基于量子 K-means 的平台聚类编组 量子增强求解方法

何一1)2) 郑寇全1)2) 荆锋1)2) 张毅军1)2) 王勋1) 刘颖1)2) 赵乐1)

1) (国防科技大学试验训练基地, 西安 710106)

2) (陕西省智能协同网络军民共建重点实验室, 西安 710106)

(2024年9月9日收到; 2024年10月27日收到修改稿)

针对联合作战战役行动中平台聚类编组问题,本文提出了一种基于量子 K-means 的量子增强求解方法. 该方法首先分别对经典 K-means 算法中的聚类类别数目设定和聚类中心点选择两部分进行了优化处理;其次,该方法针对聚类数据样本与各聚类中心点之间的欧氏距离构建对应的量子线路;然后,该方法针对聚类 数据集的误差平方和构建对应的量子线路.实验结果表明,所提方法不但有效解决了此类行动规模下的平台 聚类编组问题,与经典 K-means 算法相比,算法的时间复杂度和空间复杂度都有较大幅度降低.

关键词:量子 K-means 算法,量子增强,平台聚类 PACS: 03.67.Lx, 03.65.Aa CSTR: 32037.14.aps.73.20241265

DOI: 10.7498/aps.73.20241265

1 引 言

平台聚类编组问题作为军事 C2 组织资源调度中重点研究问题之一,一直受到国内外研究人员的广泛关注^[1,2].在联合作战中,由于涉及的平台数量庞大、类型多样,资源供给数量众多、种类丰富,在进行平台聚类编组时不但需要考虑大量的武器平台元素,还需要考虑平台位置信息、提供的资源规模等属性,因此联合作战中平台聚类编组问题属于组合优化问题中的 NP (non-deterministic polynomial)问题^[3,4].由于军事 C2 组织资源调度中一般要求任务簇与平台簇一一对应,所以通常情况下联合作战中任务分组的数量就决定了平台编组的数量,据此可以利用无监督聚类优化算法求解联合作战中平台聚类编组问题.

目前, 无监督聚类优化算法主要有 K-means 算法^[5-7]、层次聚类^[8-10]、Mean-shift 聚类算法^[11-13]、 高斯混合模型 EM 算法^[14-16]等.与其他方法相比, K-means 算法因具有算法思路简单、聚类效果优良 等特点,加之在处理大数据情况下,具有很好的伸 缩性,算法复杂度相对较低等优点,使其在实际应 用中广受青睐^[17,18].该算法作为数据挖掘领域的十 大算法之一,已成功应用于特征分析^[19-21]、商业智 能^[22-24]、图像分组^[25-27]、文档聚类^[28-30]等多个领 域.随着大数据时代的到来,问题所包含的聚类数 据量呈指数级增长,给 K-means 算法的计算速 率带来了巨大挑战.一些学者尝试将量子计算与 K-means 算法相结合,利用量子天然所具有的并行 计算能力来加速 K-means 算法的聚类过程,达到 量子增强的目的,使得 K-means 算法的时间复杂 度有较大幅度地降低^[31-33].

在上述研究成果的启发下,本文针对联合作战 战役行动中平台聚类编组问题提出了一种基于量 子 K-means 的量子增强求解方法.首先,该方法 在 K-means 算法的基础上,利用已确认的任务簇

[†] 通信作者. E-mail: zhangyijun_gfkjdx@163.com

^{© 2024} 中国物理学会 Chinese Physical Society

数量确定平台聚类的类别数量,再以每个任务簇中 所含任务的位置信息为基准,通过推导计算,求解 出每个初始聚类中心点位置;其次,以欧氏距离作 为衡量平台数据与各聚类中心点间相似度的指标, 对平台数据进行量子化处理,将平台数据转化成对 应的量子态形式表示,根据理论推导将欧氏距离求 解转化成量子态内积求解,通过设计构造通用的量 子态内积求解量子线路完成对欧氏距离的求解;接 着,根据平台数据与各聚类中心之间的相似度,将 各平台划归到对应的聚类中心并对聚类中心位置 信息进行更新,不断迭代直至达到类簇的误差平方 和数值或者预先设定的迭代阈值,完成聚类并取得 平台聚类编组结果.

2 基础理论

经典 K-means 算法的基本思想为:首先,根据 相关依据和经验,预先确定聚类类别数目 K,并从 聚类数据集中随机选取 K个初始聚类中心点;其 次,依次计算其余聚类数据元素与各聚类中心点之 间的欧氏距离,通过欧氏距离找出距离目标数据元 素最近的聚类中心点,并将该聚类数据元素划归到 该聚类中心点所对应的簇中;接着,通过求解每个 簇中的聚类数据元素的平均值,计算该簇对应的新 聚类中心点;最后,根据聚类效果,进行反复迭代, 直至聚类中点位置相对稳定或者迭代次数达到设 定次数.

2.1 经典 K-means 算法具体步骤和流程

设聚类数据集合为 $X = \{x_1, x_2, \dots, x_N\}$,其 中N为聚类数据数量,初始聚类中心点有K个, 每个聚类中心点表示为 c_k ,其中 $k \in \{1, 2, \dots, K\}$, 每个聚类数据维度为M.则可以定义如下公式:

$$D(x_i, c_k) = \sqrt{\sum_{j=1}^{M} (x_{ij} - c_{kj})^2},$$
 (1)

其中 $D(x_i, c_k)$ 表示聚类数据 x_i 与聚类中心点 c_k 之间的欧氏距离, x_{ij} 和 c_{kj} 分别表示聚类数据 x_i 和聚类中心点 c_k 第j个维度对应的属性值, $j \in$

 $\{1, 2, \cdots, M\}$.

$$c_k' = \frac{\sum_{x_i \in C_k} x_i}{N_k},\tag{2}$$

其中*c_k* 表示通过求平均值得到的新聚类中心点, *C_k* 表示聚类数据划归到聚类中心点*c_k* 所组成的集 合, *N_k* 表示聚类集合 *C_k* 所含聚类数据的数量.

SSE =
$$\sum_{k=1}^{K} \sum_{x_i \in C_k} |D(x_i, c_k)|^2$$
, (3)

其中 SSE 表示整个聚类数据集的残差平方和.

经典 K-means 算法的具体步骤如下:

步骤1 根据具体问题,结合实际经验具体分析,确定将要划分聚类中心数目 *K*;

步骤 2 在聚类数据集中,初始聚类中心点通 过随机选定的 *K* 个数据元素来确定;

步骤3 根据(1)式计算其余聚类样本元素和 各聚类中心点之间的欧氏距离;

步骤 4 对欧氏距离计算结果进行比较,将聚 类数据元素划归到与其欧氏距离最小的聚类中心 点所在的簇中;

步骤 5 确定每个簇中的聚类数据元素数目, 并通过 (2) 式对簇中聚类数据元素的均值进行计算;

步骤 6 将均值计算结果作为新的聚类中心 点位置,并进行更新;

步骤 7 根据 (3) 式, 计算整个聚类数据集的 残差平方和;

步骤 8 对聚类数据集的残差平方和变化进行判断,若变化超出所设定的范围,则跳转到步骤 3,根据新的聚类中心点重新进行聚类分簇.如果变化在所设定的范围内,则结束聚类,将聚类分簇结果进行输出.图1为K-means算法的具体聚类流程图.

2.2 聚类数据与聚类中心点之间欧氏距离 的具体推导

设 $x_i = [x_{i1}, x_{i2}, \dots, x_{iM}]^T$ 为聚类数据集合 *X* 中的任意数据元素, $c_k = [c_{k1}, c_{k2}, \dots c_{kM}]^T$ 为聚 类中心集合 C_k 中的任意一个聚类中心点. 根据 (1) 式, $x_i = c_k$ 之间的欧氏距离具体推导如下:

$$D(\mathbf{x}_{i}, \mathbf{c}_{k}) = \sqrt{\sum_{j=1}^{M} (x_{ij} - c_{kj})^{2}} = \sqrt{(x_{i1} - c_{k1})^{2} + (x_{i2} - c_{k2})^{2} + \dots + (x_{iM} - c_{kM})^{2}}$$
$$= \sqrt{(x_{i1}^{2} + x_{i2}^{2} + \dots + x_{iM}^{2}) + (c_{k1}^{2} + c_{k2}^{2} + \dots + c_{kM}^{2}) - 2(x_{i1}c_{k1} + x_{i2}c_{k2} + \dots + x_{iM}c_{kM})}.$$
(4)



图 1 K-means 算法的具体聚类流程图

Fig. 1. Specific clustering process diagram of the K-means algorithm.

根据 (4) 式, 整个聚类数据集的残差平方和 SSE 公 式也可以进一步推导:

$$SSE = \sum_{k=1}^{K} \sum_{x_i \in C_k} |D(x_i, c_k)|^2$$
$$= \sum_{k=1}^{K} \sum_{x_i \in C_k} \left[\left(x_{i1}^2 + x_{i2}^2 + \dots + x_{iM}^2 \right) + \left(c_{k1}^2 + c_{k2}^2 + \dots + c_{kM}^2 \right) - 2 \left(x_{i1}c_{k1} + x_{i2}c_{k2} + \dots + x_{iM}c_{kM} \right) \right].$$
(5)

通过对经典 K-means 算法的计算复杂度分析, 可以发现该算法对计算资源的消耗主要集中在计 算聚类数据元素与聚类中心点之间的欧氏距离部 分(步骤 3)和计算整个聚类数据集的残差平方和 部分(步骤 7).随着数据量和维度的不断增大,该 算法所消耗的计算资源会急剧增加,算法的计算复 杂度也会指数级提升.根据(4)式和(5)式的推导 结果,可以建立起欧氏距离与内积计算之间的联 系,将计算欧氏距离转化为内积计算.但是,针对 内积计算问题采用目前常用的经典算法进行求解 需要消耗较大的计算资源,并且算法的计算复杂度 都比较高.所以,本文从新兴的量子技术角度出发, 考虑在量子理论体系下,利用量子增强技术来提升 经典 K-means 算法处理问题的能力.

3 基于量子 K-means 的量子增强求解 方法

本文提出的基于量子 K-means 的量子增强求 解方法是将量子力学基础理论和量子计算相关理 论与经典 K-means 算法的主要步骤相结合,在充 分发挥量子并行计算优势基础上提升经典算法的 执行效率和准确度.针对经典 K-means 算法,本文 提出的方法主要做了 3 个方面的优化和改进.一是 根据具体问题对聚类分簇 K 值和初始聚类中心点 位置进行优化处理;二是对所有聚类数据元素和聚 类中心点进行量子化处理,建立经典内积与量子态 内积之间的联系;三是制备量子态,设计构建通用 的量子态内积求解量子线路,通过对量子态内积的 求解实现对应欧氏距离的求解.

3.1 聚类分簇和聚类中心点优化

经典 K-means 算法较为明显的缺点就是聚类 分簇 K值难以确定和初始聚类中心点随机选取. 如果聚类分簇 K值和初始聚类中心点选取不当往 往会严重影响到最终聚类结果,使其陷入局部最优.

对于聚类分簇 K值, 在算法聚类过程中, 如 果 K值选择过大, 会导致各簇之间数据特征的差 异性很小, 如果 K值选择过小, 会导致每个簇内部 的数据特征差异性很大.同时, K值过大或者过小 都会导致聚类结果不理想, 难以实现全局最优.对 平台进行聚类的目标就是想实现任务簇与平台簇 的一一对应关系, 因此本文从实际出发以已知的任 务簇数量作为初始聚类分簇 K的取值 (如图 2 所 示).这样选取的 K值不但与实际相符, 也更容易 求出全局较优解.



图 2 初始化平台集聚类分簇 K 值图

Fig. 2. K value graph of initialization platform clustering class clustering.

在 K-means 算法聚类过程中, 初始聚类中心 点的选取对聚类效果起着关键性作用, 当初始聚类 中心点位置不理想时, 不但影响聚类效果, 还会使 算法的迭代次数增加, 从而提高算法的计算复杂 度. 本文在确定聚类分簇 K值的基础上, 对每个任 务簇中的任务数据元素平均值进行计算, 并将该平 均值对应的位置作为对应平台簇的初始聚类中心 点位置. 这样选取的聚类中心点不仅与实际问题相 结合, 具体操作还比较简单, 一定程度上降低了算 法的操作复杂度.

3.2 经典内积与量子态内积

在希尔伯特空间中,一个量子位对应的量子态可以表示成一个二维向量,根据张量积运算,两个量子位对应的量子态就可以表示一个四维向量,以此类推,n个量子位对应的量子态就可以表示一个2ⁿ 维向量.当聚类数据元素的特征维度为 N'时,需要的量子位数为 [log₂N'].当量子态对应的向量形式中出现多余的维度时,将其对应的振幅数值设置为 0 即可.下面进行具体的计算推导.

设聚类数据元素与聚类中心点的特征维度都 为 N',设一个聚类数据元素 $x_1 = [x_{11}, x_{12}, \cdots, x_{1N'}]^T$ 与一个聚类中心点 $c_1 = [c_{11}, c_{12}, \cdots, c_{1N'}]^T$. 设两个量子态分别为 $|\eta\rangle = [e_1, e_2, \cdots, e_L]^T$, $|\vartheta\rangle = [f_1, f_2, \cdots, f_L]^T$,其中 $L = 2^{\lceil \log_2 N' \rceil}$.令 $E = ||x_1|| = \sqrt{\sum_{i=1}^{N'} x_{1i}^2}$, $F = ||c_1|| = \sqrt{\sum_{i=1}^{N'} c_{1i}^2}$.将 E代入 x_1 的表达式, x_1 的表达形式可以转化为

$$\boldsymbol{x}_1 = E\left[\frac{x_{11}}{E}, \frac{x_{12}}{E}, \cdots, \frac{x_{1N'}}{E}\right]^{\mathrm{T}}.$$
 (6)

将 *F* 代入 *c*₁ 的表达式, *c*₁ 的表达形式可以转 化为:

$$\boldsymbol{c}_1 = F\left[\frac{c_{11}}{F}, \frac{c_{12}}{F}, \cdots, \frac{c_{1N'}}{F}\right]^{\mathrm{T}}.$$
 (7)

当L = N'时, 令 $e_1 = \frac{x_{11}}{E}, e_2 = \frac{x_{12}}{E}, \cdots, e_L = \frac{x_{1N'}}{E}$, 令 $f_1 = \frac{c_{11}}{F}, f_2 = \frac{c_{12}}{F}, \cdots, f_L = \frac{c_{1N'}}{F}$.可以 得到如下表达式:

$$|\eta\rangle = [e_1, e_2, \cdots, e_L]^{\mathrm{T}}$$
$$= \left[\frac{x_{11}}{E}, \frac{x_{12}}{E}, \cdots, \frac{x_{1N'}}{E}\right]^{\mathrm{T}} = \frac{x_1}{E}, \qquad (8)$$
$$|\vartheta\rangle = [f_1, f_2, \cdots, f_L]^{\mathrm{T}}$$

$$= \left[\frac{c_{11}}{F}, \frac{c_{12}}{F}, \cdots, \frac{c_{1N'}}{F}\right]^{\mathrm{T}} = \frac{c_1}{F}.$$
 (9)

根据 (8) 式和 (9) 式, 计算量子态 |η⟩ 与 |ϑ⟩ 的 内积可以得到如下表达式:

$$\langle \eta \mid \vartheta \rangle = [e_1^*, e_2^*, \cdots, e_L^*] [f_1, f_2, \cdots, f_L]^{\mathrm{T}}$$

$$= \left\langle \frac{\boldsymbol{x}_1}{E}, \frac{\boldsymbol{c}_1}{F} \right\rangle = \frac{1}{EF} \left\langle \boldsymbol{x}_1, \boldsymbol{c}_1 \right\rangle.$$
(10)

当L > N'时, 令 $e_1 = \frac{x_{11}}{E}$, $e_2 = \frac{x_{12}}{E}$, \cdots , $e_{N'} = \frac{x_{1N'}}{E}$, $e_{N'+1} = 0$, $e_{N'+2} = 0$, \cdots , $e_L = 0$. 令 $f_1 = \frac{c_{11}}{F}$, $f_2 = \frac{c_{12}}{F}$, \cdots , $f_{N'} = \frac{c_{1N'}}{F}$, $f_{N'+1} = 0$, $f_{N'+2} = 0$, \cdots , $f_L = 0$. 可以得到如下表达式:

$$|\eta\rangle = [e_1, e_2, \cdots, e_L]^{\mathrm{T}}$$

$$= \left[\frac{x_{11}}{E}, \frac{x_{12}}{E}, \cdots, \frac{x_{1N'}}{E}, 0, \cdots, 0\right]^{\mathrm{T}}, \qquad (11)$$

$$|\vartheta\rangle = \left[f_1, f_2, \cdots, f_L\right]^{\mathrm{T}}$$

$$= \left[\frac{c_{11}}{F}, \frac{c_{12}}{F}, \cdots, \frac{c_{1N'}}{F}, 0, \cdots, 0\right]^{\mathrm{T}}.$$
 (12)

根据 (11) 式和 (12) 式, 计算量子态 $|\eta\rangle = |\vartheta\rangle$ 的内积可以表示为

$$\langle \eta \mid \vartheta \rangle = [e_1^*, e_2^*, \cdots, e_L^*, 0, \cdots, 0]$$
$$\times [f_1, f_2, \cdots, f_L, 0, \cdots, 0]^{\mathsf{T}} = \frac{1}{EF} \langle \boldsymbol{x}_1, \boldsymbol{c}_1 \rangle.$$
(13)

这样无论 L 大于或者等于 N',都可以建立起 经典内积与量子态内积之间的联系,使接下来构建 量子线路进行内积求解成为可能.

3.3 通用的量子态内积量子线路

设任意两个量子态分别为 $|\varphi\rangle = [q_1, q_2, \cdots, q_L]^T$, $|\phi\rangle = [p_1, p_2, \cdots, p_L]^T$, 其中 *L* 表示量子比特 位数, $\sum_{i=1}^{L} |q_i|^2 = 1$, $\sum_{i=1}^{L} |p_i|^2 = 1$. 以生成量 子态 $|\varphi\rangle$ 为例, 设计其对应的通用量子线路. 根据 量子基础理论, 量子态 $|\varphi\rangle$ 可以写成如下形式:

根据 (14) 式中 $|\varphi\rangle$ 的表达形式,参考文献 [34] 量子态制备过程中量子逻辑门选择方法,经过量子 酉变换计算推理,利用量子逻辑门设计出了制备 $|\varphi\rangle$ 对应的量子线路.如图 3 所示,在量子初始基态 $|0\rangle^{\otimes L}$ 的基础上,经过 RY(*i*)量子逻辑门、X量子 逻辑门等操作后,得到 $|\varphi\rangle$ 对应的量子态.具体计 算推导过程如下:



图 3 制备量子态 S11 量子线路概率图

Fig. 3. Probability diagram for preparing quantum circuits of quantum state $\,S_{11}$.

$$\begin{split} |0\rangle^{\otimes(L-1)} |0\rangle &\xrightarrow{\mathrm{RY}(\theta_{11})} |0\rangle^{\otimes(L-1)} \left(a_{11} |0\rangle + b_{11} |1\rangle\right) \xrightarrow{\mathrm{Controlled} - S_{12}} \\ a_{11} |0\rangle^{\otimes(L-1)} |0\rangle + b_{11} |\varphi_{12}\rangle |1\rangle \xrightarrow{X} a_{11} |0\rangle^{\otimes(L-1)} |1\rangle + b_{11} |\varphi_{12}\rangle |0\rangle \\ \xrightarrow{\mathrm{Controlled} - S_{11}} a_{11} |\varphi_{11}\rangle |1\rangle + b_{11} |\varphi_{12}\rangle |0\rangle \xrightarrow{X} a_{11} |\varphi_{11}\rangle |0\rangle + b_{11} |\varphi_{12}\rangle |1\rangle \end{split}$$

RY(i) 量子逻辑门的矩阵形式为

$$\operatorname{RY}(\theta) = e^{-i\theta Y/2} = \begin{bmatrix} \cos(\theta/2) & -\sin(\theta/2) \\ \sin(\theta/2) & \cos(\theta/2) \end{bmatrix}. (15)$$

X量子逻辑门的矩阵形式为

$$\mathbf{X} = \begin{bmatrix} 0 & 1\\ 1 & 0 \end{bmatrix}. \tag{16}$$

从上述计算推导过程和图 3显示的信息可以 看出,图中量子逻辑门模块 S₁₁和 S₁₂的主要作用 就是演化生成量子态 |φ₁₁〉和 |φ₁₂〉.在设计量子线 路过程中,直接实现量子逻辑门模块 S_{11} 和 S_{12} 比较困难,本文可以采用相同的原理,将量子态 $|\varphi_{11}\rangle$ 和 $|\varphi_{12}\rangle$ 的表达式写成 $|\varphi_{11}\rangle = a_{21} |\varphi_{21}\rangle |0\rangle + b_{21} |\varphi_{22}\rangle |1\rangle$ 和 $|\varphi_{12}\rangle = a_{22} |\varphi_{23}\rangle |0\rangle + b_{22} |\varphi_{24}\rangle |1\rangle$ 的 形式,其中 a_{21} , b_{21} , a_{22} , b_{22} 为实数,且 $a_{21}^2 + b_{21}^2 = 1$, $a_{22}^2 + b_{22}^2 = 1$.这样就可以采用递归方法,对量 子线路中较为复杂的量子逻辑门模块 S_{11} 和 S_{12} 进 行进一步分解设计,最终得到如图 4 所示的制备量 子态 $|\varphi\rangle$ 的通用量子线路图.该量子线路图可以适 用于任意量子态的制备.



图 4 制备量子态 $|0\rangle_1 |\varphi\rangle_2 |\phi\rangle_3 \xrightarrow{1:H} \frac{1}{\sqrt{2}} (|0\rangle_1 |\varphi\rangle_2 |\phi\rangle_3 + |1\rangle_1 |\varphi\rangle_2 |\phi\rangle_3)$ 的通用量子线路图 Fig. 4. The general quantum circuit diagram for preparing quantum states $|0\rangle_1 |\varphi\rangle_2 |\phi\rangle_3 \xrightarrow{1:H} \frac{1}{\sqrt{2}} (|0\rangle_1 |\varphi\rangle_2 |\phi\rangle_3 + |1\rangle_1 |\varphi\rangle_2 |\phi\rangle_3)$.

根据量子力学基础理论,量子态 | ø > 也可以写 成如下形式:

$$|\phi\rangle = c_{11} |\phi_{11}\rangle |0\rangle + d_{11} |\phi_{12}\rangle |1\rangle,$$
 (17)

其中 c₁₁, d₁₁ 为任意实数, 且 c²₁₁ + d²₁₁ = 1. 采用 同样的计算推导方法,本文也可以设计出制备量子 态 |φ〉的通用量子线路.

在完成两个量子态 |φ〉 和 |φ〉 的制备后, 接下 来就是设计实现两个量子态内积的通用量子线路. 因为受控 SWAP 量子逻辑门可以实现两个量子比 特状态之间的转换,所以在量子态演化过程中,在 量子线路中增加控制位,主要利用 H量子逻辑门、 受控 SWAP 量子逻辑门实现两个量子态 $|\varphi\rangle$ 和 $|\phi\rangle$ 之间的转换,形成量子纠缠态,通过对控制位量子 态的测量求得两个量子态的内积. $|\varphi\rangle$ 和 $|\phi\rangle$ 的量 子态内积求解线路如图 5 所示,其中 H 量子逻辑 门的矩阵形式为

$$H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1\\ 1 & -1 \end{bmatrix}.$$
 (18)

SWAP 量子逻辑门的矩阵形式为

$$SWAP = \begin{bmatrix} 1000\\0010\\0100\\0001 \end{bmatrix}.$$
 (19)

求解量子态内积的计算推导如下:
$$\begin{split} |0\rangle_{1}|\varphi\rangle_{2}|\phi\rangle_{3} \xrightarrow{1:H} \frac{1}{\sqrt{2}} (|0\rangle_{1}|\varphi\rangle_{2}|\phi\rangle_{3} + |1\rangle_{1}|\varphi\rangle_{2}|\phi\rangle_{3}) \\ \xrightarrow{2 \text{ and } 3:SWAP,控制位为1}} \frac{1}{\sqrt{2}} (|0\rangle_{1}|\varphi\rangle_{2}|\phi\rangle_{3} + |1\rangle_{1}|\phi\rangle_{2}|\varphi\rangle_{3}) \\ \xrightarrow{1:H} \frac{1}{2}|0\rangle_{1} (|\varphi\rangle_{2}|\phi\rangle_{3} + |\phi\rangle_{2}|\varphi\rangle_{3}) \\ + \frac{1}{2}|1\rangle_{1} (|\varphi\rangle_{2}|\phi\rangle_{3} - |\phi\rangle_{2}|\varphi\rangle_{3}) . \end{split}$$

用标准基态 |0> 测量控制位得到 0 的概率为

$$\begin{split} & \left[\frac{1}{2}\left(\langle\phi|_{3}\langle\varphi|_{2}+\langle\varphi|_{3}\langle\phi|_{2}\right)\langle0|_{1}+\frac{1}{2}\left(\langle\phi|_{3}\langle\varphi|_{2}-\langle\varphi|_{3}\langle\phi|_{2}\right)\langle1|_{1}\right]|0\rangle_{1}\right.\\ & \times\left\langle0|_{1}\left[\frac{1}{2}|0\rangle_{1}\left(|\varphi\rangle_{2}|\phi\rangle_{3}+|\phi\rangle_{2}|\varphi\rangle_{3}\right)+\frac{1}{2}|1\rangle_{1}\left(|\varphi\rangle_{2}|\phi\rangle_{3}-|\phi\rangle_{2}|\varphi\rangle_{3}\right)\right]\\ & =\frac{1}{2}\left(\langle\phi|_{3}\langle\varphi|_{2}+\langle\varphi|_{3}\langle\phi|_{2}\right)\langle0|_{1}|0\rangle_{1}\langle0|_{1}\frac{1}{2}|0\rangle_{1}\left(|\varphi\rangle_{2}|\phi\rangle_{3}+|\phi\rangle_{2}|\varphi\rangle_{3}\right)=\frac{1}{2}\left(1+\langle\varphi\mid\phi\rangle^{2}\right). \end{split}$$

用标准基态 |1> 测量控制位得到 1 的概率为

$$\begin{split} & \left[\frac{1}{2}\left(\langle\phi|_{3}\langle\varphi|_{2}+\langle\varphi|_{3}\langle\phi|_{2}\right)\langle0|_{1}+\frac{1}{2}\left(\langle\phi|_{3}\langle\varphi|_{2}-\langle\varphi|_{3}\langle\phi|_{2}\right)\langle1|_{1}\right]|1\rangle_{1}\right.\\ & \left.\times\left\langle1|_{1}\left[\frac{1}{2}|0\rangle_{1}\left(|\varphi\rangle_{2}|\phi\rangle_{3}+|\phi\rangle_{2}|\varphi\rangle_{3}\right)+\frac{1}{2}|1\rangle_{1}\left(|\varphi\rangle_{2}|\phi\rangle_{3}-|\phi\rangle_{2}|\varphi\rangle_{3}\right)\right]\\ & =\frac{1}{2}\left(\langle\phi|_{3}\langle\varphi|_{2}-\langle\varphi|_{3}\langle\phi|_{2}\right)\langle1|_{1}|1\rangle_{1}\langle1|_{1}\frac{1}{2}|1\rangle_{1}\left(|\varphi\rangle_{2}|\phi\rangle_{3}-|\phi\rangle_{2}|\varphi\rangle_{3}\right)=\frac{1}{2}\left(1-\langle\varphi\mid\phi\rangle^{2}\right). \end{split}$$



量子态内积求解对应的量子线路图 图 5

Fig. 5. The quantum circuit diagram corresponding to solving the inner product of quantum states.

4 实验与分析

为了对本文提出的基于 QK-means 的量子增 强求解方法进行有效性和准确性验证,本节首先选 取 UCI(University of CaliforniaIrvine)数据库中 的 Haberman, Iris, Diabetes, Wine 四个公共数据 集作为本次实验的测试数据集,并结合经典 Kmeans 算法在相同条件下进行仿真实验与分析;其 次,以某多军兵种联合作战登陆战役模拟演习中收 集的关于作战平台实验数据为研究对象,将本文提 出的方法与经典的 K-means 算法同时用于求解平 台聚类问题,并对结果进行分析研究;最后,对本 文提出的基于 QK-means 的量子增强求解方法的 计算复杂度与经典 K-means 算法的计算复杂度进 行对比分析.

4.1 公共数据集仿真实验

Haberman, Iris, Diabetes, Wine 四个公共数 据集的基本信息如表1所示,本文依次将其作用于 两种算法进行实验验证.实验结果重点从准确率、 运行时间和迭代次数等方面对两种算法进行对比 分析.由于在对这两种算法进行实验测试时,在初 始聚类中心选取环节都具有一定的随机性,造成单 次实验结果可能出现较大偏差,为了避免这种误差 影响,本文设定在每个数据集上运行算法10次, 求其平均值作为最终结果.本文提出的方法的伪代 码如表2所示.

		表	1	实验	险	数	据	集	言	息	表		
1	4	D			1	1				c			

Table 1.	Experimental dataset information table.							
数据集	样本数	特征维度数	类别数					
Haberman	306	3	2					
Iris	150	4	3					
Diabetes	768	8	2					
Wine	178	13	3					

图 6 所示为基于 QK-means 的量子增强求解 方法和经典 K-means 算法在 4 种公共数据集上进 行聚类分簇的准确率比较图.由于选用的 4 个公共 数据集都是带标签的,每个数据样本都有自身对应 的分类标签,所以这里以结果实现正确划分的数据 样本个数与数据样本总数量的比值作为衡量两个 算法的准确程度.从图中通过分析可以得出,在 4 种数据集下本文提出的方法比经典 K-means 算法 都有较为显著的提升,主要原因是经典 K-means 算 法在确定初始聚类中心点位置时,采用的是随机在 所有数据样本中选取,而初始聚类中心点位置很容 易影响到聚类结果,造成实验结果与数据集自身标 签分类结果之间有偏差,容易陷入到局部最优解.

表 2 基于 QK-means 的量子增强求解方法的伪代码 Table 2. Pseudo code of the quantum enhancement solution method based on QK-means.

算法1. 基于QK-means的量子增强求解方法的伪代码
输入:输入数据集 <i>S</i> (<i>N</i> , <i>M</i> , <i>K</i>),其中 <i>N</i> 表示数据集样本数量, <i>M</i> 表示数据样本维度, <i>K</i> 表示数据分类个数.初始化量子软件开发环境与量子云平台
输出: K个聚类分簇以及每个分簇所包含的数据样本
初始化量子软件开发环境 Q_r 与量子比特数量
1) 根据输入数据集分类个数确定聚类中心数为K
2) 结合公共数据集实际情况, 根据3.1节中所述选取聚类中心点的方法, 将每个分簇数据集合的平均值作为初始聚类中心点位置
3) 对数据样本和聚类中心点进行量子化,并给SSE赋一个较大值
4) while SSE值 ≥ 阈值 do
5) 通过基于QK-means的量子线路制备量子态, 计算数据样本与各聚类中心点之间的欧氏距离 $D(x_i, c_k)$
6) 根据 $D(x_i, c_k)$, 当 $D(x_i, c_k)$ 取到 $\min_{i,k} D(x_i, c_k)$ 时, 将 $x_i \to C_k$
7) 计算每个 N_k , 并求该聚类分簇的平均值 $c'_k = \frac{1}{N_k} \sum_{x_i \in C_k} x_i$
8) 通过 c' _k 更新聚类中心位置
9) 求解整个聚类数据集的残差平方和 SSE = $\sum_{k=1}^{K} \sum_{x_i \in C_k} D(x_i, c_k) ^2$
10) end while
11) 输出每个 C_k

而本文针对初始聚类中心点位置选取方式进行了 优化,在确保合理性的前提下,缩小每个聚类中心 点选择范围,避免了不良初始聚类中心点数据带来 的较大实验结果误差.



图 6 在 4 种公共数据集下,两种算法的准确率比较图 Fig. 6. Comparison of accuracy between two algorithms on four common datasets.

图 7 为在 4 种公共数据集下两种算法的运行 时间对比图. 从图 7 可以分析得出, 在 4 种数据下 本文提出的方法的运行时间明显低于经典 K-means 算法的运行时间, 特别是当数据样本较大或者数据 维度较大时, 两种算法的运行时间对比更为明显. 通过分析认为主要原因是本文提出的方法利用量 子增强技术使其具有并行计算的能力, 大大缩短了 算法的运行时间, 而这种并行计算能力, 随着数据 量和数据维度增大, 表现出的优势更为明显. 另一 个原因是本文提出的方法优化了初始聚类中心点 的选取方式, 一定程度上通过减少迭代次数缩短了 算法的运行时间.



图 7 在 4 种公共数据集下, 两种算法的运行时间对比图 Fig. 7. Comparison of runtime between the two algorithms on four common datasets.

图 8 为在 4 种公共数据集下两种算法的迭代 次数对比图,图中迭代次数是多次试验测试的平均 值,所以不一定是整数值.从图中分析可以得出, 在4种数据下本文提出的方法的迭代次数少于经 典 K-means 算法的迭代次数, 特别是当数据样本 较为复杂时, 与经典 K-means 算法相比, 本文提出 的方法具有较大优势. 分析原因主要是本文提出 的方法在选择初始聚类中心点时,通过缩小每个 聚类中心点选取范围,将其限制在一个较优的数 值范围,降低了随机性给聚类结果带来的影响,从 而使算法可以较早取得较优解,减少了算法的迭 代次数. 当数据样本变得复杂时, 经典 K-means 算 法通过随机方式选取初始聚类中心点的局限性和 不确定性就越为突出,造成了算法迭代次数增多, 而本文提出的方法则基本不受数据样本复杂度的 影响.



图 8 在 4 种公共数据集下, 两种算法的迭代次数对比图 Fig. 8. Comparison of iteration times of two algorithms on four common datasets.

4.2 平台聚类问题求解

如图 9 所示为某多军兵种联合作战登陆战役 模拟演习项目的作战任务区域图. 该项目作战区域 主要由 A 区域与 B 区域组成,本文只对 A 区域中 的作战平台开展联合作战战役行动中平台聚类问 题研究.实验过程中, A 区共有作战平台类型 12 种, 数量 23 个,各平台包含的具体资源属性类型有 12 种,同类型的不同平台采用不同字符进行表示. 实验所需的数据是在战役演习项目中对 A 区域作 战平台数据进行收集整理的基础上,随机选取的 3 组作战平台数据,每组数据包含 23 个平台数据 样本.



图 9 作战任务区域图 Fig. 9. Operational mission area map.

针对联合作战战役行动中平台聚类问题,实验 所需搭建环境以及所采用的实验原理和方法都与 4.1节实验基本相同,这里不再赘述.针对本文提 出方法的伪代码,程序在执行中将聚类中心点选取 方式调整为利用每个任务簇的任务数据平均值 进行确定,并将该值作为对应平台簇的初始聚类中 心点.

图 10 为在 3 组平台数据下两种算法的实验结

果比较图.图 10(a) 为基于 QK-means 的量子增强 求解方法的实验结果,图 10(b) 为经典 K-means 算法的实验结果.从图中可以分析得出,两种算法 都可以对 3 组平台数据进行聚类分簇,但本文提出 的方法在聚类效果上要优于经典 K-means 算法, 其主要原因就是经典 K-means 算法在初始聚类中 心点选取时随机性较大,容易陷入局部较优解.

图 11 为在 3 组平台数据下两种算法的运行时 间对比图. 从图中可以分析得出, 在 3 组平台数据 下本文提出的方法的运行时间低于经典 K-means 算法的运行时间, 再次验证了本文提出的方法的量 子并行计算优势. 此外, 由于 3 组平台数据的规模 相同, 所以两种算法在 3 组数据下的运行时间也基 本相等.

图 12 为在 3 组平台数据下两种算法的迭代次 数对比图.由于图中迭代次数是经过多次试验测试 得到的平均值,所以不一定是整数值.从图 12 分 析可以得出,在 3 组数据下本文提出的方法的迭代 次数少于经典 K-means 算法的迭代次数,且由于 3 组平台数据的规模相同,两种算法在 3 组数据下 的迭代次数也基本相等.



图 10 在 3 组平台数据下, 两种算法的实验结果比较图 Fig. 10. Comparison of experimental results of two algorithms under three sets of platform data.



图 11 在 3 组平台数据下,两种算法的运行时间对比图 Fig. 11. Comparison of runtime between two algorithms under three sets of platform data.



图 12 在 3 组平台数据下,两种算法的迭代次数对比图 Fig. 12. Comparison of iteration times of two algorithms under three platform data groups.

4.3 复杂度分析

4.3.1 时间复杂度分析

根据经典 K-means 算法流程图 (图 1), 经典 K-means 算法中主要消耗资源的步骤为计算聚类 数据样本与各聚类中心点之间的欧氏距离、计算聚 类数据集的残差平方和等.根据文献 [35] 相关内 容,计算聚类数据样本与各聚类中心点之间的欧氏 距离部分的时间复杂度为O(KMN),计算聚类数据 集的残差平方和部分的时间复杂度也为O(KMN), 其中 K为聚类中心点数量, M为数据样本维度 数量, N为数据样本数量.所以如果算法迭代次 数为 Z, 则经典 K-means 算法总的时间复杂度为 O(2ZKMN).而根据基于 QK-means 的量子增强 求解方法的伪代码 (表 2),该方法主要针对经典 K-means 算法的计算聚类数据样本与各聚类中心 点之间的欧氏距离部分和计算聚类数据集的残差 平方和部分进行量子增强,实现并行计算.根据文 献 [36] 相关内容,在基于 QK-means 的量子增强 求解方法中,计算聚类数据样本与各聚类中心点之 间的欧氏距离部分的时间复杂度为O(KN log M), 计算聚类数据集的残差平方和部分的时间复杂 度也为O(KN log M),其中 K为聚类中心点数量, N为数据样本数量,M为数据样本维度数量.如果 算法迭代次数为 Z,则该方法总的时间复杂度为 O(2ZKN log M).从分析的结果可以看出本文所 提方法与经典 K-means 算法相比,较大幅度地缩 短了算法的时间复杂度,实现了量子增强.

4.3.2 空间复杂度分析

通过对经典 K-means 算法运行结构进行仔细 分析,该算法在进行聚类数据样本与各聚类中心点 之间的欧氏距离计算以及聚类数据集的残差平方 和计算时需要分配较多额外的辅助存储空间. 根据 文献 [35] 相关内容, 分析得出计算聚类数据样本与 各聚类中心点之间的欧氏距离部分空间复杂度为 O(MK), 计算聚类数据集的残差平方和部分空间 复杂度为O(M(N+K)),所以经典 K-means 算法 总的空间复杂度应该取较大值为O(M(N+K)), 其中 K为聚类中心点数量, N为数据样本数量, M为数据样本维度数量,而本文提出的方法在计 算欧氏距离时需要先对数据样本量子化,通过量子 化可以减少计算所需的空间资源,根据文献 [35] 相 关内容,该方法中的计算聚类数据样本与各聚类中心 点之间的欧氏距离部分空间复杂度为 $O(K \log M)$, 计算聚类数据集的残差平方和部分空间复杂度 为 $O((N + K) \log M)$,所以该方法总的空间复杂 度为 $O((N+K)\log M)$. 与经典 K-means 算法相 比,本文提出的方法在算法空间复杂度方面具有明 显降低.

5 结 论

针对联合作战战役行动中平台聚类编组问题, 本文提出了基于 QK-means 的量子增强求解方法. 首先,该方法在经典 K-means 算法基础上,结合具 体实际问题,对算法结果影响较大的初始聚类分簇 数量与初始聚类中心点位置等环节进行了优化处 理,降低了随机选取给算法带来的误差影响,避免 算法结果陷入局部最优解;其次,该方法针对经典 K-means 算法中需要消耗较大计算资源计算聚类 数据样本与各聚类中心点之间的欧氏距离部分和 计算聚类数据集的残差平方和部分,构建对应的量 子线路进行求解,实现量子增强与加速;最后,在 公共数据集上和具体的战役平台编组中对经典 Kmeans 算法和基于 QK-means 的量子增强求解方 法进行仿真验证与分析,实验结果表明,本文提出 的方法能较好地解决联合作战战役行动中平台聚 类编组问题,与经典 K-means 算法相比,在时间复 杂度和空间复杂度两个方面都有较大幅度降低.该 方法虽然便于理解,有较高的准确率,算法的复杂 度也大幅度降低,但该方法量子化程度有限,需要 较多的计算资源.如何提升基于 QK-means 的量 子增强求解方法的量子化程度是下一步我们重点 需要研究的方向.

参考文献

- Zhang S Y, Zhang W 2016 J. Equip. Acad. 27 36 (in Chinese)
 [张守玉, 张炜 2016 装备学院学报 27 36]
- [2] Wang X, Yao P Y, Zhang J Y, Wan L J, Jia F C 2019 J. Syst. Eng. Electron. 30 110
- [3] Yang Y 2023 Telecommun. Eng. 63 941 (in Chinese) [杨宇 2023 电讯技术 63 941]
- [4] Márquez C R, Braganholo V, Ribeiro C C 2024 Ann. Oper. Res. 338 05995
- [5] Macqueen J 1967 Proceedings of the 5th Berkley Symposium on Mathematical Statistics and Probability (Berkeley: University of California Press) p281
- [6] Lin Y S, Wang K D, Ding Z G 2023 Ieee Wirel. Commun. Le. 12 1130
- [7] Moyème K D, Yao B, Kwami S S, Pidéname T, Yendoubé L 2024 Energies 17 3022
- [8] Mottier M, Chardon G, Pascal F 2024 Ieee T. Aero. Elec. Sys. 60 3639
- [9] Ayad M J, Ku R K 2021 Indon. J. Electr. Eng. Co. 24 1744
- [10] Li Y X, Liu M L, Wang W C, Zhang Y H 2020 Ieee T. Multimedia 22 1385
- [11] Rani R S, Madhavan P, Prakash A 2022 Circ. Syst. Signal Pr. 41 3882

- [12] Al-Rahayfeh A, Atiewi S, Abuhussein A, Almiani M 2019 *Future Internet* 11 109
- [13] Tang D, Man J P, Tang L, Feng Y, Yang Q W 2020 Ad Hoc Netw. 102 102145
- [14] Pu Y N, Sun J, Tang N S, Xu Z B 2023 Image Vision Comput. 135 104707
- [15] Borzooei S, Miranda Ge H B, Abolfathi S, Scibilia G, Meucci L, Zanetti M C 2020 Water Sci. Technol. 81 1541
- [16] Culos A E, Andrews J L, Afshari H 2020 Commun. Stat-Simul C. 51 5658
- [17] Barkha N, Poonam V, Priya K 2016 IJLTET 7 121
- [18] Ikotun A M, Ezugwu A E, Abualigah L, Abuhaija B, Jia H E 2022 Inform. Sci. 622 178
- [19] Zhang Z B, Ling B W, Huang G H 2024 Ieee T. Signal Proces. 72 1348
- [20] Capó M, Pérez A, Lozano J A 2021 Ieee T. Neur. Net. Lear.
 32 2195
- [21] Wan B T, Huang W K, Pierre B, Cheng W W, Zhou S F 2024 Granular Comput. 9 45
- [22] Hamzehi M, Hosseini S 2022 Multimed. Tools Appl. 81 33233
- [23] Serkan T, Fatih O 2022 Appl. Sci. 12 11
- [24] Eissa M A Q 2022 Tehnički Glasnik 16 3
- [25] Wei R K, Liu Y, Song J K, Xie Y Z, Zhou K 2024 *Ieee T. Image Process.* **33** 1768
- [26] Pavan P, Vani B 2022 ECS Transactions 107 13055
- [27] Crognale M, Iuliis M D, Rinaldi C, Gattulli V 2023 Earthq. Eng. Eng. Vib. 22 333
- [28] Mohit M, Madhur M, Ketan L 2020 Int. J. Futur. Gener. Co. 13 28
- [29] Ibrahem A W, Hashim H A, AbdulKhaleq N Y, Jalal A A 2022 Indon. J. Electr. Eng. Co. 27 1151
- [30] Bezdan T, Stoean C, Naamany A A, Bacanin N, Rashid T A, Zivkovic M, Venkatachalam K 2021 Mathematics 9 1929
- [31] Tomesh T, Gokhale P, Anschuetz E R, Chong F T C 2021 *Electronics* 10 1690
- [32] Ouedrhiri O, Banouar O, Hadaj S E, Raghay S 2022 Concurr. Comp-Pract E. 34 e6943
- [33] Gong C G, Dong Z Y, Gani A, Han Q 2021 Quantum Inf. Process. 20 130
- [34] Zhang Y J, Mu X D, Guo L M, Zhang P, Zhao D, Bai W H
 2023 Acta Phys. Sin. 72 070302 (in Chinese) [张毅军, 慕晓冬,
 郭乐勐, 张朋, 赵导, 白文华 2023 物理学报 72 070302]
- [35] Liu X J, Yuan J B, Xu J, Duan B J 2018 J. Jilin Univ. (Eng. Ed.) 48 539 (in Chinese) [刘雪娟, 袁家斌, 许娟, 段博佳 2018
 吉林大学学报 (工学版) 48 539]
- [36] Rebentrost P, Mohseni M, Lloyd S 2014 Phys. Rev. Lett. 113 130503

Quantum enhanced solution method for platform clustering grouping based on quantum K-means

He Yi¹⁾²⁾ Zheng Kou-Quan¹⁾²⁾ Jing Feng¹⁾²⁾ Zhang Yi-Jun^{1)2)†} Wang Xun¹⁾ Liu Ying¹⁾²⁾ Zhao Le¹⁾

1) (Experimental Training Base, National University of Defense Technology, Xi'an 710106, China)

2) (Shaanxi Provincial Key Laboratory of Intelligent Collaborative Network Military-Civil Joint Construction, Xi'an 710106, China)

(Received 9 September 2024; revised manuscript received 27 October 2024)

Abstract

The paper proposes a quantum enhanced solution method based on quantum K-means for platform clustering and grouping in joint operations campaigns. The method first calculates the number of categories for platform clustering based on the determined number of task clusters, and sets the number of clustering categories in the classical K-means algorithm. By using the location information of the tasks, the clustering center points are calculated and derived. Secondly, the Euclidean distance is used as an indicator to measure the distance between the platform data and each cluster center point. The platform data are quantized and transformed into their corresponding quantum state representations. According to theoretical derivation, the Euclidean distance solution is transformed into the quantum state inner product solution. By designing and constructing a universal quantum state inner product solution quantum circuit, the Euclidean distance solution is completed. Then, based on the sum of squared errors of the clustering dataset, the corresponding quantum circuits are constructed through calculation and deduction. The experimental results show that compared with the classical K-means algorithm, the proposed method not only effectively solves the platform clustering and grouping problem on such action scales, but also significantly reduces the time and space complexity of the algorithm.

Keywords: QK-means algorithm, quantum enhancement, platform clustering

PACS: 03.67.Lx, 03.65.Aa

DOI: 10.7498/aps.73.20241265

CSTR: 32037.14.aps.73.20241265

[†] Corresponding author. E-mail: zhangyijun_gfkjdx@163.com

物理学报Acta Physica Sinica





Institute of Physics, CAS

基于量子K-means的平台聚类编组量子增强求解方法

何一 郑寇全 荆锋 张毅军 王勋 刘颖 赵乐

Quantum enhanced solution method for platform clustering grouping based on quantum K-means He Yi Zheng Kou-Quan Jing Feng Zhang Yi-Jun Wang Xun Liu Ying Zhao Le 引用信息 Citation: Acta Physica Sinica, 73, 230305 (2024) DOI: 10.7498/aps.73.20241265 在线阅读 View online: https://doi.org/10.7498/aps.73.20241265 当期内容 View table of contents: http://wulixb.iphy.ac.cn

您可能感兴趣的其他文章

Articles you may be interested in

用于等离子体相干模式自动识别的谱聚类算法实现

Implementation of spectral clustering algorithm for automatic identification of plasma coherence patterns 物理学报. 2022, 71(15): 155202 https://doi.org/10.7498/aps.71.20220367

量子态在两体和k体下基于min相对熵的量子关联测度 Quantum correlation measure based on min relative entropy for two-partition and k-partition 物理学报. 2023, 72(1): 010302 https://doi.org/10.7498/aps.72.20221293

量子噪声对Shor算法的影响

Effects of quantum noise on Shor's algorithm 物理学报. 2024, 73(5): 050301 https://doi.org/10.7498/aps.73.20231414

机器学习辅助绝热量子算法设计

Machine learning assisted quantum adiabatic algorithm design 物理学报. 2021, 70(14): 140306 https://doi.org/10.7498/aps.70.20210831

基于辅助单比特测量的量子态读取算法

A quantum state readout method based on a single ancilla qubit 物理学报. 2021, 70(21): 210303 https://doi.org/10.7498/aps.70.20211066

基于压缩态光场的量子增强型光学相位追踪

Quantum-enhanced optical phase tracking via squeezed state 物理学报. 2024, 73(5): 054203 https://doi.org/10.7498/aps.73.20231835