

基于量子计算的粗糙集核属性 求解算法

段隆振¹ 谢旭明^{1,2} 邱桃荣¹ 杨舒晴¹

摘要 粗糙集的核属性求解问题在经典计算中是一个 NP 问题. 现有的方法中最优的时间复杂度也需要 $O(|C||U|)$ (U 为论域、 C 为属性列数). 由于量子计算的并行性特点, 本文致力于采用量子计算的方法来求解粗糙集的核属性, 拟提出了一种基于量子计算的粗糙集核属性求解算法. 经过仿真实验, 在任何情况下, 该算法都能以 1 的总概

率得到目标分量; 且通过理论分析证明了算法的时间复杂度不会高于 $O\left(\left|\frac{\pi}{2 \arcsin \sqrt{\frac{M}{|U|}}}\right| + 1\right|U|)$.

关键词 量子计算, 粗糙集, 核属性, 算法设计

引用格式 段隆振, 谢旭明, 邱桃荣, 杨舒晴. 基于量子计算的粗糙集核属性求解算法. 自动化学报, 2020, 46(8): 1753–1758

DOI 10.16383/j.aas.2018.c170328

开放科学 (资源服务) 标识码 (OSID):



A Quantum Algorithm of the Computation of Core in Rough Set

DUAN Long-Zhen¹ XIE Xu-Ming^{1,2} QIU Tao-Rong¹
YANG Shu-Qing¹

Abstract The computation of core in rough set is an NP-hard problem in classic computing. The best time complexity of the existing algorithms is $O(|C||U|)$ (U stands for Universe, C stands for the number of attributes). Due to its ability to compute in parallel, quantum computing, is adopted to find the core in rough set in this research. Thus, a quantum algorithm of the computation of core in rough set is proposed. Based on simulation experiment, the proposed algorithm is able to get one target out of the target set with the probability of 1. And the time complexity of the new algorithm is proved to be no more than $O\left(\left|\frac{\pi}{2 \arcsin \sqrt{\frac{M}{|U|}}}\right| + 1\right|U|)$ by theoretical analysis.

Key words Quantum computing, rough set, core, algorithm design

Citation Duan Long-Zhen, Xie Xu-Ming, Qiu Tao-Rong, Yang Shu-Qing. A quantum algorithm of the computation of core in rough set. *Acta Automatica Sinica*, 2020, 46(8): 1753–1758

粗糙集理论^[1] 是波兰科学家 Pawlak 于 1982 年提出的一种刻画不完整性、不精确性、不完备性的数学理论. 自提出以来, 粗糙集理论在许多领域得到应用和发展. 黎明等^[2]

利用粗糙集数据分析方法构建了一种神经网络模型. 张迎春等^[3] 将新能量公式和粗糙集相结合来处理水平集图像分割问题. 在粗糙集理论研究中, 属性约简是核心内容之一, 而大多数属性约简算法都以核属性为基础. 目前常见的求核算法包括基于差别矩阵的求核算法、基于信息熵的求核算法、基于正区域的求核算法. Hu 等^[4] 提出了一种利用差别矩阵来求粗糙集核属性的方法. 叶东毅等^[5] 利用反例证明了 Hu 等的求核方法存在缺陷, 并提出一个基于新的差别矩阵求核算法. 王国胤^[6] 指出叶东毅等对 Hu 等方法缺陷分析的不正确, 并结合代数论和信息论深入分析了 Hu 等方法缺陷产生的原因, 且补充地提出一种基于信息熵的求核算法, 最后指明了 Hu 等方法、叶东毅等方法和提出算法的各自适用情况. 以上各种求核算法的时间复杂度最好也只能达到 $O(|C|^2|U|\log_2|U|)$. 后来, 徐章艳等^[7] 利用信息熵来简化差别矩阵, 提出的新算法将时间复杂度降为 $\max(O(|C||U/C|^2), O(|C||U|))$. 周江卫^[8] 采用基数排序思想计算粗糙集的正区域, 然后基于正区域来求解核属性, 该算法将时间复杂度降低为 $O(|C||U|)$. 此外, 尹林子等^[9] 将可辨识矩阵用在增量式属性约简中, 但其求核部分的时间复杂度并不优于已有算法. 虽然在经典计算中粗糙集的核属性能求出解, 但当粗糙集的属性列较多时, 需要的计算消耗也相当大.

量子计算^[10] 是量子力学和计算机科学相结合的一门新兴学科, 具有叠加、并行、纠缠等特性. 1985 年, Deutsch^[11] 提出的第一个量子算法, 高效地解决了判别一个函数是否为平衡函数的问题, 显示了量子计算强大的并行能力. 此后, Shor^[12] 提出的量子算法亦利用量子计算的并行性将大数质分解问题的时间复杂度从指数级降到多项式级. 另一个经典且具有代表性的量子算法 — Grover 算法^[13], 最初被设计来解决无序数据库搜索问题, 通过一个量子黑盒对同一个叠加态中的目标分量及非目标分量采用不同的操作, 以此来提高目标分量的概率幅, 压缩非目标分量的概率幅, 最终实现了以低的时间复杂度高概率地得到目标分量. 在 Grover 算法的启发下, 孙国栋等^[14] 提出 RF-Grover 算法, 将求根问题中的根作为目标分量, 其他作为非目标分量, 实现了求根问题的量子计算, 并将时间复杂度降为 $O\left(\sqrt{\frac{M}{k}}\right)$. 经分析, 求核问题的本质也是无序搜索的问题. 那么只要将核属性定义为目标分量, 非核属性定义为非目标分量, 再通过构建合适的量子黑盒来对两种分量进行不同操作, 求核问题则能用 Grover 算法来求解, 且有望在时间复杂度方面得到优化.

但 Grover 算法在目标分量占比不满足特定条件的时候不能以 1 的概率得到目标分量; 而且当目标分量占比超过半数时, 算法失效. 并且, 后续提出的改进算法中也存在相应的问题. Tulsi 等^[15] 后来提出一种基于固定 $\frac{\pi}{3}$ 相位旋转的改进算法, 算法整体提升了得到目标分量的概率, 但时间复杂度也增加了不少. 张煜东等^[16] 提出一种改进的 Grover 算法, 算法将搜索范围扩大了一倍, 解决了原算法失效的问题, 但仍不可以在任何情况下都以 1 的概率得到目标分量, 且算法的时间复杂度也增加了. 李盼池等^[17] 提出可变旋转轴的 Grover 改进算法, 通过改变两个算子的相位旋转角度, 实现了当目标分量与所有分量的比值大于 $\frac{3-\sqrt{5}}{8}$ 时, 总能以 1 的概率得到目标分量, 但当这个比值小于 $\frac{3-\sqrt{5}}{8}$ 时却不能保证以 1 的概率得到目标分量. 朱皖宁等^[18] 通过在 Grover 算法的模型上增加相位因子实现了算法迭代次数的自适应性, 但其目标分量的概率却不会优于 Grover 算法.

收稿日期 2017-06-16 录用日期 2018-01-01
Manuscript received June 16, 2017; accepted January 1, 2018
国家自然科学基金 (61070139, 81460769, 61762045), 江西省科技化项目 (20112BBG70087) 资助
Supported by National Natural Science Foundation of China (61070139, 81460769, 61762045), Scientific and Technological Project of Jiangxi Province (20112BBG70087)
本文责任编辑 张敏灵
Recommended by Associate Editor ZHANG Min-Ling
1. 南昌大学信息工程学院 南昌 330031 2. 南昌大学图书馆 南昌 330031
1. College of Information Engineering, Nanchang University, Nanchang 330031 2. Library of Nanchang University, Nanchang 330031

本文拟利用改进的 Grover 算法来求解粗糙集核属性, 主要研究内容包括两个方面: 一方面, 改进现有的 Grover 算法, 使得新算法能以概率 1 得到目标分量; 另一方面, 将改进的 Grover 算法应用在粗糙集的核属性求解上. 结合上述两研究内容得出的基于量子计算的求核算法能以 1 的总概率和不高于 $O\left(\left|\frac{\pi}{2 \arcsin \sqrt{\frac{M}{C}}}\right| + 1\right)\|U\|$ 的时间复杂度得到目标分量.

1 基础知识

1.1 粗糙集核属性

粗糙集的数据表示通常为一个信息系统. 下面引入信息系统及其核属性的定义.

定义 1. $S = (U, A, V, f)$ 是一个信息系统, 其中 U 表示非空有限对象集, 称为论域; A 是非空有限属性集; $V = \bigcup_{(a \in A)} V_a$, V_a 是属性 a 的值域; $f: U \times A \rightarrow V$ 表示一个信息函数, 它为每一个对象的每个属性赋予一个信息值, 即 $\forall a \in A, x \in U, f(x, a) \in V_a$.

任意属性子集 $B \subseteq A$ 决定了一个二元不可辨识关系 $IND(B)$.

$$IND(B) = \{(x, y) \in U \times U \mid \forall a \in A, f(x, a) = f(y, a)\} \quad (1)$$

定义 2. 设 $S = (U, A, V, f)$ 是一个信息系统, 对于 $a \in A$, 如果 $IND(A - \{a\}) = IND(A)$, 则可以认定 a 在 A 中是多余的属性 (或称为非核属性), 否则, 称 a 是 A 的必要的属性 (或称核属性).

非核属性在信息系统中起不到决定性的作用. 如果将一系列非核属性从信息系统中删除, 信息系统的分类能力不会受到影响; 若将一系列核属性从信息系统中删除, 那么它的分类能力一定会被改变.

1.2 Grover 算法

1.2.1 量子比特和等权重叠加态

在经典信息理论中, 位是信息量的基本单位. 而在量子信息理论中, 量子位是信息量的基本单位. 经典位只能处于 0 或 1 中的一种状态. 而量子位可以以叠加的方式同时处于量子 $|0\rangle$ 和 $|1\rangle$ 态. 处于叠加态的量子位可以表示为 $|\varphi\rangle = \alpha|0\rangle + \beta|1\rangle$, 其中复数 α 和 β 必须满足 $\alpha^2 + \beta^2 = 1$. 这个叠加态就意味着在测量的时候出现 $|0\rangle$ 的概率为 α^2 , 出现 $|1\rangle$ 的概率为 β^2 . 上面的叠加态中如有 $\alpha = \beta$, 那么就可以称这个量子位处于等权重叠加态, 即 $|0\rangle$ 和 $|1\rangle$ 出现的概率相同.

n 量子位的等权重叠加态可以通过 $H^{\otimes n}$ (H 门为一个单量子比特门) 作用到 n 量子位 $|0\rangle^{\otimes n}$ 得到. 等权重叠加态是实现 Shor 算法和 Grover 算法极为重要的一步.

1.2.2 G_a 算子 (量子黑盒)

量子黑盒是 G 算子的重要组成部分, 设其可以接收叠加态的输入. 首先, 量子黑盒需要结合实际处理的问题实现函数功能, 即判断分量. 若用于搜索无序数据库, 则要看分量上对应的结果是否为要找的值; 若用于求根, 则要看分量上对应的结果是不是根; 本文则是要判断分量上对应的属性是否为核属性. 根据函数确定目标分量和非目标分量后, 量子黑盒对应给出 1 和 0 的函数输出. Grover 算法量子黑盒在搜索

无序数据库时需要实现的函数功能如下:

$$\begin{cases} f(x) = 1, & x \text{ 对应的数据是目标结果} \\ f(x) = 0, & x \text{ 对应的数据非目标结果} \end{cases}$$

在量子黑盒实现上述函数之后, 再通过一定的操作实现只对目标分量的相位取反. 经过量子黑盒处理后的叠加态, 在目标分量上的相位与输入时相反, 在非目标分量上保持不变. 量子黑盒实现的功能以投影算子的形式可写作 $G_a: (I - 2|a\rangle\langle a|)$, 其中 $|a\rangle$ 为等权重叠加态中的目标分量.

1.2.3 G_s 算子

G_s 算子是一个均值翻转算子, 它的作用是将经过 G_a 处理后的新叠加态中的每一个分量的概率幅以该叠加态的平均幅值翻转. G_s 以投影算子的形式可写作 $G_s: (2|s\rangle\langle s| - I)$, 其中 $|s\rangle$ 为需要制备的等权重叠加态.

G_a 和 G_s 算子共同组成 G 算子, 其内部结构图如图 1 所示.

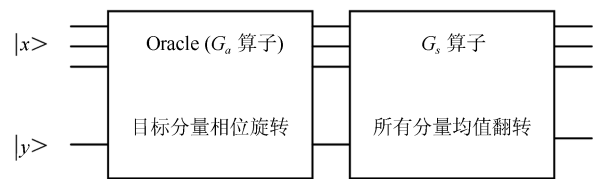


图 1 G 算子内部结构

Fig. 1 Inner structure of G

图 1 中叠加态 $|x\rangle$ 和 $|y\rangle$ 作为 G 算子的输入, $|x\rangle$ 包含目标和非目标分量, $|y\rangle$ 是用来辅助 G_a 算子实现目标分量取反的一个单量子位叠加态. 首先, G_a 算子作用在 $|x\rangle$ 上, 实现目标分量上的取反操作. 然后, G_s 算子作用在得到的叠加态上, 实现均值翻转. 那么, 经过 G 算子作用后, 除去某些特殊情况 (例如, 当目标与非目标分量个数相同时, 就只能实现相位的改变, 而没有幅值的改变), 目标分量的概率幅值就会不同于非目标分量的概率幅值. 经过一定次数的 G 算子迭代, 目标分量就能有个高的概率幅值, 从而算法得以实现.

1.2.4 Grover 算法描述

结合上述内容, Grover 算法的简化示意图可以绘制为图 2.

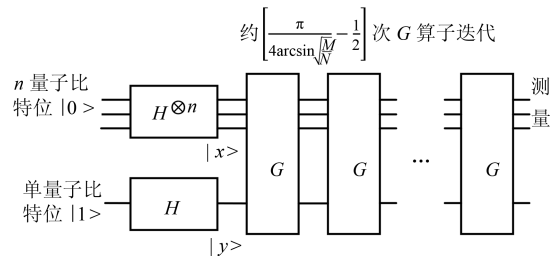


图 2 Grover 算法示意图

Fig. 2 Diagram of Grover algorithm

图 2 中 N 与 n 满足关系 $N = 2^n$, N 代表解空间所有分量的个数. $H^{\otimes n}$ 作用到 n 量子位的位态 $|0\rangle^{\otimes n}$ 上可以得到一个等权重叠加态 $|x\rangle$, $|x\rangle$ 再经过一定次数的 G 算子迭代, 然后得到一个目标分量概率幅较大的叠加态.

分析运算过程可以得出, G 算子迭代次数至关重要, 只有迭代合适的次数才能达到目标分量幅值的峰值. 由文献

[13] 可知, 若所有分量的个数为 N , 目标分量的个数为 M , 则可以推算出 Grover 算法的迭代次数约为式 (2) 的整数次.

$$\frac{\pi}{4 \arcsin \sqrt{\frac{M}{N}}} - \frac{1}{2} \quad (2)$$

2 基于量子计算的粗糙集核属性求解

本节先提出关于 Grover 算法的两个定理且给出证明, 再给出构建求核算子的方法, 最后描述了基于量子计算的求核算法.

2.1 定理及证明

定理 1. 当目标分量的个数为 M 时, 若存在正整数 T 使得 $\frac{M}{N} = \left(\sin \frac{\pi}{4T+2}\right)^2$ 成立, 那么 Grover 算法在迭代 T 次后总能以 1 的总概率得到目标分量.

证明. 假设一共有 N 个分量, 其量子态表示为 $|s\rangle$, 其中 M 个为目标分量, 其量子态表示为 $|a\rangle$. 那么在 $|a\rangle$ 和 $|s\rangle$ 张开的平面上, $|s\rangle$ 与 $|a\rangle$ 的垂直量子态 $|a^\perp\rangle$ 的夹角 θ 就满足 $\sin \theta = \sqrt{\frac{M}{N}}$.

G_a 的作用是将 $|s\rangle$ 中 $|a\rangle$ 分量转动 π 的相位, 即 $|a\rangle$ 变成 $-|a\rangle$. 所以当 G_a 作用于 $|s\rangle$ 时, 就相当于 $|s\rangle$ 相对 $|a^\perp\rangle$ 反射到 $|s'\rangle$. 相似地, G_s 将 $|s'\rangle$ 相对 $|s\rangle$ 反射到 $|s''\rangle$. 那么迭代一次后的 $|s''\rangle$ 与 $|a^\perp\rangle$ 的夹角就变成了 3θ . 以此类推, 迭代 T 次后的 $|s^T\rangle$ 与 $|a^\perp\rangle$ 的夹角即为 $(2T+1)\theta$.

当存在正整数 T 使得 $\frac{M}{N} = \left(\sin \frac{\pi}{4T+2}\right)^2$ 成立时, 将 $\sin \theta = \sqrt{\frac{M}{N}}$ 代入 $(2T+1)\theta$ 中, 则可计算出 $|s^T\rangle$ 与 $|a^\perp\rangle$ 的夹角为 $\frac{\pi}{2}$, 这就是说 $|s^T\rangle$ 和 $|a\rangle$ 正好重叠, 也就是说 $|s^T\rangle$ 中 $|a\rangle$ 分量的概率幅为 1. \square

定理 2. 当目标分量的个数为 M 时, 若不存在正整数 T 使得 $\frac{M}{N} = \left(\sin \frac{\pi}{4T+2}\right)^2$ 成立, 那么一定存在一个相位旋转角度 φ , 在 Grover 算法迭代 T' (T' 满足 $(2T'+1)\theta < \frac{\pi}{2} < (2T'+3)\theta$) 次后, 再以 φ 的相位旋转角度构造算子 G'_a 和 G'_s . 迭代一次, 则可以以 1 的总概率得到目标分量.

证明. 分析定理 1 的推理过程, 很容易得出 Grover 算法迭代 T' 次后 $|s^{T'}\rangle$ 与 $|a\rangle$ 的夹角 $\theta' = (2T'+1)\theta$, 且 θ' 的取值范围为 $\frac{(2T'+1)\pi}{4T'+6} < \theta' < \frac{\pi}{2}$. 此时, 目标分量上的概率幅为 $\frac{e^{i\varphi} \sin \theta'}{\sqrt{M}}$, 非目标分量上的概率幅为 $\frac{\cos \theta'}{\sqrt{N-M}}$.

以 φ 的相位旋转角度构造的算子 G'_a 和 G'_s , 可以表示为将原有的两个算子分别改成 $I - (1 - e^{i\varphi})|a\rangle\langle a|$ 和 $(1 - e^{i\varphi})|s^{T'}\rangle\langle s^{T'}| - I$. 根据叠加态的归一化定理, 若要目标分量的概率幅为 1, 那么非目标分量的概率幅就必须为 0. $|s^{T'}\rangle$ 经过 G'_a 作用后的叠加态中非目标分量的概率幅保持不变, 即为 $\frac{\cos \theta'}{\sqrt{N-M}}$; 再经过 G'_s 作用后的叠加态中非目标分量的概率幅变成 $\frac{\cos \theta'}{\sqrt{N-M}} \left(\frac{(1 - e^{i\varphi})(\cos \theta')^2}{N-M} - 1 \right) + \frac{\cos \theta'}{\sqrt{N-M}} \frac{N-M-1}{N-M} (1 - e^{i\varphi}) (\cos \theta')^2 + \frac{e^{i\varphi}(1 - e^{i\varphi})(\sin \theta')^2 \cos \theta'}{\sqrt{N-M}}$. 那么只要令非目标分量的概率幅为 0 则可以求出相位旋转角度 $\varphi = \arccos \left(1 - \frac{1}{2(\sin \theta')^2} \right)$, 由于 $\frac{(2T'+1)\pi}{4T'+6} < \theta' < \frac{\pi}{2}$, 所以 φ 必有解. \square

2.2 求核算法的 G 算子及 G' 算子

本文求核算法要实现的函数功能类似 Grover 算法的函

数功能. 需要实现的功能是利用输入的粗糙集来实现对分量是否为目标分量进行判断, 函数可以表示如下:

$$\begin{cases} f(x) = 1, & x \text{ 对应属性是核属性} \\ f(x) = 0, & x \text{ 对应属性非核属性} \end{cases}$$

根据粗糙集来判断一系列属性是否为核属性, 经典计算中已经给出了很多判别方法. 把粗糙集结合判别逻辑融入到量子黑盒中, 那么上述函数就可以实现了. 与 Grover 算法的量子黑盒相同, 该求核算法的量子黑盒也要求能接收叠加态的输入. 量子黑盒的函数功能实现和接收量子态的实现, 以现有的技术并做不到, 所以这里进行的是理论研究. 另外, 该文的仿真实验也是仿真量子运算的逻辑过程.

2.2.1 求核 G 算子

根据构造的量子黑盒函数, 制备 G_a 算子使其实现 $G_a : (I - 2|a\rangle\langle a|)$, 制备 G_s 算子使其实现 $G_s : (2|s\rangle\langle s| - I)$. G 算子由 G_a 和 G_s 算子组成.

2.2.2 求核 G' 算子

根据定理 2 和构造的量子黑盒函数, 制备 G'_a 算子使其实现 $G'_a : I - (1 - e^{i\varphi})|a\rangle\langle a|$, 制备 G'_s 算子使其实现 $G'_s : (1 - e^{i\varphi})|s\rangle\langle s| - I$. G' 算子由 G'_a 和 G'_s 算子组成.

G'_a 和 G'_s 算子在 $\varphi = \pi$ 时等价于 G_a 和 G_s 算子. 关于 Grover 算法中 π 相位旋转和其他相位角旋转的相关知识可以查阅文献 [15].

2.3 基于量子计算的求核算法

设待求解的粗糙集为 $S = (U, A, V, f)$, 并设粗糙集的属性列数为 C (C 对应 Grover 算法里的 N), C 为 2 的 n 次幂 (当 C 不满足 2 的 n 次幂时, 增加不影响分类关系的属性列使其满足 2 的 n 次幂要求). 根据上面的定理 1 和定理 2, 粗糙集核属性求解算法具体步骤如下:

- 1) 制备 n 量子位的等权重叠加态 $|s\rangle$, 依据粗糙集 S 制备 $G_a : (I - 2|a\rangle\langle a|)$ 算子, 根据 $|s\rangle$ 制备 $G_s : (2|s\rangle\langle s| - I)$ 算子.
- 2) 通过文献 [19] 中的算法确定 S 中核属性的个数 M .
- 3) 若存在正整数 T 使得 $\frac{M}{N} = \left(\sin \frac{\pi}{4T+2}\right)^2$ 成立, 则跳转到步骤 4), 否则跳转到步骤 6).
- 4) 将 $|s\rangle$ 执行 T 次 G 算子迭代, 其中包括:
 - a) 将 G_a 作用于 $|s\rangle$ 得到 $|s'\rangle$.
 - b) 将 G_s 作用于 $|s'\rangle$ 得到 $|s''\rangle$.
 - c) 令 $|s\rangle = |s''\rangle$, $|s''\rangle$ 为前 n 个量子位经过 $G_s G_a$ 算子作用后的叠加态.
- 5) T 次 G 算子迭代后转到步骤 10).
- 6) 依据 M 求出迭代次数 T' 和相位旋转角度 φ .
- 7) 将 $|s\rangle$ 执行 T' 次 G 算子迭代, 其中包括:
 - a) 将 G_a 作用于 $|s\rangle$ 得到 $|s'\rangle$.
 - b) 将 G_s 作用于 $|s'\rangle$ 得到 $|s''\rangle$.
 - c) 令 $|s\rangle = |s''\rangle$, $|s''\rangle$ 为前 n 个量子位经过 $G_s G_a$ 算子作用后的叠加态.
- 8) 依据 φ 和粗糙集 S 制备 $G'_a : I - (1 - e^{i\varphi})|a\rangle\langle a|$ 算子, 根据 φ 和 $|s\rangle$ 制备 $G'_s : (1 - e^{i\varphi})|s\rangle\langle s| - I$ 算子.
- 9) 将 $|s\rangle$ 执行 1 次 G' 算子迭代, 其中包括:
 - a) 将 G'_a 作用于 $|s\rangle$ 得到 $|s'\rangle$.
 - b) 将 G'_s 作用于 $|s'\rangle$ 得到 $|s''\rangle$.
 - c) 令 $|s\rangle = |s''\rangle$, $|s''\rangle$ 为前 n 个量子位经过 $G'_s G'_a$ 算子作用后的叠加态.

10) 算法结束, 测量 $|s\rangle$.

3 仿真实验及分析

本节将本文算法与主要的几种量子搜索算法进行了仿真实验, 并对比分析了仿真实验的结果, 借此来说明基于量子计算的核属性求解算法的有效性. 在测试数据集方面, 本文先选了部分 UCI 数据集; 考虑到 UCI 数据集不能体现本文算法的全面情况, 实验也采用了一组自建数据集.

3.1 数据集

1) UCI 数据集

本实验选取了 UCI 数据集中的 Iris 数据集、Soybean 数据集、Dermatology 数据集、Balloons 数据集和 Breast Cancer 数据集.

2) 自建数据集

实验者自建数据集为: 构造一个 64 行 64 列的矩阵, 矩阵的斜对角线上由整数 0 至 63 构成, 矩阵其他位置的数值均为 0.

$$\begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 63 \end{bmatrix} \quad (3)$$

数据集 I_i 由上述矩阵的第一行至第 $i+1$ 行构成, 其中, $i = 1, 2, \dots, 63$. 由此, 上述矩阵可以产生 63 个数据集 (数据集 I_1, I_2, \dots, I_{63}). 显而易见, 数据集 I_i 的核属性集包含第 2 至第 $i+1$ 列共 i 列属性.

3.2 仿真环境

系统: 64 位 Windows7 系统、2GB 内存、Intel Core i5 处理器.

软件: MATLAB 2012R.

3.3 测试结果及分析

1) UCI 数据集测试结果及分析

该部分实验将本文算法与基于 Grover 算法、文献 [15] 以及文献 [16] 的求核算法进行比较. 数据集采用的是第 3.1 节 1) 部分选取的 UCI 数据集. 实验中对比了不同算法的量子迭代次数和最终得到叠加态中目标分量的总概率. 实验结果如表 1 所示.

从表 1 中可以看出: 1) Grover 算法存在失效的问题, 当算法不失效的时候, 都能以较高的概率得到目标分量. Grover 算法的平均迭代次数是最少的. 2) 文献 [15] 和文献 [16] 中的算法解决了 Grover 算法失效的问题, 但在目标分量的概率上并不是总高于 Grover 算法. 除此之外, 这两种算法的迭代次数要明显高于 Grover 算法. 3) 本文算法在目标分量概率上表现突出, 不仅不存在失效问题, 而且总能以 1 的概率得到目标分量. 在平均迭代次数方面, 本文算法略高于 Grover 算法, 但明显少于文献 [15] 和文献 [16] 中的算法.

虽然 UCI 数据集的实验结果已经可以部分地说明本文算法的优势, 但要了解整体的情况, 还需要更全面的数据集. 很多文献 [15-17] 都体现出 Grover 算法的结果都与目标分量占总分量的比例有关. 因此, 本实验构造了核属性占比不同的 63 个数据集 (数据集 I_i 包含 i 个核属性), 其实验结果见下文.

表 1 UCI 数据集实验结果

Table 1 The test result of UCI sets

数据集	算法	迭代次数	概率
Iris 数据集	Grover 算法	失效	失效
Iris 数据集	文献 [15] 算法	2	0.9686
Iris 数据集	文献 [16] 算法	1	1.0000
Iris 数据集	本文算法	1	1.0000
Soybean 数据集	Grover 算法	1	0.8815
Soybean 数据集	文献 [15] 算法	3	0.9723
Soybean 数据集	文献 [16] 算法	2	0.9766
Soybean 数据集	本文算法	2	1.0000
Dermatology 数据集	Grover 算法	6	0.9983
Dermatology 数据集	文献 [15] 算法	12	1.0000
Dermatology 数据集	文献 [16] 算法	8	0.9978
Dermatology 数据集	本文算法	6	1.0000
Balloons 数据集	Grover 算法	失效	失效
Balloons 数据集	文献 [15] 算法	2	0.9686
Balloons 数据集	文献 [16] 算法	1	1.0000
Balloons 数据集	本文算法	1	1.0000
Breast Cancer 数据集	Grover 算法	1	1.0000
Breast Cancer 数据集	文献 [15] 算法	3	0.9688
Breast Cancer 数据集	文献 [16] 算法	2	0.9451
Breast Cancer 数据集	本文算法	1	1.0000

2) 自建数据集测试结果及分析

这部分实验依然是将本文算法与基于 Grover 算法、文献 [15] 以及文献 [16] 的求核算法进行比较. 为了更全面地体现量子迭代次数和最终得到叠加态中目标分量的总概率与核属性不同占比的关系, 这部分实验采用了第 3.1 节 2) 部分的自建数据集. 实验得到的数据被绘成图 3 和图 4.

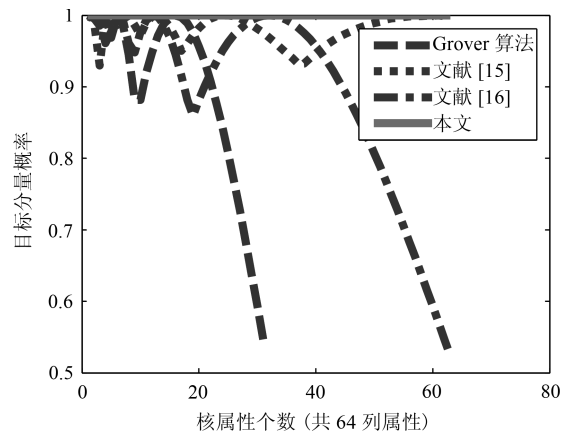


图 3 核属性个数与目标分量概率的关系

Fig. 3 Comparing the number of core attributes and percentage of targets

从图 3 中可以看出: Grover 算法在核属性不超过半数时总能以 0.5 以上的概率得到目标分量, 当核属性超过半数时算法失效; 文献 [15] 的算法明显优于 Grover 算法, 总能以 0.9 以上的概率得到目标分量; 文献 [16] 中的算法解决了 Grover 算法中的失效问题, 且在目标分量概率上的表现优于 Grover 算法, 但不如文献 [15] 的算法; 本文算法总能以 1 的概率得到目标分量. 总的来说, 在目标分量概率方面, 本文算法明显优于已见的几种算法.

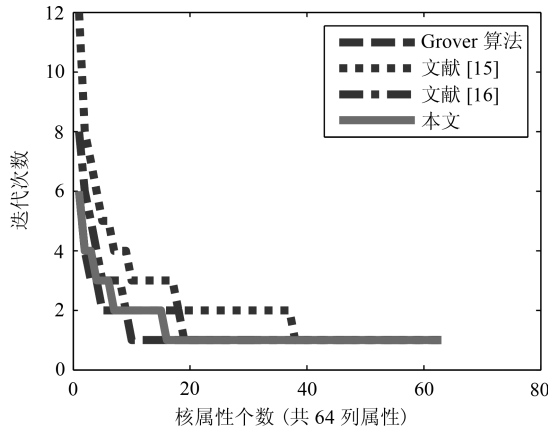


图 4 核属性个数与迭代次数的关系

Fig. 4 Comparing the number of core attributes and the number of iterations

从图 4 中可以看出: Grover 算法的算子迭代次数最少; 文献 [15] 和文献 [16] 中算法的算子迭代次数要明显高于 Grover 算法; 本文算法的算子迭代次数, 明显低于文献 [15] 和文献 [16] 中算法的算子迭代次数, 略高于 Grover 算法的算子迭代次数。

3.4 小结

本节对本文算法和几种现有算法进行了对比实验。从实验中可以得出, 在目标分量概率方面, 本文算法总能以 1 的概率得到目标分量, 优于其他几种算法; 在算子迭代次数方面, 本文算法略次于 Grover 算法, 但优于其他的改进算法。

4 算法效率分析

由于量子计算机在目前还没有真正实现, 故无法用真正的量子计算机来完成本实验。量子计算的仿真实验是通过分别计算每一个分量的结果再组合到一起来实现算子。所以在仿真实验中只能体现出算子的迭代次数及目标分量概率上的情况, 而不能体现出整体运算时间的情况。因此, 我们用理论证明的方式来分析本文算法的效率。

本文算法的时间复杂度 = 求解核属性个数 M 的时间复杂度 + 算子的时间复杂度 \times 算子的迭代次数。先设待求解粗糙集的属性个数为 C , 那么有求解核属性个数 M 的时间复杂度、算子的时间复杂度和算子的迭代次数如下:

1) 求解核属性个数 M 的时间复杂度

通过文献 [19] 可以得出求解 M 的时间复杂度不高于

$$O\left(\left|\frac{\pi}{4 \arcsin \sqrt{\frac{M}{C}}} + \frac{1}{2}\right| |U|\right).$$

2) 算子的时间复杂度

在经典计算中采用正区域的求核方法需要 $O(|C||U|)$ 的时间复杂度去验证每一列属性从而确定粗糙集的核属性。在量子计算中, 由于所有属性都叠加到一个叠加态当中, 验证属性是否为核属性的方法被制作成一个算子, 那么对于一个算子来说只是验证了一个叠加的属性是否为核属性。在这里我们设粗糙集的属性个数正好满足 $C = 2^n$, 那么算子的时间复杂度则为 $O(|U|)$ 。

3) 算子的迭代次数

通过定理 1 和定理 2 可得: a) 理想状态下, 本文算法的迭代次数正好等于 Grover 算法的迭代次数; b) 非理想状态下, 本文算法迭代次数要么等于 Grover 算法的迭代次数,

要么比 Grover 算法的迭代次数多一次。结合 a) 和 b), 以及 Grover 算法的迭代次数 $O\left(\frac{\pi}{4 \arcsin \sqrt{\frac{M}{C}}} - \frac{1}{2}\right)$, 本文算法的算子迭代次数一定不多于 $O\left(\frac{\pi}{4 \arcsin \sqrt{\frac{M}{C}}} + \frac{1}{2}\right)$ 次。

综上所述, 本文算法的时间复杂度一定不会高于 $O\left(\left|\frac{\pi}{2 \arcsin \sqrt{\frac{M}{C}}} + 1\right| |U|\right)$ 。通过时间复杂度的公式可以看出, 本文算法的时间复杂度并不是单独受属性列规模的影响, 而是与核属性列数与总的属性列数比值相关, 这个比值越大, 时间复杂度越小, 反之亦然。

5 结束语

粗糙集的核属性求解在经典计算中属于一个 NP 问题。由于量子计算的并行性特性, 本文尝试将量子计算引入粗糙集的核属性求解上, 以期以 1 的概率得到目标分量。仿真实验结果表明, 在核属性个数任意的情况下, 本文算法总能以 1 的概率得到目标分量, 且迭代次数的均值只略高于 Grover 算法。此外, 通过理论分析, 证明了本文算法的时间复杂度不会高于 $O\left(\left|\frac{\pi}{2 \arcsin \sqrt{\frac{M}{C}}} + 1\right| |U|\right)$ 。

该研究的不足之处在于, 目前其只能利用量子算法解决核属性求解的问题, 但对于求解核属性后如何做粗糙集的属性约简还不能进行有效的处理。因此, 未来的研究方向为在求解核属性之后, 如何利用量子计算再对粗糙集进行属性约简。

References

- 1 Pawlak Z. Rough sets. *International Journal of Computer & Information Sciences*, 1982, **11**(5): 341–356
- 2 Li Ming, Zhang Hua-Guang. Research on the method of neural network modeling based on rough sets theory. *Acta Automatica Sinica*, 2002, **28**(1): 27–33
(黎明, 张化光. 基于粗糙集的神经网络建模方法研究. *自动化学报*, 2002, **28**(1): 27–33)
- 3 Zhang Ying-Chun, Guo He. Level set image segmentation based on rough set and new energy formula. *Acta Automatica Sinica*, 2015, **41**(11): 1913–1925
(张迎春, 郭禾. 基于粗糙集和新能量公式的水平集图像分割. *自动化学报*, 2015, **41**(11): 1913–1925)
- 4 Hu X H, Cercone N. Learning in relational databases: A rough set approach. *Computational Intelligence*, 1995, **11**(2): 323–338
- 5 Ye Dong-Yi, Chen Zhao-Jiong. A new discernibility matrix and the computation of a core. *Acta Electronica Sinica*, 2002, **30**(7): 1086–1088
(叶东毅, 陈昭炯. 一个新的差别矩阵及其求核方法. *电子学报*, 2002, **30**(7): 1086–1088)
- 6 Wang Guo-Yin. Calculation methods for core attributes of decision table. *Chinese Journal of Computers*, 2003, **26**(5): 611–615
(王国胤. 决策表核属性的计算方法. *计算机学报*, 2003, **26**(5): 611–615)
- 7 Xu Zhang-Yan, Yang Bing-Ru, Cai Wei-Dong, Cui Wei, Gu Dong-Yuan. Quick algorithm for computing core based on

- the positive region. *Systems Engineering and Electronics*, 2006, **28**(12): 1902–1905, 1931
(徐章艳, 杨炳儒, 蔡卫东, 崔巍, 谷冬元. 一个基于正区域的快速求核算法. *系统工程与电子技术*, 2006, **28**(12): 1902–1905, 1931)
- 8 Zhou Jiang-Wei, Feng Bo-Qin, Liu Yang. New algorithm for quick computing core. *Journal of Xi'an Jiaotong University*, 2007, **41**(6): 688–691
(周江卫, 冯博琴, 刘洋. 一种新的快速求核算法. *西安交通大学学报*, 2007, **41**(6): 688–691)
- 9 Yin Lin-Zi, Yang Chun-Hua, Wang Xiao-Li, Gui Wei-Hua. An incremental algorithm for attribute reduction based on labeled discernibility matrix. *Acta Automatica Sinica*, 2014, **40**(3): 397–404
(尹林子, 阳春华, 王晓丽, 桂卫华. 基于标记可辨识矩阵的增量式属性约简算法. *自动化学报*, 2014, **40**(3): 397–404)
- 10 Zhang Jing, Li Chun-Wen, Wu Re-Bing. Coherent control modeling of quantum computers in open environments. *Acta Automatica Sinica*, 2005, **31**(5): 759–764
(张靖, 李春文, 吴热冰. 开放环境下多比特量子计算机的相干控制模型. *自动化学报*, 2005, **31**(5): 759–764)
- 11 Deutsch D. Quantum theory, the church-turing principle and the universal quantum computer. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 1985, **400**(1818): 97–117
- 12 Shor P W. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Journal on Computing*, 1997, **26**(5): 1484–1509
- 13 Grover L. A fast quantum mechanical algorithm for database search. In: *Proceedings of the 28th Annual ACM Symposium on the Theory of Computing*. New York, USA: IEEE, 1996. 212–219
- 14 Sun Guo-Dong, Su Sheng-Hui, Xu Mao-Zhi. Quantum mechanical algorithms for solving root finding problem. *Journal of Beijing University of Technology*, 2015, **41**(3): 366–371
(孙国栋, 苏胜辉, 徐茂智. 求根问题的量子计算算法. *北京工业大学学报*, 2015, **41**(3): 366–371)
- 15 Tulsi T, Grover L K, Patel A. A new algorithm for fixed point quantum search. *Quantum Information & Computation*, 2006, **6**(6): 483–494
- 16 Zhang Yu-Dong, Wei Geng, Wu Le-Nan. An improved Grover quantum searching algorithm. *Signal Processing*, 2009, **25**(2): 256–259
(张煜东, 韦耿, 吴乐南. 一种改进的 Grover 量子搜索算法. *信号处理*, 2009, **25**(2): 256–259)
- 17 Li Pan-Chi, Li Shi-Yong. Grover's quantum search algorithm based on adaptive phase rotation. *Journal of System Simulation*, 2009, **21**(12): 3557–3560
(李盼池, 李士勇. 基于自适应相位旋转的 Grover 量子搜索算法. *系统仿真学报*, 2009, **21**(12): 3557–3560)
- 18 Zhu Wan-Ning, Chen Han-Wu. Grover auto-control searching algorithm. *Acta Electronica Sinica*, 2016, **44**(12): 2975–2980
(朱皖宁, 陈汉武. 迭代次数自适应的 Grover 算法. *电子学报*, 2016, **44**(12): 2975–2980)
- 19 Brassard C H, Hoyer P, Tapp A. Quantum counting. In: *Proceedings of the 25th International Colloquium Automata Language and Programming*. Berlin, Germany: Springer-Verlag, 1998. 820–831
- 段隆振 南昌大学教授. 主要研究方向为数据挖掘, 并行计算与大数据.
E-mail: lzhduan@126.com
(DUAN Long-Zhen Professor at Nanchang University. His research interest covers data mining, parallel computing and big data.)
- 谢旭明 南昌大学博士研究生. 主要研究方向为粗糙集与量子计算.
E-mail: xiexuming@ncu.edu.cn
(XIE Xu-Ming Ph. D. candidate at Nanchang University. His research interest covers rough set and quantum computing.)
- 邱桃荣 南昌大学教授. 主要研究方向为粗糙集, 粒计算与量子计算. 本文通信作者.
E-mail: qiutaorong@ncu.edu.cn
(QIU Tao-Rong Professor at Nanchang University. His research interest covers rough set, granular computing and quantum computing. Corresponding author of this paper.)
- 杨舒晴 南昌大学博士研究生. 主要研究方向为粗糙集与量子计算.
E-mail: ysqsinuo@163.com
(YANG Shu-Qing Ph. D. candidate at Nanchang University. Her research interest covers rough set and quantum computing.)