arXiv:2302.02994v1 [quant-ph] 6 Feb 2023

多分类 SWAP 测试分类器

S. M. Pillay[1], I. Sinayskiy[2,3],E. Jembere[1],和 F. Petruccione[2,3,4]

[1]夸祖鲁-纳塔尔大学数学、统计学与计算机科学学院，南非德班 4001[2]夸祖鲁-纳塔尔大学化学与物理学院，南非德班 4001[3]国家理论与计算科学研究所（NITheCS），南非斯泰伦博斯。[4]斯泰伦博斯大学数据科学与计算思维学院，南非斯泰伦博斯 7604。

摘要

多分类问题是科研和工业等众多领域的基础。为解决多分类问题，可以采用一对一或一对多等启发式策略。然而，这些策略要求开发的二元分类模型数量随类别数增加而增长。近年来，量子机器学习领域发展出了多分类量子分类器，通过学习数据与一组标签态之间的映射，避免了这种增长。本文提出了首个多分类 SWAP 测试分类器，其设计受到二元版本及近期研究中标签态应用的启发。采用该分类器可以避免开发多个模型的成本。与以往工作不同，所需量子比特数、测量策略以及所用电路的拓扑结构均不随类别数变化。此外，与其他多分类量子分类器架构不同，仅重构单个量子比特的状态便可获得足够的信息，以完成多分类任务。解析结果和数值模拟均表明，该分类器不仅适用于各种分类问题，而且在某些噪声条件下也具有鲁棒性。

1 引言

量子机器学习中备受关注的一类问题是分类。人们开发了量子核方法，利用量子计算机在大型希尔伯特空间中高效执行运算的能力来解决这些问题 [1, 2, 3]。SWAP 测试分类器是一种量子核方法，它能并行估算给定测试数据与所有训练数据之间核值的加权和 [3]。这个和是对一种以特定格式编码测试数据、训练数据及其各自标签的量子态执行 SWAP-Test 的结果；该测试仅涉及一次量子比特测量[4]。与其他量子核方法相比，这种方法无需在每次估算核值时都准备测试数据。此外，无需借助经典子程序进行训练，即可获得利用该加权和的决策规则。

SWAP 测试分类器是一种二元分类器，只能将数据分为两类。然而，图像识别、自然语言处理等大多数现实问题都需要能够处理两类以上类别的多类分类器。要使用 SWAP 测试分类器进行多类分类，可以采用多个二元分类模型，并使用一对多策略等启发式策略 [5]。但是，这些策略存在局限性，而且所需的二元分类器数量会随着类别数量的增加而增加。

近期量子机器学习领域的一些研究开发出了多类别量子

分类器，能够规避这种增长 [6, 7]。实现方式是借助经典子程序训练参数化电路，将经典数据映射到希尔伯特空间中彼此分离的区域。在一些方法中，这些区域以标签态为中心；标签态是一组与各类别相对应的量子态[6]。在另一些方法中，则通过训练构建这些区域或簇，使它们在希尔伯特空间中尽可能彼此远离 [7]。在这两种方法中，标签态的选择与数据编码策略密不可分，反之亦然。按照提议，若选择希尔伯特空间中的正交态作为标签态，那么数据编码策略就必须使用足够数量的量子比特来容纳这些正交态。或者，若使用大量量子比特进行嵌入，则需要在大型希尔伯特空间中围绕标签态形成簇。在这种情况下，要在这些广阔的空间中寻找数据之间的相似性可能并不容易 [8]。

本文提出了首个多分类 SWAP 测试分类器。我们的工作受到二元 SWAP-Test 分类器以及先前研究中标签态用法的启发。借助这一多分类量子分类器，我们能够避免构建多个二元分类器来执行多类分类所需的成本。无论数据编码策略和类别数量如何，仅使用单量子比特标签态是本工作的创新之处。此外，标签态与数据分开存储，确保标签态的选择独立于数据编码策略。其优点在于，所需量子比特的数量以及所用电路的拓扑结构无需随类别数量变化。重要的是，无论类别数量或数据编码策略如何，都只需重构一个量子比特的态。因此，测量策略也独立于类别数量。与其他多类别量子分类器架构不同，重构单个量子比特的态即可获得多类分类任务所需的全部信息 [9]。

对于某个多类分类问题，分类器通过以下方式实现：首先制备一个量子态，以特定格式编码测试数据、训练数据及其各自的标签。选择标签态，使其对应的布洛赫向量，（我们称之为标签向量）在布洛赫球上彼此相距最远。随后，对制备好的量子态执行改进的 SWAP 测试，其中涉及对存储标签态的量子比特进行态重构。该操作实际上会得到标签向量的线性组合。每个标签向量的贡献，是测试数据与所有具有该标签的训练数据之间核值的加权和。该加权和中的核值与二分类 SWAP 测试分类器一样，可并行计算。通过改变数据编码策略，可以调整所计算的核的类型，以适应分类问题。最后，在经典计算中计算所得向量与每个标签向量之间的重叠度。将测试数据分配给与其重叠度最高的标签向量所对应的标签。与二分类 SWAP 测试分类器一样，这种多分类 SWAP 测试分类器本身并不依赖经典训练子程序。

通过将多分类 SWAP 测试分类器应用于若干不同的分类问题，展示了其有效性；这些问题的特征维度和类别数量各不相同。所使用的数据集包括生成的异或数据集，以及鸢尾花数据集（3 个类别）、葡萄酒数据集（3 个类别）和手写数字数据集（10 个类别）等真实世界数据集 [10]。我们从解析上证明，多分类分类器在这些真实世界数据集上能够达到较高的准确率。即使无需训练，并且仅采用标准的数据编码策略，这些实验也表明，多分类 SWAP 测试分类器具有非凡的强大能力。

随后，我们考察了多分类 SWAP 测试分类器在现实条件下的性能。为此，我们展示了该分类器对有限采样和噪声的鲁棒性。通过方差分析，我们表明，在单个量子比特上能够被准确区分的标签态数量，会随着所需测量的重复次数增加而线性增长。我们还表明，在某些退极化噪声条件下，分类过程不受影响。这些理论结果通过融入退极化噪声和有限采样的数值实验得到了验证。

本文组织结构如下：第 2 节概述了所提出的多类别交换测试分类器的构成步骤，讨论了其对噪声的鲁棒性，并评估了使用该分类器可由单个量子比特存储的标签态数量

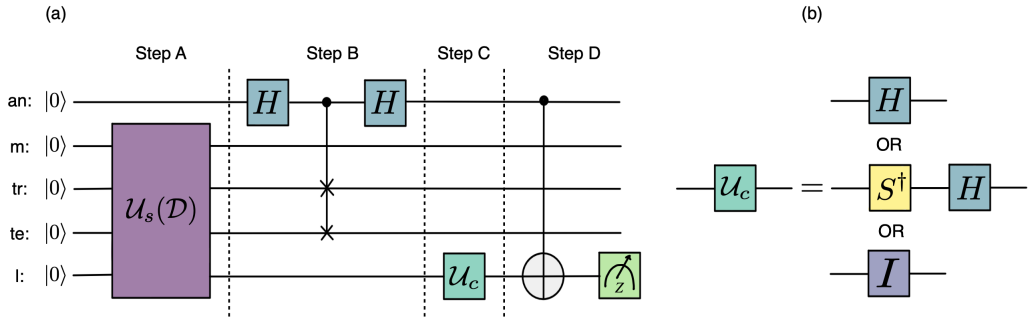


图 1：（a）**多分类 SWAP 测试分类器所需的电路**。第一个寄存器（an）存储辅助量子位。第二个寄存器存储索引寄存器（m），用于将训练寄存器（tr）中的训练数据与标签量子比特（l）上的相应标签态关联起来。测试数据存储在（te）中。为在电路末尾对（l）进行量子态层析，将准备三个执行步骤 A 和 B 的电路。在每个电路中，步骤 A 应用 $\mathcal{U}_s(\mathcal{D})$，将测试数据、训练数据和训练标签制备为方程（2）中给出的量子态 $|\Psi_i\rangle$。随后，步骤 B 交换包含测试数据和训练数据的寄存器。在每个电路中，步骤 C 应用（b）中的一种门序列，将基变换到 X 基或 Y 基，或者保持为 Z 基。这三个电路计算出预测向量$\mathbf{y}_{pred}$，随后将其用于分配函数，以对测试数据进行分类。

。第 3 节介绍了在各种数据集上开展的实验方法与结果。最后，第 4 节总结全文，并指出未来可能的研究方向。

2 结果

2.1 使用多分类 SWAP 测试分类器进行分类

分类是机器学习中的一个基础问题。给定一个数据集

$$\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{M} \subset \mathbb{R}^N \times \{y_i\}_{i=1}^{L}, \tag{1}$$

由训练数据 $\{\mathbf{x}_i\}_{i=1}^{M}$ 及其各自的标签 $\{y_i\}_{i=1}^{M}$组成，监督分类的目标是开发一个用于对未标记数据进行分类的模型。开发这些模型的算法称为分类器。本节介绍构成多分类 SWAP 测试分类器的步骤。图 1 也概述了这些步骤。

对于某个未标记的测试数据 $\tilde{\mathbf{x}}$，多分类 SWAP 测试分类器要求将该测试数据、训练数据及其各自的标签编码为量子态

$$|\Psi_i\rangle = \sum_{m=1}^{M} \sqrt{w_m}\,|0\rangle\,|\tilde{\mathbf{x}}\rangle\,|\mathbf{x}_m\rangle\,|y_m\rangle\,|m\rangle\,. \tag{2}$$

第一个量子比特初始化为基态 $|0\rangle$，之后将作为改进的 SWAP 测试中的辅助量子位。索引寄存器 $\sum_{m=1}^{M}\sqrt{w_m}\,|m\rangle$ 及其 $\sum_{m=1}^{M} w_m = 1$用于将训练数据与各自的标签关联起来。测试数据和训练数据分别编码在不同的寄存器中。某个酉算符 $\mathcal{U}_{\Phi(\mathbf{x})}$ 将测试数据和训练数据分别编码为量子态：$|\tilde{\mathbf{x}}\rangle = \mathcal{U}_{\Phi(\tilde{\mathbf{x}})}|0\rangle$和 $|\mathbf{x}_m\rangle = \mathcal{U}_{\phi(\mathbf{x}_m)}|0\rangle$。这一操作可理解为对经典数据应用特征映射 [1]。算符的选择，即编码方法的选择，决定了分类器将计算的核的类型 [11, 12]。标签量子比特 $|y_m\rangle$ 用于存储每个标签y_m 所映射到的标签态。每个标签态

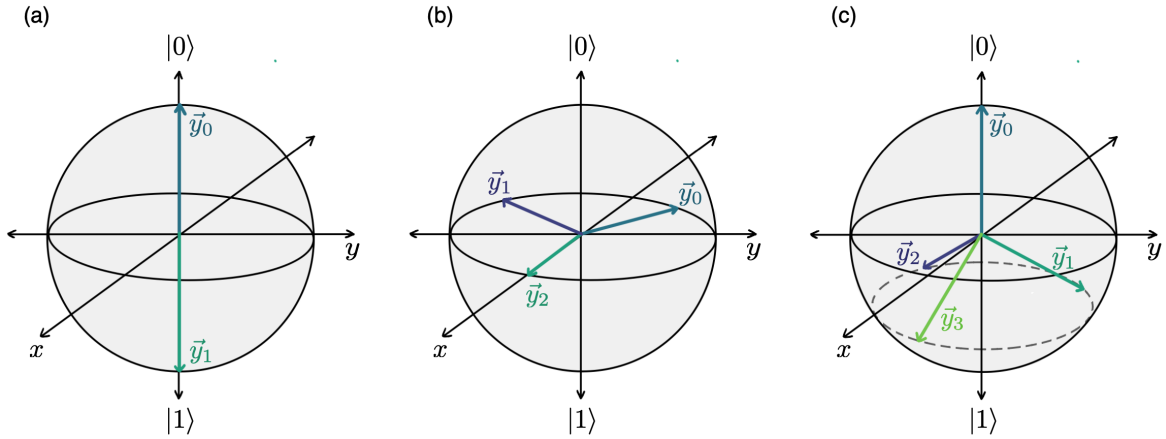


图 2：**标签向量的最优选取**：(a) 2 个类别 $\{y_0$：$[0,0,1]$，y_1：$[0,0,-1]\}$；(b) 3 个类别 $\{y_0$:$[1,0,0]$，y_1：$[-0.5,0.866,0]$，y_2：$[-0.5,-0.866,0]\}$；以及 (c) 4 个类别 $\{y_0$：$[0,0,1]$，y_1：$[-0.471,0.861,-0.333]$，y_2:$[-0.471,-0.861,-0.333]$，y_2：$[0.943,0,-0.333]\}$。这些向量指向塔姆斯问题的解。

$$|y_m\rangle = \cos\left(\frac{\theta_{y_m}}{2}\right)|0\rangle + \mathrm{e}^{i\phi_{y_m}}\sin\left(\frac{\theta_{y_m}}{2}\right)|1\rangle \quad (3)$$

与 $0 \le \theta_{y_m} \le \pi$ 和 $0 \le \phi_{y_m} \le 2\pi$ 相关联，并具有一个布洛赫向量

$$\mathbf{y}_m = \begin{pmatrix} \cos\phi_{y_m}\sin\theta_{y_m} \\ \sin\phi_{y_m}\sin\theta_{y_m} \\ \cos\theta_{y_m} \end{pmatrix} \quad (4)$$

以下将其称为 y_i 的标签向量。标签态的选取方式是，使其布洛赫向量在布洛赫球上彼此尽可能远离。对于此分类器，我们提出，应通过塔姆斯问题的解来确定布洛赫向量的最优位置：该问题旨在将 L 个数据点放置在单位球面上，使距离最近的两个数据点尽可能远 [13]。这样可以使布洛赫球上的向量彼此最大程度地分开。图 2 展示了其中一些位置，包括 $L = 2$，3 类和 4 类的情况。

完成所需的态制备后，对测试数据寄存器和训练数据寄存器执行改进的 SWAP 测试。一个由辅助量子位控制的受控 SWAP 门置于施加在辅助量子位上的两个哈达玛门之间，用于交换这两个寄存器。结果为

$$\begin{aligned} |\Psi_f\rangle &= H_a.\texttt{C-SWAP}.H_a|\Psi_i\rangle \\ &= \sum_{m=1}^{M} \frac{\sqrt{w_m}}{2}(|0\rangle|\psi_+\rangle + |1\rangle|\psi_-\rangle)|y_m\rangle|m\rangle \end{aligned} \quad (5)$$

其中 $|\psi_\pm\rangle = |\tilde{\mathbf{x}}\rangle|\mathbf{x}_m\rangle \pm |\mathbf{x}_m\rangle|\tilde{\mathbf{x}}\rangle$。下标 a 表示哈达玛门 H 作用于辅助量子位。

为了进行必要的量子态层析，需要将上述量子态制备三次。然后，可以对这三个量子态分别执行所需的基变换：

$$\begin{aligned} |\Psi_{fx}\rangle &= H_l|\Psi_f\rangle, \\ |\Psi_{fy}\rangle &= H_l S_l^\dagger|\Psi_f\rangle, \\ |\Psi_{fz}\rangle &= |\Psi_f\rangle. \end{aligned} \quad (6)$$

这里，下标 l 表示将门 H 和 $S^\dagger$作用于标签量子比特。

在对量子态进行任何测量之前，对每个量子态施加一个以辅助量子位为控制、以标签量子比特为目标的受控非门操作：

$$\begin{aligned}\left|\tilde{\Psi}_{fx}\right\rangle &= \texttt{c-not}_{a,l}\left|\Psi_{fx}\right\rangle,\\ \left|\tilde{\Psi}_{fy}\right\rangle &= \texttt{c-not}_{a,l}\left|\Psi_{fy}\right\rangle,\\ \left|\tilde{\Psi}_{fz}\right\rangle &= \texttt{c-not}_{a,l}\left|\Psi_{fz}\right\rangle.\end{aligned} \tag{7}$$

这样，每种情况下原本需要进行的双量子比特测量便转化为单量子比特测量 [14]。最后，对每个˜ ˜量子态 $\rho_{fs} = |\Psi_{fs}\rangle\langle\Psi_{fs}|$，在 $s \in \{x, y, z\}$ 上执行一次单量子比特可观测量$\langle\sigma_z^l\rangle$的测量，其中上标 l 表示该算符仅作用于标签量子比特。利用这些测量结果构造一个向量，我们将其称为预测向量$\mathbf{y}^{pred}$

$$\begin{aligned}\mathbf{y}_{pred} &= \begin{pmatrix}\mathrm{Tr}\left(\sigma_z^{(l)}\rho_{fx}\right)\\ \mathrm{Tr}\left(\sigma_z^{(l)}\rho_{fy}\right)\\ \mathrm{Tr}\left(\sigma_z^{(l)}\rho_{fz}\right)\end{pmatrix},\\ &= \begin{pmatrix}\sum_m \mathrm{w}_m|\langle\tilde{\mathbf{x}}|\mathbf{x}_m\rangle|^2\cos\phi_{y_m}\sin\theta_{y_m}\\ \sum_m \mathrm{w}_m|\langle\tilde{\mathbf{x}}|\mathbf{x}_m\rangle|^2\sin\phi_{y_m}\sin\theta_{y_m}\\ \sum_m \mathrm{w}_m|\langle\tilde{\mathbf{x}}|\mathbf{x}_m\rangle|^2\cos\theta_{y_m}\end{pmatrix}.\end{aligned} \tag{8}$$

起初，$\mathbf{y}^{pred}$ 的重要性可能并不明显。然而，我们可以利用这样一个事实：改进的 SWAP-Test 所得的保真度构成了一个有效的核 $k(\tilde{\mathbf{x}}, \mathbf{x}_m) = |\langle\tilde{\mathbf{x}}|\mathbf{x}_m\rangle|^2$。于是，若 $\alpha_i = \sum_{m|\mathrm{y}_m=i} \mathrm{w}_m k(\tilde{\mathbf{x}}, \mathbf{x}_m)$，则预测向量可表示为

$$\mathbf{y}_{pred} = \sum_{i=1}^{L} \alpha_i \begin{pmatrix}\cos\phi_{y_i}\sin\theta_{y_i}\\ \sin\phi_{y_i}\sin\theta_{y_i}\\ \cos\theta_{y_i}\end{pmatrix} \tag{9}$$

或者等价地，

$$\mathbf{y}_{pred} = \sum_{i=1}^{L} \alpha_i \mathbf{y_i}. \tag{10}$$

现在，显然预测向量是标签向量的线性组合。每个标签向量 α_i 的权重，是测试数据与具有该标签的训练数据之间核值的总和。较高的 α_i 会增大$\mathbf{y}^{pred}$ 与$\mathbf{y}_i$之间的重叠度并表示测试数据与属于类别 y_i的训练数据之间具有较高的相似度。

量子电路估计得到的预测向量随后用于以下赋值函数

$$\tilde{y} = \max_{y_i}\{\mathbf{y_i} \cdot \mathbf{y}_{pred}\}. \tag{11}$$

实际上，当标签向量$\mathbf{y_i}$ 与预测向量$\mathbf{y}^{pred}$ 的内积最大时，测试数据将被分配到类别 y_i。这表明$\mathbf{y_i}$ 与$\mathbf{y}^{pred}$ 的重叠度最高。图3展示了分类器可能得到的若干预测向量示例，以及生成该预测向量的测试数据将如何被分类。

2.2 多分类 SWAP 测试分类器分析

尽管量子硬件的开发已取得长足进展，但在真实设备上执行量子算法时，不可避免地会遇到两种误差来源：噪声和有限采样。本节讨论多分类 SWAP 测试分类器对噪声的鲁棒性。我们还将介绍单个量子比特上能够准确区分的标签态数量，以及噪声如何影响这一数量。

2.2.1 噪声鲁棒性

退极化噪声是量子系统中一种简单但常用的噪声模型 [15]。该模型由量子信道给出

$$\mathcal{E}_d(\rho) = \frac{pI}{2} + (1-p)\rho, \tag{12}$$

它描述了量子系统的信息以某种概率 $0 \leq p \leq 1$丢失的情况。退极化参数 p 对应于退极化噪声的“程度”，其中 $p = 0$ 表示没有退极化噪声，而 $p = 1$ 表示系统信息完全丢失。

在分析退极化噪声对分类器结果的影响时，我们考虑与 [4] 中用于展示二分类 SWAP 测试分类器对退极化噪声具有鲁棒性时相同的条件

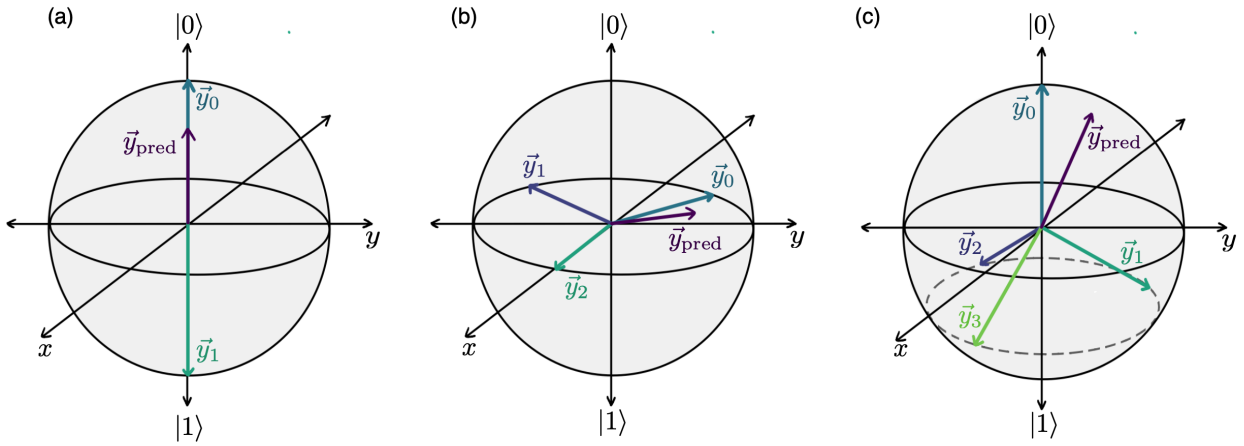


图3：**该多分类 SWAP 测试分类器在 (a) 2 个类别、(b) 3 个类别和 (c) 4 个类别的问题中可能得到的预测向量示意图。**根据公式11中给出的分配函数，当测试点的标签向量$\mathbf{y_i}$ 与所得预测向量的内积最大时，该测试点将被分配到类别 y_i 。根据内积的定义，当向量之间的角度最小时，内积最大。在这些示意图中，测试点将被分配到类别 y_0。注意：这些预测向量并非来自任何具体的分类问题。

。我们考虑公式（12）中的单量子比特退极化通道，但采用 Kraus形式 [16],

$$\mathcal{E}_d(\rho) = \sum_{k=1}^{4} E_k^l \ \rho \ \left(E_k^l\right)^\dagger, \tag{13}$$

其中 Kraus算子的集合为

$$E = \left\{ \sqrt{1 - \frac{3p}{4}} I, \sqrt{\frac{p}{4}}\sigma_x, \sqrt{\frac{p}{4}}\sigma_y, \sqrt{\frac{p}{4}}\sigma_z \right\}, \tag{14}$$

上标 l 表示 Kraus算子E_k 仅作用于标签量子比特。Kraus算子还满足 $\sum_{k=1}^{4} \left(E_k^l\right)^\dagger E_k^l = \mathbb{I}_2$。

接着，我们评估该信道仅作用于测量前的标签量子比特时所产生的影响，也就是说，作用于量子态 ρ_{fx}、ρ_{fy} 和 ρ_{fx}中的标签量子比特。该信道对 ρ_{fx} 的影响为

$$\begin{aligned}
\mathrm{Tr}\left(\sigma_z^{(l)} \mathcal{E}_d(\rho_{fx})\right) &= \mathrm{Tr}\left(\mathcal{E}_d(\rho_{fx})\right)\sigma_z^{(l)}\right) \\
&= \mathrm{Tr}\left(\sum_{k=1}^{4} E_k^l \rho_{fx} \left(E_k^l\right)^\dagger \sigma_z^{(l)}\right) \\
&= \sum_{k=1}^{4} \mathrm{Tr}\left(\rho_{fx} E_k^l \sigma_z^{(l)} \left(E_k^l\right)^\dagger\right) \\
&= (1-p)\mathrm{Tr}\left(\sigma_z^{(l)} \rho_{fx}\right)
\end{aligned} \tag{15}$$

类似地，该信道对 ρ_{fy} 和 ρ_{fz} 的影响为

$$\begin{aligned}
\mathrm{Tr}\left(\sigma_z^{(l)} \ \mathcal{E}_d(\rho_{fy})\right) &= (1-p)\mathrm{Tr}\left(\ \sigma_z^{(l)} \rho_{fy}\right) \\
\mathrm{Tr}\left(\sigma_z^{(l)} \ \mathcal{E}_d(\rho_{fz})\right) &= (1-p)\mathrm{Tr}\left(\ \sigma_z^{(l)} \rho_{fz}\right)
\end{aligned} \tag{16}$$

这些结果可用于构造一个向量

$$(1-p)\begin{pmatrix} \mathrm{Tr}\left(\sigma_z^{(l)} \rho_{fx}\right) \\ \mathrm{Tr}\left(\sigma_z^{(l)} \rho_{fy}\right) \\ \mathrm{Tr}\left(\sigma_z^{(l)} \rho_{fz}\right) \end{pmatrix} = (1-p)\mathbf{y}_{pred}. \tag{17}$$

这表明，在所考虑的退极化噪声条件下，预测向量仅被缩放了 $(1-p)$ 倍。图4展示了缩放预测向量的效果示例。

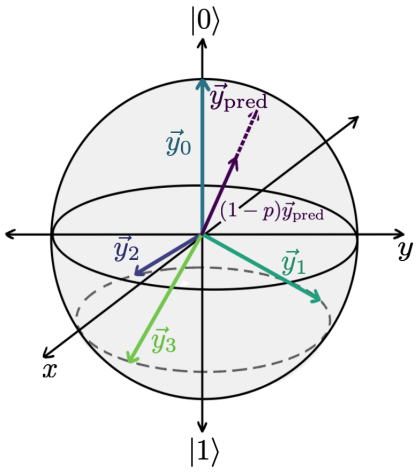


图4：**一个示例，说明所概述的退极化噪声条件会如何影响预测向量**。预测向量将按因子 $(1-p)$ 缩放。这不会影响预测向量与标签向量之间的角度，因此不会影响分类结果。注意：此示例与任何具体的分类问题无关。

If we rewrite the assignment function in equation (11) as

$$\begin{aligned}\tilde{y} &= \max_{y_i}\{|\mathbf{y}_i||\mathbf{y}_{pred}|\cos\lambda_i\} \\ &= \max_{y_i}\{\cos\lambda_i\},\end{aligned} \tag{18}$$

我们可以看到，分配函数仅取决于 λ_i 标签向量$\mathbf{y_i}$ 与预测向量$\mathbf{y}^{pred}$之间的角度。因此显然，将预测向量缩放 $(1-p)$ 倍不会影响分配函数。这样，分类器的输出便不受这些退极化噪声条件的影响。

2.2.2 标签态数量

接下来，我们考虑单个量子比特上可以存储多少个标签态。由于分类器的设计，这一数量取决于我们准确测量预测向量的能力。我们能够在布洛赫球上区分的预测向量数量，就是应该存储在单个标签量子比特上的标签态数量。

为了确定我们能够区分的预测向量数量，我们首先估算所需测量中 Δ 的标准误差。为简洁起见，我们作如下定义： $x = \text{Tr}(\sigma_z^{(l)}\rho_{fx})$、$y = \text{Tr}(\sigma_z^{(l)}\rho_{fy})$ 和 $z = \text{Tr}(\sigma_z^{(l)}\rho_{fz})$。那么， $s \in \{x, y, z\}$ 的标准误差为

$$\Delta s = \sqrt{\frac{4p_s(1-p_s)}{R}} \tag{19}$$

其中 $p_s = \frac{1}{2}\left(\text{Tr}\left(\sigma_z^{(l)}\rho_{fs}\right)+1\right)$，且 R 为重复次数。

使用标准误差传播公式，我们估算 $\Delta\theta$ 和 $\Delta\phi$。这使我们能够估算椭球体的面积，该椭球体表示预测向量测量结果的不确定性。图5展示了这个椭球体。将布洛赫球的表面积除以该椭球体的面积，我们便可求得能够存储的标签态数量（N_s）：

$$\begin{aligned}N_s &= \frac{\text{Area of Bloch Sphere}}{\text{Area of Ellipsoid}} \\ &= \frac{4\pi||\vec{r}||^2}{\pi\Delta\theta\Delta\phi}\end{aligned} \tag{20}$$

其中 $\vec{r} = (x,\ y,\ z)$。

代入必要的数值后，我们得到

$$N_s = \mathcal{O}(R). \tag{21}$$

这意味着，能够存储的标签态数量会随着所需测量的重复次数线性增长。

如果考虑退极化噪声对该数量的影响，那么在这些条件下可存储的标签态数量$\tilde{(N_s)}$为：

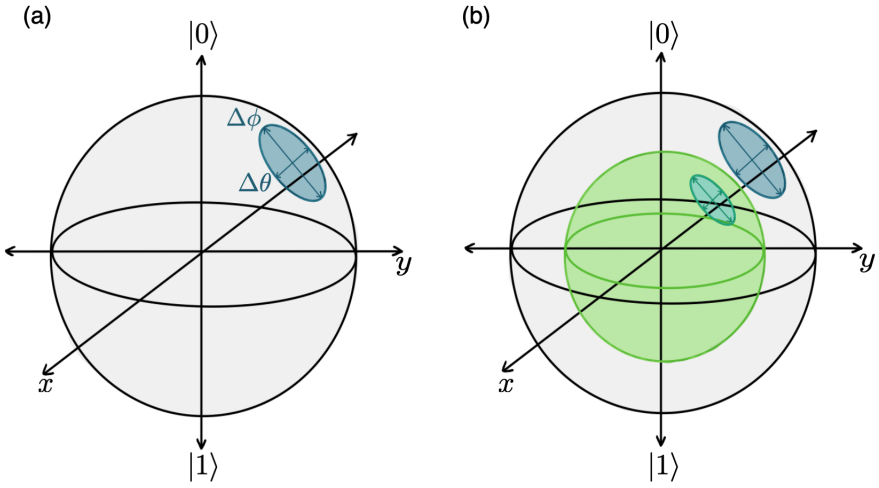


图5：**表示预测向量测量标准误差的椭球**。(a) 展示了布洛赫球上的椭球，(b) 展示了在某些退极化噪声条件下可能出现的缩小布洛赫球上的椭球。在这两种情况下，将球体的面积除以椭球的面积，即可计算出应存储在标签量子比特上的标签态数量。

$$\begin{aligned}\tilde{N}_s &= \frac{4\pi||(1-p)\vec{r}||^2}{\text{Area of Ellipsoid}} \\ &= \frac{4\pi(1-p)^2||\vec{r}||^2}{A(p)},\end{aligned} \tag{22}$$

此时，椭球体的面积成为函数 $A(p)$，现在还取决于退极化参数 p。

在原点（$p=0$）附近对 $A(p)$ 进行泰勒级数展开，保留至二阶，可得：

$$A(p) = A_0 + A_1 p + A_2 p^2 + \mathcal{O}(p^3), \tag{23}$$

where $A_1 = \frac{\partial A(p)}{\partial p}|_{p=0}$ and $A_1 = \frac{1}{2}\frac{\partial^2 A(p)}{\partial p^2}|_{p=0}$.

将此展开式代入方程 (22)，我们得到一个考虑退极化噪声的标签态数量表达式：

$$\begin{aligned}\tilde{N}_s &= \frac{4\pi(1-p)^2||\vec{r}||^2}{A(p)} \\ &\simeq \frac{4\pi(1-p)^2||\vec{r}||^2}{\frac{1}{A_0}(1+\frac{A_1}{A_0}p+\frac{A_2}{A_0}p^2)} \\ &\simeq \frac{4\pi||\vec{r}||^2}{A_0}(1-2p+p^2)[1-\frac{A_1}{A_0}p-(\frac{A_2}{A_0}-\frac{A_1^2}{A_0^2}p^2)] \\ &\simeq N_s\left[1-(\frac{A_1}{A_0}+2)p-(\frac{A_2}{A_0}-\frac{A_1^2}{A_0^2}-2\frac{A_1}{A_0}-1)p^2\right].\end{aligned} \tag{24}$$

通过数值搜索，求得 $(\frac{A_1}{A_0}+2)$ 和 $(\frac{A_2}{A_0}-\frac{A_1^2}{A_0^2}-2\frac{A_1}{A_0}-1)$ 的最大值，其中 $-1 \le x,\ y,\ z \le 1$ 和 $\sqrt{x^2+y^2+z^2} \le 1$ yields

$$(\frac{A_1}{A_0}+2) \le 5 \tag{25}$$

$$(\frac{A_2}{A_0}-\frac{A_1^2}{A_0^2}-2\frac{A_1}{A_0}-1) \le -6. \tag{26}$$

这表明，在最坏情况下：

$$\tilde{N}_s \sim N_s(1 - 5p + 6p^2). \tag{27}$$

这些退极化噪声条件最多会使标签态数量减少 ($1-5p+6p^2$)。我们预计真实实验中的退极化参数为 $p << 0.1$。根据方程 (27)，这对可存储的标签态数量影响微乎其微。

3 方法

我们将多分类 SWAP 测试分类器应用于多个不同的多分类问题，以展示其有效性。

在数值模拟中，我们将该分类器应用于一个包含 4 个类别的三维生成数据集；每个类别包含 16 个数据点。如图 6 所示，该数据集被称为 4-XOR，因为它的设计受到二维 XOR 数据集的启发。与二维 XOR 数据集一样，4-XOR 不是线性可分的。

为了评估多分类 SWAP 测试分类器在该数据集上的性能，我们进行了留一交叉验证。对于每个数据点，我们使用 Qiskit 对构造预测向量的电路进行数值模拟。随后，将预测向量代入方程（11）给出的分配函数，以对该数据点进行分类。首先，我们在理想条件下对电路进行数值模拟，不考虑噪声或有限采样。然后，我们在有限采样条件下，并根据第 2.2.1 节所述的退极化噪声条件，对电路进行数值模拟。图 7 展示了在退极化噪声下进行模拟的电路，图 8 则对退极化噪声的模拟进行了更详细的说明[17]。

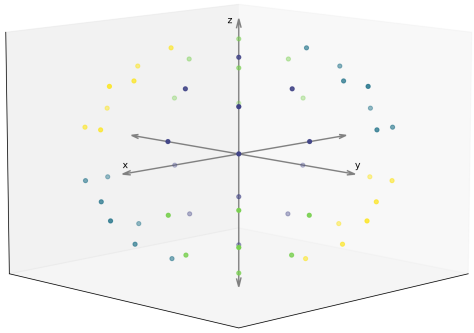


图 6： **4-XOR数据集（3 个特征、4 个类别和 64 个数据点）**。生成该数据集时，同一类别的数据点被安排在球面上彼此正对的位置，以使式 (29) 中的核发挥最大效用。

对测试数据或训练数据进行编码时，我们首先将数据归一化，使得 $\|\mathbf{x}\|^2 = \sum_i |x_i|^2 = 1$。然后，将归一化数据的每个分量编码到相应寄存器中不同的计算基态：

$$\mathbf{x} \to \sum_{i=1}^{2^n} x_i \left|i\right\rangle . \tag{28}$$

这种策略通常称为振幅编码，它会产生以下量子核：

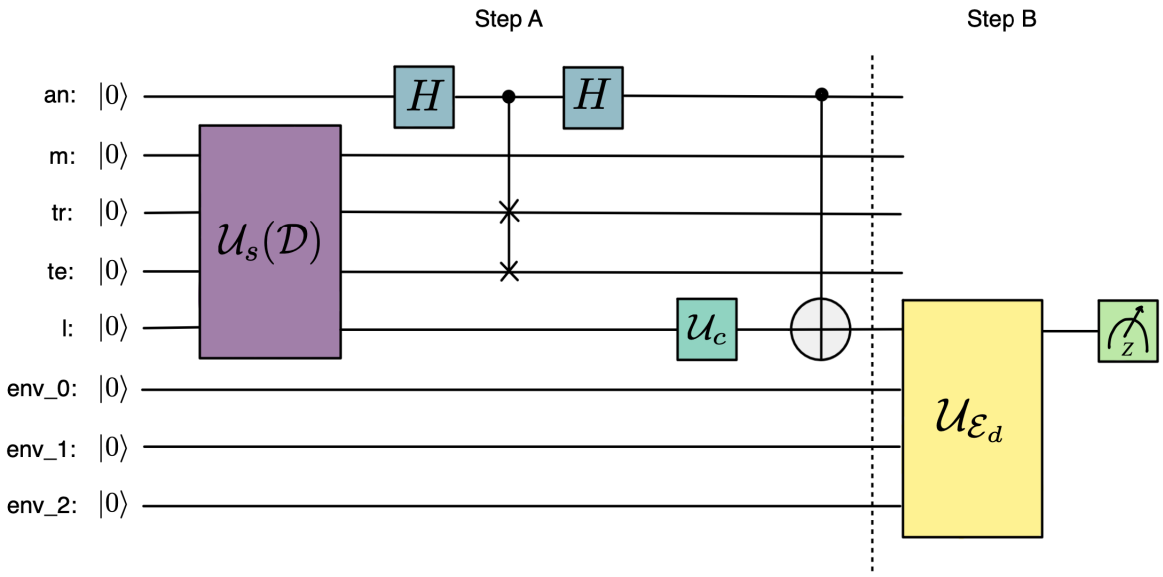


图7：**在退极化噪声下模拟多分类 SWAP 测试分类器所需的电路**。步骤 A 包括态制备、改进的 SWAP 测试和基变换。步骤 B 包括应用 $\mathcal{U}_{\mathcal{E}_d}$ 来模拟退极化噪声。

$$k(\mathbf{x}, \mathbf{z}) = |\langle \mathbf{x}|\mathbf{z}\rangle|^2, \tag{29}$$

它就是线性核 [1] 的模平方。还应注意，训练权重采用均匀分布，即：$\sum_{m=1}^{M} w_m = 1, \quad w_m = 1/M \; \forall \; M$。

我们的模拟表明，在理想条件以及存在退极化噪声和有限采样的现实条件下执行时，分类器均能达到较高的准确率（100%）。这些实验的结果如表 1 所示。分析在退极化噪声下模拟得到的预测向量后，证实了第 2.2.2 节的结论。我们发现，在退极化噪声条件下，得到的预测向量仅发生缩放，这不会影响分类过程。如图 9所示。表 1 也展示了这一点：预测向量的平均范数减少了 $(1-p)$ 倍。

我们进一步评估多类分类器在其他几个数据集上的表现，并采用五折交叉验证。这些数据集包括三个生成的异或数据集，以及 scikit-learn 提供的鸢尾花数据集、葡萄酒数据集和手写数字数据集。需要注意的是，葡萄酒数据集和手写数字数据集经过了人工平衡。对于葡萄酒数据集，我们从每个数据集的每个类别中均匀抽取 48 个数据点。对于手写数字数据集，则从每个类别中均匀抽取 174 个数据点。

对于这些数据集中的每个测试点，我们使用方程 (10) 以经典方式计算预测向量。为此，我们直接计算每个 $\alpha_i = \sum_{m|y_m=i} w_m |\langle \tilde{\mathbf{x}}|\mathbf{x}_m\rangle|^2$，即 $\alpha_i = \sum_{m|y_m=i} w_m k(\tilde{\mathbf{x}}, \mathbf{x}_m)$。这之所以可行，仅仅是因为我们用于将测试数据和训练数据编码为量子态的方法，会产生可通过经典方式计算的核 $k(\tilde{\mathbf{x}}, \mathbf{x}_m)$。在某些情况下，我们使用核 $k(\mathbf{x}, \mathbf{z}) = |\langle \mathbf{x}|\mathbf{z}\rangle|^2$，它源自振幅编码。在其他情况下，我们使用核 $k(\mathbf{x}, \mathbf{z}) = \prod_{k=1}^{n} |\cos(\mathrm{x_k} - \mathrm{z_k})|^2$，它源自角度编码。构造出预测向量后，接着将其用于分配

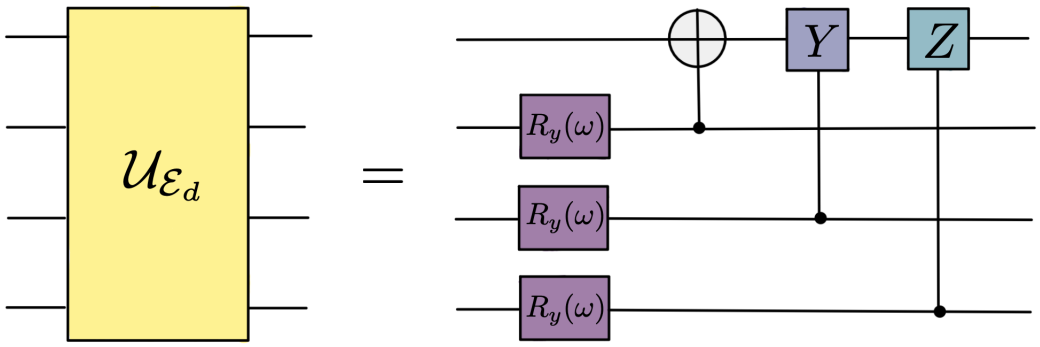


图 8：**算符 $\mathcal{U}_{\mathcal{E}_d}$ 的详细示意图，该算符用于模拟电路中的退极化噪声**。此处，R_y 是一个 Y 轴旋转，其旋转角为 $\omega = \frac{1}{2}\arccos(1-2p)$。

表 1：多分类 SWAP 测试分类器应用于生成的 4-XOR 数据集时所获得的准确率。所列准确率来自包含有限采样和退极化噪声的数值模拟。预测向量的平均范数来自仅包含退极化噪声的数值模拟，用于展示第 2.2.1 节的结果

退极化率（p）	准确率（%）	预测向量的平均范数
0	100	0.1356
0.02	100	0.1331
0.04	100	0.1302
0.06	100	0.1275
0.08	100	0.1248
0.1	100	0.1223

函数，即方程 (11) 中给出的函数，以对测试点进行分类。这些实验的结果见表 2。可以看出，准确率很高，每个数据集的准确率都超过了 90%。

4 结论

我们提出了一种多分类量子分类器，由一组量子电路和一个以经典方式计算的分配函数组成。量子电路估计测试数据与训练数据之间核值的加权和，而分配函数则使我们能够有意义地解释这个总和。通过分析和实验，已证明多分类 SWAP 测试分类器功能强大且对噪声具有鲁棒性。研究还表明，该分类器能够处理 $\mathcal{O}(R)$ 个类别，其中 R 是所需电路的执行次数。在我们的数值实验中，我们采用了标准编码策略，并证明其对生成的数据集有效。值得注意的是，未来可以选择能够产生经典上难以模拟的核函数的编码策略 [2] 。与二元 SWAP-Test 分类器一样，该分类器也可以将评估得到的核函数提升至 n 次方，代价是需要 n 份

表 2：多分类 SWAP 测试分类器应用于各个数据集时获得的准确率。

数据集	类别数	特征数	数据点数	编码	准确率（%）
XOR	2	2	100	振幅	100
	4	3	200	振幅	99
	8	4	400	振幅	99
Iris	3	4	150	角度	95
Wine	3	13	144	角度	92
手写数字数据集	10	64	1740	角度	92

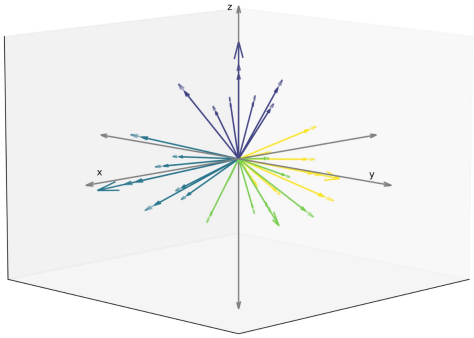


图9：**通过对4-XOR 数据集**进行分类得到的预测向量。这里，较深色的向量是加入退极化噪声（$p = 0.1$）后评估得到的预测向量，较浅色的向量则是在未加入退极化噪声时评估得到的预测向量。可以看到，预测向量只是缩短了，但向量之间的角度并未改变。

测试数据和训练数据。我们的数值实验也采用了均匀训练权重，但未来的工作可以开发一种使用可训练或非均匀训练权重的策略。未来另一个值得研究的方向是使用真实量子设备，利用该分类器解决多分类问题。

致谢

本研究基于南非共和国国家研究基金会资助的研究。谨此感谢 NICIS（国家综合网络基础设施系统）电子科研资助项目QICSA 的支持。衷心感谢 CSIR DSI 跨机构奖学金支持（IBS）计划的资助。感谢人工智能研究中心提供的支持。感谢 IBM Quantum 提供模拟器供我们使用。本文所表达的观点仅代表作者个人，不反映 IBM 或 IBM Quantum 团队的官方政策或立场。我们感谢 A. W. Pillay 先生和 I. J. David 先生协助校对稿件。

参考文献

[1] Maria Schuld 和 Nathan Killoran. 特征希尔伯特空间中的 Quantum 机器学习。*PhysicalReview Letters*, 122(4):040504，2019年。

[2] Vojtˇech Havlíˇcek, Antonio D Córcoles, Kristan Temme, Aram W Harrow, Abhinav Kandala, Jerry M Chow 和 Jay M Gambetta。利用量子增强特征空间进行监督学习。*Nature*, 567(7747):209–212, 2019年。

[3]Carsten Blank, Daniel K Park, June-Koo Kevin Rhee 和 Francesco Petruccione。采用定制量子核的量子分类器。*npjQuantumInformation*, 6(1):1–7, 2020。

[4]Daniel K Park, Carsten Blank 和 Francesco Petruccione。基于量子核的二元分类器理论。*PhysicsLetters A*, 384(21):126422, 2020。

[5] Ethem Alpaydin。机器学习导论。MIT出版社，2020。

[6] Nhat A Nghiem、Samuel Yen-Chi Chen 和 Tzu- Chieh Wei。量子分类的统一框架。*Physical ReviewResearch*，3(3):033056，2021。[7] Adrián Pérez-Salinas、Alba Cervera-Lierta、Elies Gil-Fuster 和 José I Latorre。用于通用量子分类器的数据重上传。*Quantum*，4:226，2020。[8] Hsin-Yuan Huang、Michael Broughton、Masoud Mohseni、Ryan Babbush、Sergio Boixo、Hartmut Neven 和 Jarrod R McClean。数据在量子机器学习中的力量。*Nature communica-tions*，12(1):1–9，2021。[9]Yi Zeng、Hao Wang、Jin He、Qijun Huang 和 Sheng Chang。使用全量子比特多可观测量测量策略的多分类混合量子神经网络。*Entropy*，24(3):394，2022。[10] Dheeru Dua 和 Casey Graff。UCI 机器学习知识库，2017。[11]Maria Schuld 和 Francesco Petruccione。使用量子计算机进行监督学习，第17卷。Springer，2018。[12] Maria Schuld。监督式量子机器学习模型是核方法。*arXiv* 预印本 *arXiv:2101.11020*，2021。[13] Pieter Merkus Lambertus Tammes。关于花粉粒表面出口位置的数量和排列的起源。*Recueil des travauxbotaniquesnéerlandais*，27(1):1–84，1930。[14]Daniel K Park、Carsten Blank 和 Francesco Petruccione。开销最小的鲁棒量子分类器。载于*2021 International JointCon-ference on Neural Networks (IJCNN)*，第1–7页。IEEE，2021。[15] Michael A Nielsen 和 Isaac Chuang。量子计算与量子信息。美国物理教师协会，2002。[16] Karl Kraus。量子理论中的一般状态变化。*AnnalsofPhysics*，64(2):311–335，1971。[17]Guillermo García-Pérez、Matteo AC Rossi 和 Sabrina Maniscalco。IBM Q Experience 是模拟开放量子系统的多功能实验测试平台。*npj QuantumInformation*，6(1):1–10，2020。