A. Makelov, L. Schmidt, D. Tsipras, A. Vladu, arXiv preprint arXiv:1706.06083 (2017)
11. R. Salakhutdinov, G. Hinton, in *Artificial Intelligence and Statistics* (PMLR, 2007), pp. 412–419
12. C. Blank, D.K. Park, J.K.K. Rhee, F. Petruccione, npj Quantum Information **6**(1), 1 (2020)
13. E. Grant, M. Benedetti, S. Cao, A. Hallam, J. Lockhart, V. Stojevic, A.G. Green, S. Severini, npj Quantum Information **4**(1), 1 (2018)
14. A. Pérez-Salinas, A. Cervera-Lierta, E. Gil-Fuster, J.I. Latorre, Quantum **4**, 226 (2020)
15. M. Schuld, R. Sweke, J.J. Meyer, Physical Review A **103**(3), 032430 (2021)
16. C. Zoufal, A. Lucchi, S. Woerner, npj Quantum Information **5**(1), 1 (2019)
17. A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J.M. Chow, J.M. Gambetta, Nature **549**(7671), 242 (2017)
18. T. Miyato, S.i. Maeda, M. Koyama, S. Ishii, IEEE transactions on pattern analysis and machine intelligence **41**(8), 1979 (2018)
19. A. Kurakin, I. Goodfellow, S. Bengio, arXiv preprint arXiv:1611.01236 (2016)
20. J.R. McClean, M.E. Kimchi-Schwartz, J. Carter, W.A. De Jong, Physical Review A **95**(4), 042308 (2017)
21. G.E. Crooks, arXiv preprint arXiv:1905.13311 (2019)
22. M. Schuld, V. Bergholm, C. Gogolin, J. Izaac, N. Killoran, Physical Review A **99**(3), 032331 (2019)
23. K. Mitarai, M. Negoro, M. Kitagawa, K. Fujii, Physical Review A **98**(3), 032309 (2018)

(a) A schematic of the quantum adversarial metric learning

(b) Dimension reduction circuit (c) Quantum-classical loop for adversarial metric learning
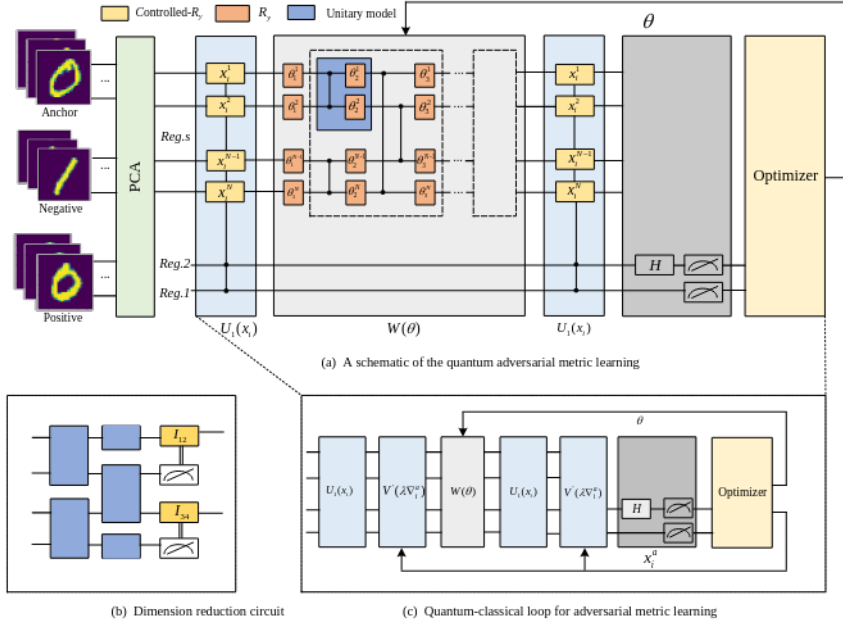
Fig. 3 Overview of quantum adversarial metric learning (QAML) model. Panel (a) shows the framework of quantum adversarial metric learning. $Reg.s$ is the sample register that stores triplet sets, and $Reg.1$ and $Reg.2$ are ancilla registers used distinguishing different samples. The model firstly adopts principal component analysis (PCA) to reduce the input dimension. Subsequently, anchor, negative and positive samples are encoded into a quantum superposition state by controlled qubit encoding. The transformation of Hilbert space is implemented by parameterized quantum circuit $W(\theta)$ and the subsequent qubit encoding $U_1(x_i)$. Finally, Hadamard and measurement operations act on ancilla registers to simultaneously compute the inner products for the positive and negative sample pairs, and the triplet loss function is further obtained. In each iteration, the parameters θ are updated by optimizing the triplet loss function with a classical optimizer. Panel (b) shows the quantum dimension reduction circuit to reduce the number of output qubits. In each module, only one qubit is measured, and the controlled unitary based on its measurement result acts on another qubit. Panel (c) shows another case of the QAML model, where adversarial samples are built and added to the training process. $V^{'}(\lambda\nabla_i^a)$ is the unitary operation based on the gradient of anchor sample x_i^a and acts on the encoded quantum states to produce its adversarial sample. In the QAML model training process, natural and adversarial samples alternatively serve as input.

24. V. Bergholm, J. Izaac, M. Schuld, C. Gogolin, M.S. Alam, S. Ahmed, J.M. Arrazola, C. Blank, A. Delgado, S. Jahangiri, et al., arXiv preprint arXiv:1811.04968 (2018)
25. M.C. Mukkamala, M. Hein, in *International conference on machine learning* (PMLR, 2017), pp. 2545–2553
26. J. Guan, W. Fang, M. Ying, CoRR (2020)