北京邮电大学

第9讲 量子关联规则挖掘算法

高　飞

网络空间安全学院

提纲

- ➢ 量子关联规则挖掘算法
  - ❑ 关联规则挖掘
  - ❑ 量子算法
  - ❑ 复杂度分析
  - ❑ 总结与展望

问题描述

- 问题：从交易数据库中找出频繁被顾客购买的商品组合（项集），即找出购买频率（支持度）超过某个人为设定的阈值（如50%）的商品组合（频繁项集）。
- 有趣的例子：沃尔玛超市数据分析人员发现{啤酒、尿不湿}频繁被一起购买，后来调查得知原因：已婚男子下班后在给小孩买尿不湿的同时，顺带也买了些啤酒。后来，工作人员根据这一现象将啤酒、尿不湿放在一起摆放，增加销售量。

交易	商品（项）
T_0	面包、奶酪、牛奶
T_1	面包、黄油
T_2	奶酪、牛奶
T_3	面包、奶酪
T_4	奶酪、黄油、牛奶

k-项集：k个项（商品）的集合

例如，设定阈值=50%

频繁1-项集有：

{面包} 60%, {奶酪} 80%, {牛奶} 60%

频繁2-项集有：

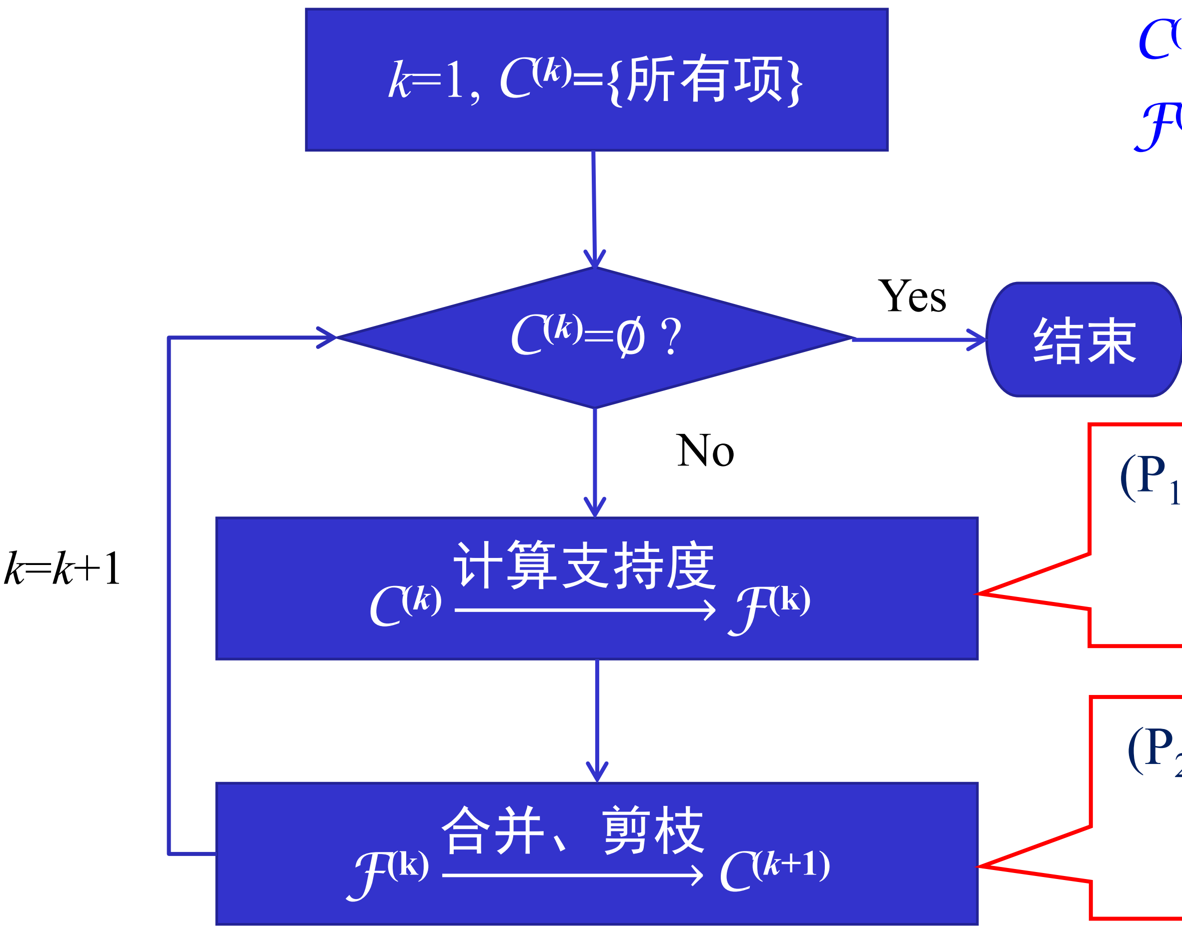
{奶酪，牛奶} 60%

经典Apriori算法

- 从数据库中找出所有的频繁项集（频繁1-项集、频繁2-项集、…）

$\mathcal{C}^{(k)}$：候选k-项集集合

$\mathcal{F}^{(k)}$：频繁k-项集集合



(P_1)计算每个候选k-项集的支持度并找出频繁k-项集
计算量高、核心步骤

(P_2)利用找到的频繁k-项集产生候选(k+1)项集
计算量低

经典Apriori算法：例子

交易	商品（项）
T_0	面包、奶酪、牛奶
T_1	面包、黄油
T_2	奶酪、牛奶
T_3	面包、奶酪
T_4	奶酪、黄油、牛奶

设定阈值=50%

候选1-项集：{面包(60%), 奶酪(80%), 牛奶(60%), 黄油(40%)}

频繁1-项集：{面包(60%), 奶酪(80%), 牛奶(60%)}

核心步骤

候选2-项集：{{面包 奶酪}(40%), {面包 牛奶}(20%), {奶酪, 牛奶}(60%)}

频繁2-项集：{{奶酪, 牛奶}(60%)}

核心步骤

候选3-项集：Ø

提纲

- 量子关联规则挖掘算法
  - ❑关联规则挖掘
  - ❑量子算法
  - ❑复杂度分析
  - ❑总结与展望

量子算法：基本思路

- 从所有候选 *k*-项集中找出频繁*k*-项集的过程包括两个子过程
  - (1) 计算每个候选*k*-项集的支持度
  - (2) 找出支持度大于阈值的候选*k*-项集，即频繁*k*-项集

那么如何设计量子算法快速的实现这两个子过程？

量子算法：基本思路

- 从所有候选 k-项集中找出频繁k-项集的过程包括两个子过程
  - (1) 计算每个候选k-项集的支持度
  - (2) 找出支持度大于阈值的候选k-项集，即频繁k-项集
- 直观思路
  - 在子过程(1)中，对每个候选 k-项集，用量子幅度估计（量子计数）估计出支持度

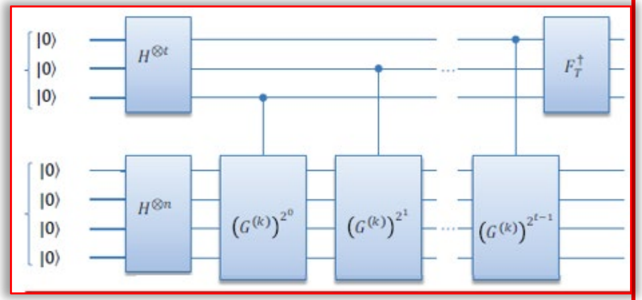
PE两要素：	初态$\|\psi\rangle$：交易记录序号均匀叠加态	G算子：标记解&关于初态翻转

  - 然后执行上述子过程(2)找出频繁k-项集

量子幅度估计：$|\psi\rangle = \sin(\theta)\,|\psi_1\rangle + \cos(\theta)\,|\psi_0\rangle$，估计$\sin^2(\theta) = \frac{M}{N}$($N$是搜索空间大小，$M$是目标解个数)

$G = (2|\psi\rangle\langle\psi| - I)(2|\psi_1\rangle\langle\psi_1| - I)$有特征值$\lambda_\pm = e^{\pm 2i\theta}$和

特征向量$|\psi_\pm\rangle = \frac{|\psi_0\rangle \mp i|\psi_1\rangle}{\sqrt{2}}$，且 $|\psi\rangle = \frac{e^{i\theta}|\psi_+\rangle + e^{-i\theta}|\psi_-\rangle}{\sqrt{2}}$



$$|0\rangle|\psi\rangle \xrightarrow{\text{相位估计: U}\rightarrow\text{G, } \sum_u c_u|u\rangle \rightarrow \frac{e^{i\theta}|\psi_+\rangle + e^{-i\theta}|\psi_-\rangle}{\sqrt{2}}} \frac{e^{i\theta}|2\tilde{\theta}\rangle|\psi_+\rangle + e^{-i\theta}|2\pi - 2\tilde{\theta}\rangle|\psi_-\rangle}{\sqrt{2}},$$

测量，可得到 $2\tilde{\theta}$ 或 $2\pi - 2\tilde{\theta}$，因此可得$\frac{M}{N} = \sin^2(\theta) = \sin^2(\pi - \theta)$

量子算法：基本思路

- 从所有候选 k-项集中找出频繁k-项集的过程包括两个子过程
  - (1) 计算每个候选k-项集的支持度
  - (2) 找出支持度大于阈值的候选k-项集，即频繁k-项集
- 直观思路

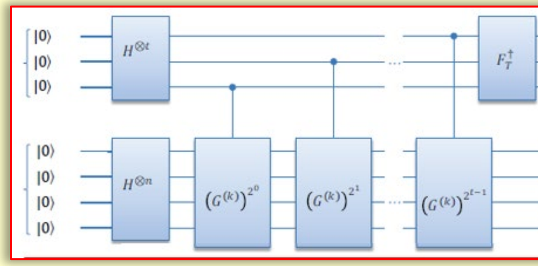
Trivial，只对(1)有加速

  - 在子过程(1)中，对每个候选 k-项集，用量子幅度估计（量子计数）估计出支持度
  - 然后执行上述子过程(2)找出频繁k-项集

量子幅度估计：$|\psi\rangle = \sin(\theta)\,|\psi_1\rangle + \cos(\theta)\,|\psi_0\rangle$，估计$\sin^2(\theta) = \frac{M}{N}$

$G = (2|\psi\rangle\langle\psi| - I)(2|\psi_1\rangle\langle\psi_1| - I)$有特征值$\lambda_{\pm}=e^{\pm 2i\theta}$和

特征向量$|\psi_{\pm}\rangle=\frac{|\psi_0\rangle \mp i|\psi_1\rangle}{\sqrt{2}}$，且$|\psi\rangle = \frac{e^{i\theta}|\psi_+\rangle + e^{-i\theta}|\psi_-\rangle}{\sqrt{2}}$



$$|0\rangle|\psi\rangle \xrightarrow{\text{相位估计}: \mathrm{U}\to\mathrm{G},\ \sum_u c_u|u\rangle \to \frac{e^{i\theta}|\psi_+\rangle + e^{-i\theta}|\psi_-\rangle}{\sqrt{2}}} \frac{e^{i\theta}|2\tilde{\theta}\rangle|\psi_+\rangle + e^{-i\theta}|2\pi - 2\tilde{\theta}\rangle|\psi_-\rangle}{\sqrt{2}}$$，测量R1

- 在子过程(1)中，对每个候选 k-项集，用量子幅度估计（量子计数）估计出支持度
- 然后执行上述子过程(2)找出频繁k-项集

➢ 能否再提速？

- 子过程(1)：能否对所有候选k-项集（用叠加态）并行计算出支持度？

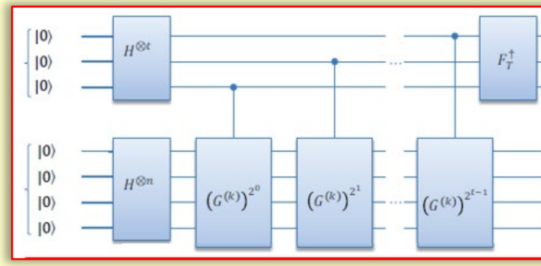
目标：产生存储所有候选k-项集及其支持度的量子态 $\sum_{j=1}^{M_c^{(k)}} |s_j^{(k)}\rangle |C_j^{(k)}\rangle$

这里$|s_j^{(k)}\rangle$就对应于上面$|2\tilde{\theta}\rangle$和$|2\pi - 2\tilde{\theta}\rangle$

量子幅度估计：$|\psi\rangle = \sin(\theta)|\psi_1\rangle + \cos(\theta)|\psi_0\rangle$，估计$\sin^2(\theta) = \frac{M}{N}$

$G = (2|\psi\rangle\langle\psi| - I)(2|\psi_1\rangle\langle\psi_1| - I)$有特征值$\lambda_\pm = e^{\pm 2i\theta}$和

特征向量$|\psi_\pm\rangle = \frac{|\psi_0\rangle \mp i|\psi_1\rangle}{\sqrt{2}}$，且$|\psi\rangle = \frac{e^{i\theta}|\psi_+\rangle + e^{-i\theta}|\psi_-\rangle}{\sqrt{2}}$



$|0\rangle|\psi\rangle \xrightarrow{\text{相位估计: U}\rightarrow\text{G}, \sum_u c_u|u\rangle \rightarrow \frac{e^{i\theta}|\psi_+\rangle + e^{-i\theta}|\psi_-\rangle}{\sqrt{2}}} \frac{e^{i\theta}|2\tilde{\theta}\rangle|\psi_+\rangle + e^{-i\theta}|2\pi - 2\tilde{\theta}\rangle|\psi_-\rangle}{\sqrt{2}}$，测量R1

- 在子过程(1)中，对每个候选 k-项集，用量子幅度估计（量子计数）估计出支持度
- 然后执行上述子过程(2)找出频繁k-项集

➢ 能否再提速？

- 子过程(1)：能否对所有候选k-项集（用叠加态）并行计算出支持度？

目标：产生存储所有候选k-项集及其支持度的量子态 $\sum_{j=1}^{M_c^{(k)}} |s_j^{(k)}\rangle|C_j^{(k)}\rangle$

这里$|s_j^{(k)}\rangle$就对应于上面$|2\tilde{\theta}\rangle$和$|2\pi - 2\tilde{\theta}\rangle$

- 子过程(2)：能否用幅度放大出找出频繁k-项集？

找出$s_j^{(k)}$大于阈值的$C_j^{(k)}$，$\sum_{s_j^{(k)} \geq s_m} |s_j^{(k)}\rangle|C_j^{(k)}\rangle$，测得频繁$k$-项集及其支持度

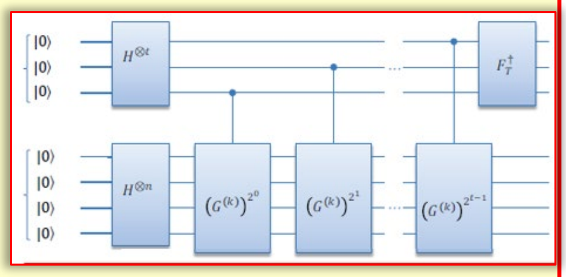
不难做到

即放大$|2\tilde{\theta}\rangle$和$|2\pi - 2\tilde{\theta}\rangle$满足$\sin^2(\theta) = \sin^2(\pi - \theta) = \frac{M}{N}$大于一定阈值的那些项

量子幅度估计：$|\psi\rangle = \sin(\theta)\,|\psi_1\rangle + \cos(\theta)\,|\psi_0\rangle$，估计$\sin^2(\theta) = \frac{M}{N}$

$G = (2|\psi\rangle\langle\psi| - I)(2|\psi_1\rangle\langle\psi_1| - I)$有特征值$\lambda_{\pm}=e^{\pm 2i\theta}$和

特征向量$|\psi_{\pm}\rangle=\frac{|\psi_0\rangle\mp i|\psi_1\rangle}{\sqrt{2}}$，且$|\psi\rangle = \frac{e^{i\theta}|\psi_+\rangle+e^{-i\theta}|\psi_-\rangle}{\sqrt{2}}$



$$|0\rangle|\psi\rangle \xrightarrow{\text{相位估计: U}\to\text{G, }\sum_u c_u|u\rangle\to\frac{e^{i\theta}|\psi_+\rangle+e^{-i\theta}|\psi_-\rangle}{\sqrt{2}}} \frac{e^{i\theta}|2\tilde{\theta}\rangle|\psi_+\rangle+e^{-i\theta}|2\pi-2\tilde{\theta}\rangle|\psi_-\rangle}{\sqrt{2}}$$，测量R1

- 在子过程(1)中，对每个候选 k-项集，用量子幅度估计（量子计数）估计出支持度
- 然后执行上述子过程(2)找出频繁k-项集

➢ 能否再提速？

- 子过程(1)：能否对所有候选k-项集（用叠加态）并行计算出支持度？

目标：产生存储所有候选k-项集及其支持度的量子态 $\sum_{j=1}^{M_c^{(k)}} |s_j^{(k)}\rangle|C_j^{(k)}\rangle$

这里$|s_j^{(k)}\rangle$就对应于上面$|2\tilde{\theta}\rangle$和$|2\pi-2\tilde{\theta}\rangle$

- 子过程(2)：能否用幅度放大出找出频繁k-项集？

找出$s_j^{(k)}$大于阈值的$C_j^{(k)}$，$\sum_{s_j^{(k)}\geq s_m} |s_j^{(k)}\rangle|C_j^{(k)}\rangle$，测得频繁$k$-项集及其支持度 不难做到

加速: $M_c^{(k)} \mapsto O(\sqrt{M_c^{(k)}M_f^{(k)}})$；当$M_f^{(k)} \ll M_c^{(k)}$时关于$M_c^{(k)}$平方加速

注：需找到所有解，复杂度 $O(\sqrt{N/M}) \mapsto O(\sqrt{NM})$

$s_j^{(k)}$：候选k-项集$C_j^{(k)}$的支持度；$M_c^{(k)}$：候选k-项集数；$M_f^{(k)}$：频繁k-项集数

量子算法：基本思路

- 从所有候选 k-项集中找出频繁k-项集的过程包括两个子过程
  - (1) 计算每个候选k-项集的支持度
  - (2) 找出支持度大于阈值的候选k-项集，即频繁k-项集
- 直观思路

Trivial，只对(1)有加速

  - 在子过程(1)中，对每个候选 k-项集，用量子幅度估计（量子计数）估计出支持度
  - 然后执行上述子过程(2)找出频繁k-项集
- 能否再提速？
  - 子过程(1)：能否对所有候选k-项集（用叠加态）并行计算出支持度？

目标：产生存储所有候选k-项集及其支持度的量子态 $\sum_{j=1}^{M_c^{(k)}} |s_j^{(k)}\rangle|C_j^{(k)}\rangle$

关键：能否做到？

  - 子过程(2)：能否用幅度放大出找出频繁k-项集？

找出$s_j^{(k)}$大于阈值的$C_j^{(k)}$，$\sum_{s_j^{(k)} \geq s_m} |s_j^{(k)}\rangle|C_j^{(k)}\rangle$，测得频繁$k$-项集及其支持度

不难做到

量子算法：基本思路

➢ 能否再提速？

❑ 子过程(1)：能否对所有候选k-项集（用叠加态）并行计算出支持度？

目标：产生存储所有候选k-项集及其支持度的量子态 $\sum_{j=1}^{M_c^{(k)}} |s_j^{(k)}\rangle|C_j^{(k)}\rangle$ 核心任务

多个候选 k 项集，并行做PE，是否可行？

$$\sum_{i=0}^{N-1}\sum_{j=1}^{M_c^{(k)}}(-1)^{f(x)}|i\rangle\left|C_j^{(k)}\right\rangle$$

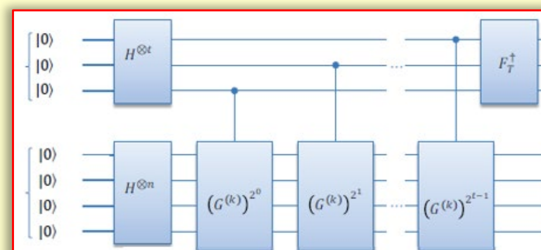
❑ 核心：构造 $G=(2|\psi\rangle\langle\psi|-I|)(2|\psi_1\rangle\langle\psi_1|-I)$

- 不同：每条记录是不是解，跟候选 k 项集有关；此时不管是G算子还是量子态，都带着$|C_j^{(k)}\rangle$这个尾巴（原本的幅度估计只有一个解标准，所以没有尾巴）
- 此时相位估计还能否实现？

量子幅度估计：$|\psi\rangle=\sin(\theta)|\psi_1\rangle+\cos(\theta)|\psi_0\rangle$，估计$\sin^2(\theta)=\frac{M}{N}$

$G=(2|\psi\rangle\langle\psi|-I)(2|\psi_1\rangle\langle\psi_1|-I)$有特征值$\lambda_\pm=e^{\pm 2i\theta}$和

特征向量$|\psi_\pm\rangle=\frac{|\psi_0\rangle\mp i|\psi_1\rangle}{\sqrt{2}}$，且 $|\psi\rangle=\frac{e^{i\theta}|\psi_+\rangle+e^{-i\theta}|\psi_-\rangle}{\sqrt{2}}$



$|0\rangle|\psi\rangle \xrightarrow{\text{相位估计}: \mathrm{U}\rightarrow\mathrm{G},\ \sum_u c_u|u\rangle\rightarrow\frac{e^{i\theta}|\psi_+\rangle+e^{-i\theta}|\psi_-\rangle}{\sqrt{2}}} \frac{e^{i\theta}|2\tilde{\theta}\rangle|\psi_+\rangle+e^{-i\theta}|2\pi-2\tilde{\theta}\rangle|\psi_-\rangle}{\sqrt{2}}$，测量R1

量子算法：构建量子黑盒

➢ 交易数据格式

交易	商品（项）
T_0	面包、奶酪、牛奶
T_1	面包、黄油
T_2	奶酪、牛奶
T_3	面包、奶酪
T_4	奶酪、黄油、牛奶

$$\begin{array}{cc} & \begin{array}{cccc} \text{面包} & \text{奶酪} & \text{黄油} & \text{牛奶} \end{array} \\ D = & \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 \end{pmatrix} \end{array}$$

数据用二元矩阵表示

行代表交易条目（指标为 i ）

列代表货物项（指标为 j）

➢ 假设存在用来读数据的基础量子黑盒O

$$O|i\rangle|j\rangle|a\rangle = |i\rangle|j\rangle|a \oplus D_{ij}\rangle$$

量子算法：构建量子黑盒

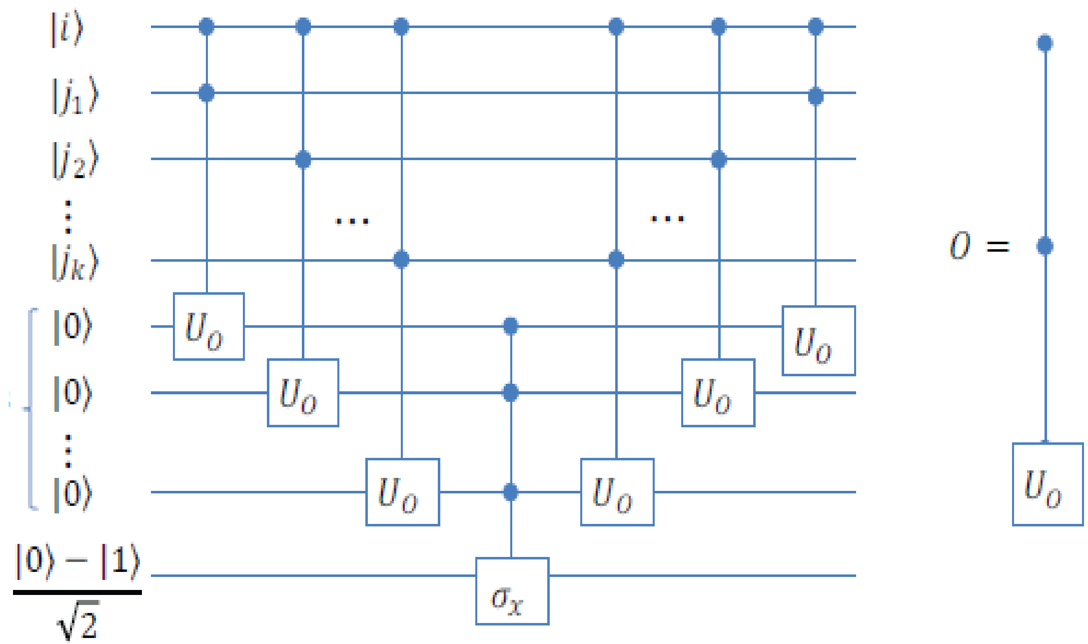
➢ 利用O构建判断T_i是否包含k-项集$X=\{I_{j_l}|l=1,2,\cdots,k\}$的量子黑盒$O^{(k)}$

$$O^{(k)}|i\rangle|j_1\rangle|j_2\rangle\cdots|j_k\rangle$$
$$=(-1)^{\tau(i,X)}|i\rangle|j_1\rangle|j_2\rangle\cdots|j_k\rangle$$

$j_1,j_2,\ldots,j_k$分别表示k-项集中每个商品的列号

$$\tau(i,X)=D_{ij_1}\cdots D_{ij_k}=\begin{cases}1, & X\subset T_i\\ 0, & X\not\subset T_i\end{cases}$$

即第 i 条交易包含这 k 个商品时 τ 为1，反之为0



实现$O^{(k)}$步骤

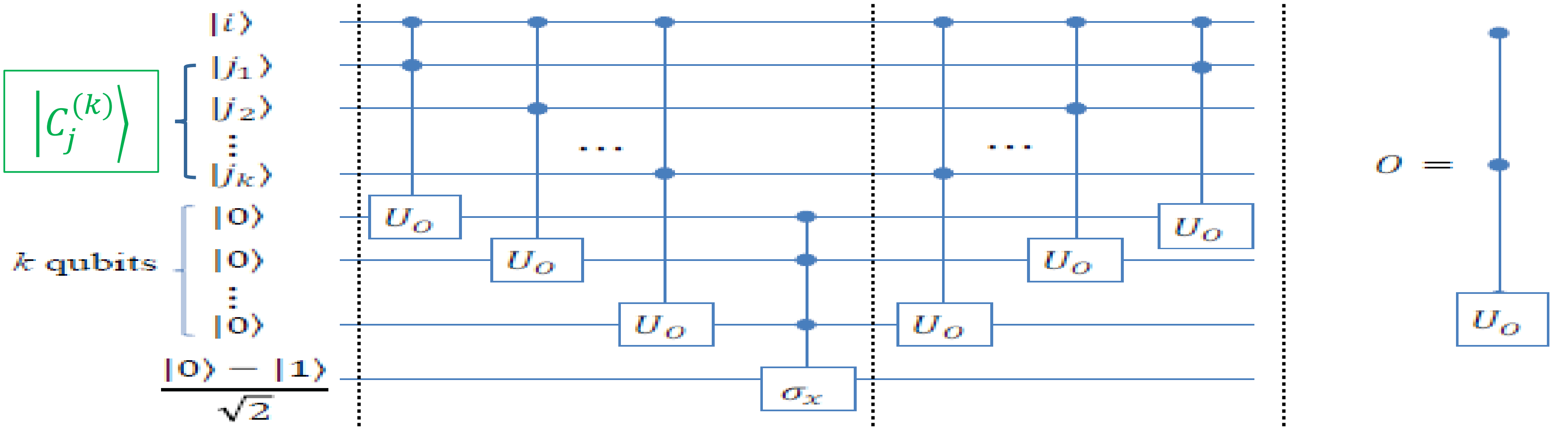
(1)（态制备）制备 $|i\rangle|j_1\rangle\cdots|j_k\rangle|0\rangle\cdots|0\rangle|-\rangle$

(2)（读数据）对$|i\rangle|j_l\rangle|0\rangle$执行$O(l=1\cdots k)$得到$|i\rangle|j_1\rangle\cdots|j_k\rangle\left|D_{ij_1}\right\rangle\cdots\left|D_{ij_k}\right\rangle|-\rangle$

(3)（全1则负）多比特受控非门得$(-1)^{D_{ij_1}\cdots D_{ij_k}}|i\rangle|j_1\rangle\cdots|j_k\rangle\left|D_{ij_1}\right\rangle\cdots\left|D_{ij_k}\right\rangle|-\rangle$

(4)（逆操作）重复(2)得 $(-1)^{D_{ij_1}\cdots D_{ij_k}}|i\rangle|j_1\rangle\cdots|j_k\rangle|0\rangle\cdots|0\rangle|-\rangle$

消耗 $2k$ 个基础黑盒O，以及$\Theta(k)$个基本门



基态

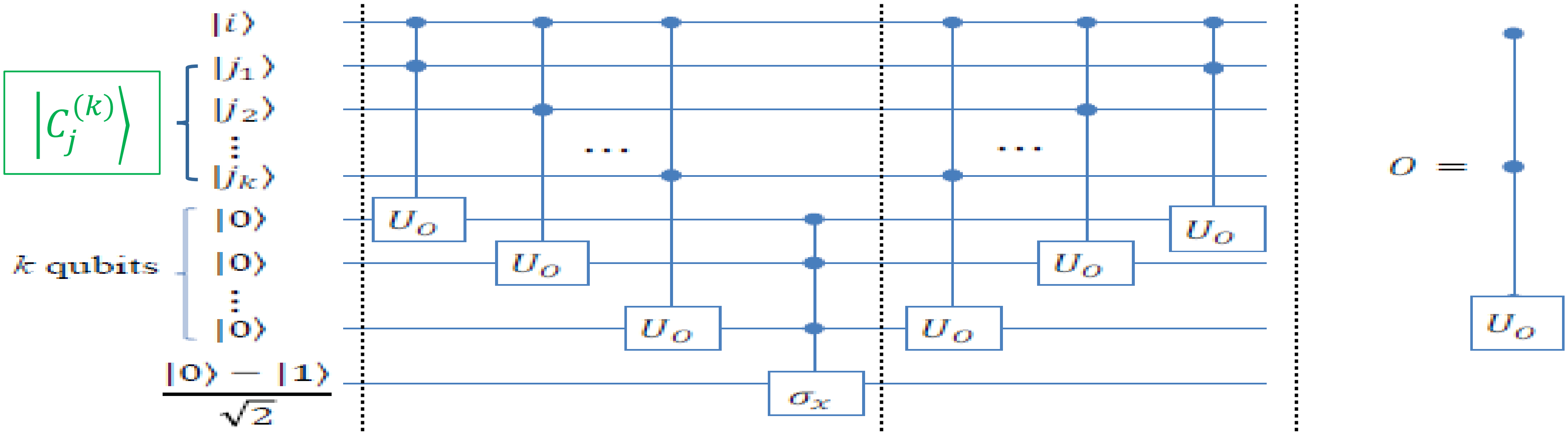
$|i\rangle\left|C_j^{(k)}\right\rangle|0\rangle\cdots|0\rangle|-\rangle$

$(-1)^{D_{ij_1}\cdots D_{ij_k}}|i\rangle\left|C_j^{(k)}\right\rangle\left|D_{ij_1}\right\rangle\cdots\left|D_{ij_k}\right\rangle|-\rangle$

$(-1)^{D_{ij_1}\cdots D_{ij_k}}|i\rangle\left|C_j^{(k)}\right\rangle|0\rangle\cdots|0\rangle|-\rangle$

R2,R3,R4全解纠缠出去，完成标记解操作

注：$|j_1\rangle\cdots|j_k\rangle$代表候选$k$-项集，所以这里态中换成了$\left|C_j^{(k)}\right\rangle$



基态

$|i\rangle \left|C_j^{(k)}\right\rangle |0\rangle \cdots |0\rangle |-\rangle$

$(-1)^{D_{ij_1}\cdots D_{ij_k}} |i\rangle \left|C_j^{(k)}\right\rangle \left|D_{ij_1}\right\rangle \cdots \left|D_{ij_k}\right\rangle |-\rangle$

$(-1)^{D_{ij_1}\cdots D_{ij_k}} |i\rangle \left|C_j^{(k)}\right\rangle |0\rangle \cdots |0\rangle |-\rangle$

R2,R3,R4全解纠缠出去，完成标记解操作

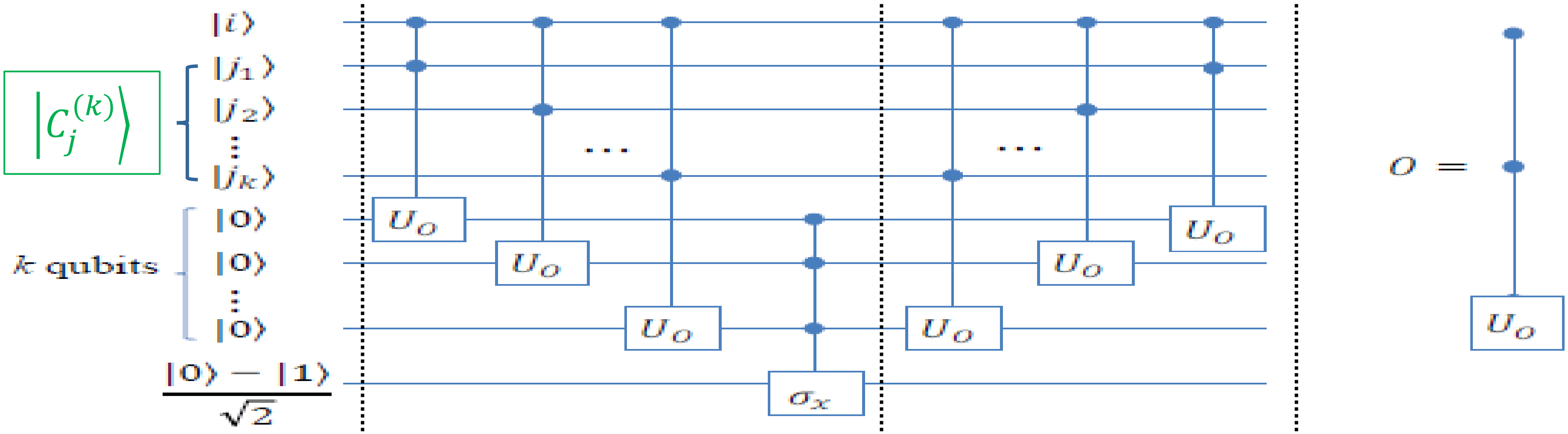
R1叠加态

$\sum_{i=0}^{N-1} |i\rangle \left|C_j^{(k)}\right\rangle |0\rangle \cdots |0\rangle |-\rangle$

$\sum_{i=0}^{N-1} (-1)^{D_{ij_1}\cdots D_{ij_k}} |i\rangle |j_1\rangle \cdots |j_k\rangle \left|D_{ij_1}\right\rangle \cdots \left|D_{ij_k}\right\rangle |-\rangle$

$\sum_{i=0}^{N-1} (-1)^{D_{ij_1}\cdots D_{ij_k}} |i\rangle \left|C_j^{(k)}\right\rangle |0\rangle \cdots |0\rangle |-\rangle$

R2,R3,R4全解纠缠出去，完成标记解操作



基态

$|i\rangle\left|C_j^{(k)}\right\rangle|0\rangle\cdots|0\rangle|-\rangle$

$(-1)^{D_{ij_1}\cdots D_{ij_k}}|i\rangle\left|C_j^{(k)}\right\rangle\left|D_{ij_1}\right\rangle\cdots\left|D_{ij_k}\right\rangle|-\rangle$

$(-1)^{D_{ij_1}\cdots D_{ij_k}}|i\rangle\left|C_j^{(k)}\right\rangle|0\rangle\cdots|0\rangle|-\rangle$

R2,R3,R4全解纠缠出去，完成标记解操作

R1叠加态

$\sum_{i=0}^{N-1}|i\rangle\left|C_j^{(k)}\right\rangle|0\rangle\cdots|0\rangle|-\rangle$

$\sum_{i=0}^{N-1}(-1)^{D_{ij_1}\cdots D_{ij_k}}|i\rangle|j_1\rangle\cdots|j_k\rangle\left|D_{ij_1}\right\rangle\cdots\left|D_{ij_k}\right\rangle|-\rangle$

$\sum_{i=0}^{N-1}(-1)^{D_{ij_1}\cdots D_{ij_k}}|i\rangle\left|C_j^{(k)}\right\rangle|0\rangle\cdots|0\rangle|-\rangle$

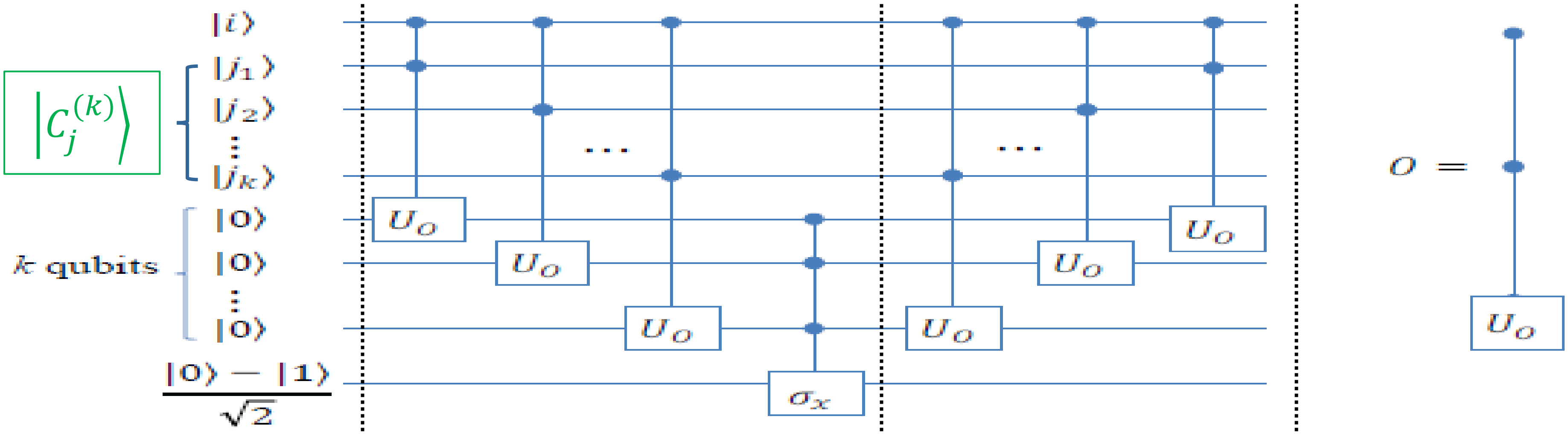
R2,R3,R4全解纠缠出去，完成标记解操作

R2叠加态

$\sum_{j=1}^{M_c^{(k)}}\left|C_j^{(k)}\right\rangle$

$C_j^{(k)}$是代表第j个候选k-项集，比如k=3，$C_1^{(k)}=\{1,2,4\}$，　$C_2^{(k)}=\{2,5,6\}$，则

$\left|C_1^{(k)}\right\rangle=|1\rangle|2\rangle|4\rangle$，　$\left|C_2^{(k)}\right\rangle=|2\rangle|5\rangle|6\rangle$，两者叠加即 $|1\rangle|2\rangle|4\rangle+|2\rangle|5\rangle|6\rangle$



基态

$$|i\rangle\left|C_j^{(k)}\right\rangle|0\rangle\cdots|0\rangle|-\rangle$$

$$(-1)^{D_{ij_1}\cdots D_{ij_k}}|i\rangle\left|C_j^{(k)}\right\rangle|0\rangle\cdots|0\rangle|-\rangle$$

$$(-1)^{D_{ij_1}\cdots D_{ij_k}}|i\rangle\left|C_j^{(k)}\right\rangle\left|D_{ij_1}\right\rangle\cdots\left|D_{ij_k}\right\rangle|-\rangle$$

R2,R3,R4全解纠缠出去，完成标记解操作

R1叠加态

$$\sum_{i=0}^{N-1}|i\rangle\left|C_j^{(k)}\right\rangle|0\rangle\cdots|0\rangle|-\rangle$$

$$\sum_{i=0}^{N-1}(-1)^{D_{ij_1}\cdots D_{ij_k}}|i\rangle\left|C_j^{(k)}\right\rangle|0\rangle\cdots|0\rangle|-\rangle$$

$$\sum_{i=0}^{N-1}(-1)^{D_{ij_1}\cdots D_{ij_k}}|i\rangle|j_1\rangle\cdots|j_k\rangle\left|D_{ij_1}\right\rangle\cdots\left|D_{ij_k}\right\rangle|-\rangle$$

R2,R3,R4全解纠缠出去，完成标记解操作

R2叠加态

$$\sum_{i=0}^{N-1}\sum_{j=1}^{M_c^{(k)}}|i\rangle\left|C_j^{(k)}\right\rangle|0\rangle\cdots|0\rangle|-\rangle$$

$$\sum_{i=0}^{N-1}\sum_{j=1}^{M_c^{(k)}}(-1)^{D_{ij_1}\cdots D_{ij_k}}|i\rangle\left|C_j^{(k)}\right\rangle|0\rangle\cdots|0\rangle|-\rangle$$

$$\sum_{i=0}^{N-1}\sum_{j=1}^{M_c^{(k)}}(-1)^{D_{ij_1}\cdots D_{ij_k}}|i\rangle\left|C_j^{(k)}\right\rangle\left|D_{ij_1}\right\rangle\cdots\left|D_{ij_k}\right\rangle|-\rangle$$

R3,R4解纠缠，完成标记解操作，但R2无法解纠缠

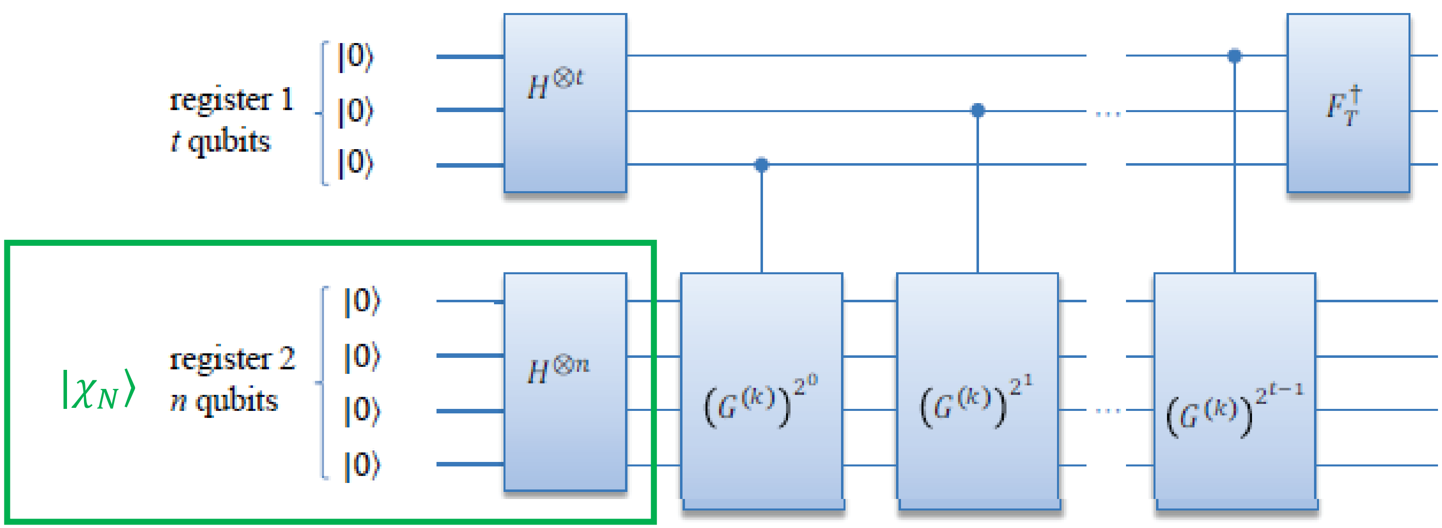
R2不叠加时，R2可解纠缠，相当于构造了标记解操作 $O_j^{(k)}|i\rangle = (-1)^{\tau\left(i,c_j^{(k)}\right)}|i\rangle$

相应的Grover算子为$G_j^{(k)} = (2|\chi_N\rangle\langle\chi_N|-I)O_j^{(k)}$， 这里$|\chi_N\rangle$为均匀叠加态

$G_j^{(k)}$有特征值$e^{\pm 2i\theta_j^{(k)}}$和特征向量$\left|\phi_{j\pm}^{(k)}\right\rangle$，$|\chi_N\rangle = \frac{\sum_{n=0}^{N-1}|n\rangle}{\sqrt{N}} = \frac{e^{i\theta_j^{(k)}}\left|\phi_{j+}^{(k)}\right\rangle - e^{-i\theta_j^{(k)}}\left|\phi_{j-}^{(k)}\right\rangle}{\sqrt{2}}$

$c_j^{(k)}$的支持度$s_j^{(k)} = \frac{M}{N} = \sin^2\left(\theta_j^{(k)}\right)$，以$|\chi_N\rangle$为初态对$G_j^{(k)}$做相位估计：

$$|0\rangle|\chi_N\rangle \xrightarrow{\text{相位估计}} e^{i\theta_j^{(k)}}|2\theta_j^{(k)}\rangle\left|\phi_{j+}^{(k)}\right\rangle - e^{-i\theta_j^{(k)}}\left|2\pi - 2\theta_j^{(k)}\right\rangle\left|\phi_{j-}^{(k)}\right\rangle$$



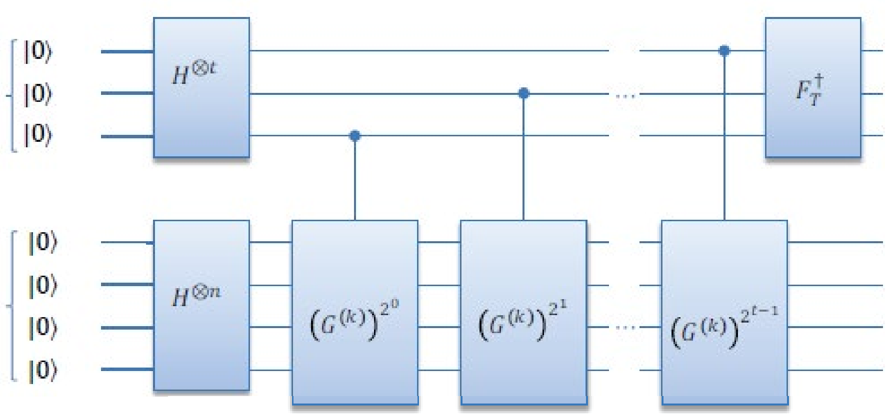
$$\sum_{i=0}^{N-1}\sum_{j=1}^{M_c^{(k)}}(-1)^{D_{ij_1}\cdots D_{ij_k}}|i\rangle\left|C_j^{(k)}\right\rangle|0\rangle\cdots|0\rangle|-\rangle$$

R2不叠加时

标记解 $O_j^{(k)}|i\rangle=(-1)^{\tau\left(i,C_j^{(k)}\right)}|i\rangle$

$$G_j^{(k)}=(2|\chi_N\rangle\langle\chi_N|-I)O_j^{(k)}$$

$$|0\rangle|\chi_N\rangle\xrightarrow{\text{相位估计}}e^{i\theta_j^{(k)}}\left|2\theta_j^{(k)}\right\rangle\left|\phi_{j+}^{(k)}\right\rangle-e^{-i\theta_j^{(k)}}\left|2\pi-2\theta_j^{(k)}\right\rangle\left|\phi_{j-}^{(k)}\right\rangle$$



R2叠加时

标记解 $O^{(k)}|i\rangle\left|C_j^{(k)}\right\rangle=(-1)^{\tau\left(i,C_j^{(k)}\right)}|i\rangle\left|C_j^{(k)}\right\rangle$

$$G^{(k)}=[(2|\chi_N\rangle\langle\chi_N|-I)\otimes I]O^{(k)}$$

这个$G^{(k)}$还能不能做相位估计？

北京邮电大学 Beijing University of Posts and Telecommunications

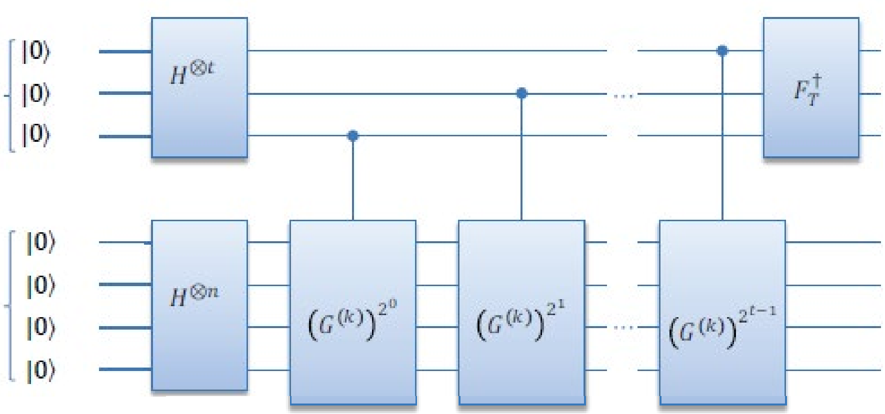
这意味着对任意$C_j^{(k)}$，均可用$G^{(k)}$执行量子幅度估计以估计其支持度

R2不叠加时

标记解 $O_j^{(k)}|i\rangle = (-1)^{\tau\left(i,C_j^{(k)}\right)}|i\rangle$

$$G_j^{(k)} = (2|\chi_N\rangle\langle\chi_N| - I)O_j^{(k)}$$

$$|0\rangle|\chi_N\rangle \xrightarrow{\text{相位估计}} e^{i\theta_j^{(k)}}\left|2\theta_j^{(k)}\right\rangle\left|\phi_{j+}^{(k)}\right\rangle - e^{-i\theta_j^{(k)}}\left|2\pi - 2\theta_j^{(k)}\right\rangle\left|\phi_{j-}^{(k)}\right\rangle$$



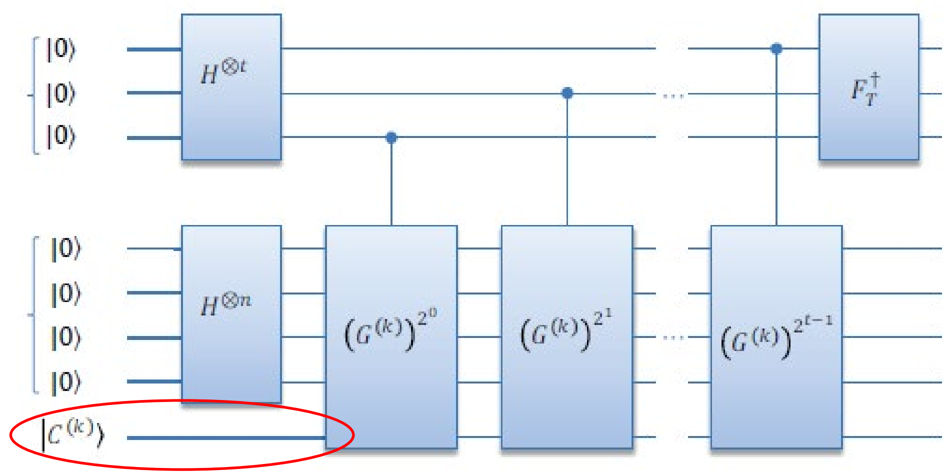
R2叠加时

标记解 $O^{(k)}|i\rangle\left|C_j^{(k)}\right\rangle = (-1)^{\tau\left(i,C_j^{(k)}\right)}|i\rangle\left|C_j^{(k)}\right\rangle$

$$G^{(k)} = [(2|\chi_N\rangle\langle\chi_N| - I) \otimes I]O^{(k)}$$

这个$G^{(k)}$还能不能做相位估计？ 可以！

对于任意一个候选k-项集$C_j^{(k)}$，$G^{(k)}$和$G_j^{(k)}$功能一样：$G^{(k)}(|\chi_N\rangle|C_j^{(k)}\rangle) = (G_j^{(k)}|\chi_N\rangle)|C_j^{(k)}\rangle$

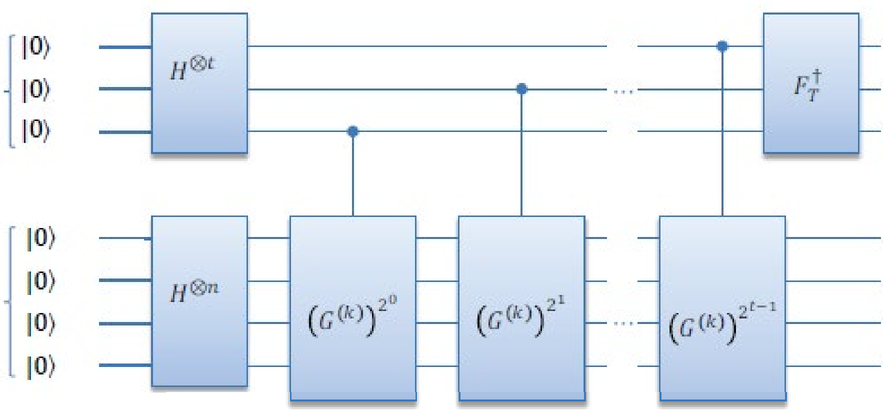


R2不叠加时

标记解 $O_j^{(k)}|i\rangle = (-1)^{\tau\left(i,C_j^{(k)}\right)}|i\rangle$

$$G_j^{(k)} = (2|\chi_N\rangle\langle\chi_N|-I)O_j^{(k)}$$

$$|0\rangle|\chi_N\rangle \xrightarrow{\text{相位估计}} e^{i\theta_j^{(k)}}\left|2\theta_j^{(k)}\right\rangle\left|\phi_{j+}^{(k)}\right\rangle - e^{-i\theta_j^{(k)}}\left|2\pi-2\theta_j^{(k)}\right\rangle\left|\phi_{j-}^{(k)}\right\rangle$$

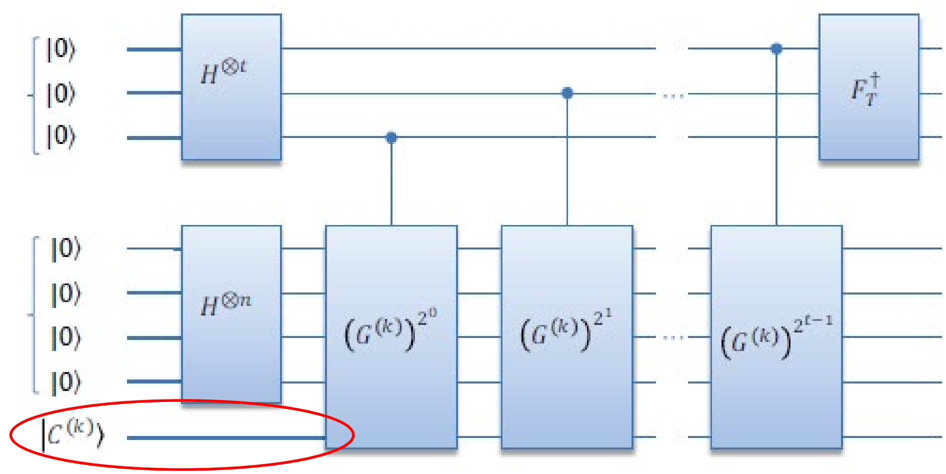


R2叠加时

标记解 $O^{(k)}|i\rangle\left|C_j^{(k)}\right\rangle = (-1)^{\tau\left(i,C_j^{(k)}\right)}|i\rangle\left|C_j^{(k)}\right\rangle$

$$G^{(k)} = [(2|\chi_N\rangle\langle\chi_N|-I)\otimes I]O^{(k)}$$

$$|0\rangle|\chi_N\rangle \xrightarrow{\text{相位估计}} \sum_{j=1}^{M_c^{(k)}}\left[e^{i\theta_j^{(k)}}|2\theta_j^{(k)}\rangle\left|\phi_{j+}^{(k)}\right\rangle - e^{-i\theta_j^{(k)}}\left|2\pi-2\theta_j^{(k)}\right\rangle\left|\phi_{j-}^{(k)}\right\rangle\right]\left|C_j^{(k)}\right\rangle$$

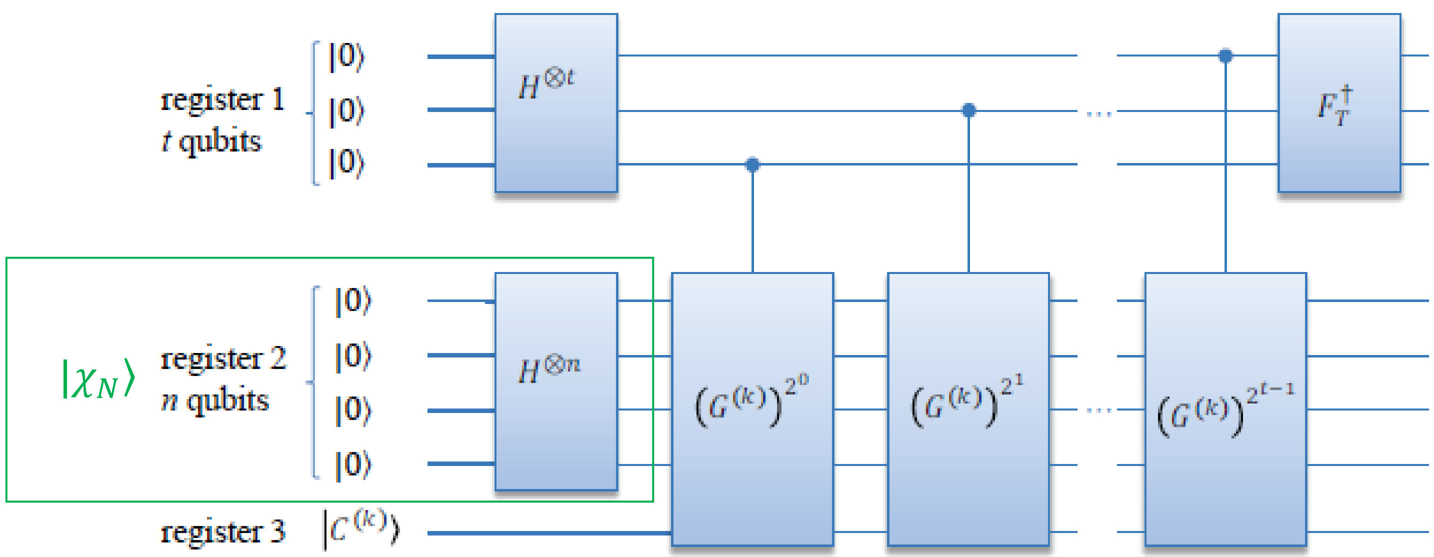


量子算法：步骤

(1) 制备$\left(\frac{\sum_{t=0}^{T-1}|t\rangle}{\sqrt{T}}\right)|\chi_N\rangle\left(\sum_{j=1}^{M_c^{(k)}}\left|C_j^{(k)}\right\rangle/\sqrt{M_c^{(k)}}\right)$，其中$M_c^{(k)}$为候选$k$-项集数目

(2) 利用构建的$G^{(k)}$执行量子相位估计：$\sum_{t=0}^{T-1}|t\rangle\langle t|\otimes\left(G^{(k)}\right)^t$、逆QFT，得

$$\sum_{j=1}^{M_c^{(k)}}\left[e^{i\theta_j^{(k)}}|2\theta_j^{(k)}\rangle\left|\phi_{j+}^{(k)}\right\rangle-e^{-i\theta_j^{(k)}}\left|2\pi-2\theta_j^{(k)}\right\rangle\left|\phi_{j-}^{(k)}\right\rangle\right]\left|C_j^{(k)}\right\rangle/\sqrt{2M_c^{(k)}}$$



$G^{(k)}$已知，易实现受控$G^{(k)}$，复杂度不变

量子算法：步骤

(1) 制备$\left(\frac{\sum_{t=0}^{T-1}|t\rangle}{\sqrt{T}}\right)|\chi_N\rangle\left(\sum_{j=1}^{M_c^{(k)}}\left|C_j^{(k)}\right\rangle/\sqrt{M_c^{(k)}}\right)$，其中$M_c^{(k)}$为候选$k$-项集数目

(2) 利用构建的$G^{(k)}$执行量子相位估计：$\sum_{t=0}^{T-1}|t\rangle\langle t|\otimes\left(G^{(k)}\right)^t$、逆QFT，得

$$\sum_{j=1}^{M_c^{(k)}}\left[e^{i\theta_j^{(k)}}|2\theta_j^{(k)}\rangle\left|\phi_{j+}^{(k)}\right\rangle-e^{-i\theta_j^{(k)}}\left|2\pi-2\theta_j^{(k)}\right\rangle\left|\phi_{j-}^{(k)}\right\rangle\right]\left|C_j^{(k)}\right\rangle/\sqrt{2M_c^{(k)}}$$

(3) 利用幅度放大搜索$s_j^{(k)}=\sin^2\left(\theta_j^{(k)}\right)=\sin^2\left(\pi-\theta_j^{(k)}\right)\geq s_m$ (阈值)的项

$$\sum_{s_j^{(k)}\geq s_m}\left[e^{i\theta_j^{(k)}}|2\theta_j^{(k)}\rangle\left|\phi_{j+}^{(k)}\right\rangle-e^{-i\theta_j^{(k)}}\left|2\pi-2\theta_j^{(k)}\right\rangle\left|\phi_{j-}^{(k)}\right\rangle\right]\left|C_j^{(k)}\right\rangle/\sqrt{2M_f^{(k)}}$$

其中$M_f^{(k)}$为频繁k-项集数目

(4) 测量第1、3寄存器获得频繁k-项集及其支持度（每次测得1个）

提纲

- 量子关联规则挖掘算法
  - 关联规则挖掘
  - 量子算法
  - 复杂度分析
  - 总结与展望

复杂度分析

(1) 制备$\left(\frac{\sum_{t=0}^{T-1}|t\rangle}{\sqrt{T}}\right)|\chi_N\rangle\left(\sum_{j=1}^{M_c^{(k)}}\left|C_j^{(k)}\right\rangle/\sqrt{M_c^{(k)}}\right)$，其中$M_c^{(k)}$为候选$k$-项集数目

- 门复杂度为$O\left(\log T+\log N+k\log M_c^{(k)}\right)$：$\sum_{j=1}^{M_c^{(k)}}\left|C_j^{(k)}\right\rangle/\sqrt{M_c^{(k)}}$可用QRAM[1]产生的$\sum_{j=1}^{M_c^{(k)}}|j\rangle\left|C_j^{(k)}\right\rangle/\sqrt{M_c^{(k)}}$代替，该过程复杂度为$O\left(k\log M_c^{(k)}\right)$

QRAM：$\sum_i|i\rangle|0\rangle\rightarrow\sum_i|i\rangle|b_i\rangle$

(2) 利用构建的$G^{(k)}$执行量子相位估计：$\sum_{t=0}^{T-1}|t\rangle\langle t|\otimes\left(G^{(k)}\right)^t$、逆QFT

- 查询复杂度为$O(Tk)$：执行了$T-1$个受控$G^{(k)}=[(2|\chi_N\rangle\langle\chi_N|-I)\otimes I]O^{(k)}$，每个包含1个$O^{(k)}$，而1个$O^{(k)}$需要$2k$个基础黑盒$O$
- 门复杂度为$O(T\log N)$：$G^{(k)}=[(2|\chi_N\rangle\langle\chi_N|-I)\otimes I]O^{(k)}$中需要$O(\log N)$门复杂度实现$(2|\chi_N\rangle\langle\chi_N|-I)$
- 误差：若$\theta_j^{(k)}$估计误差为ϵ，则$s_j^{(k)}$的估计误差[2]为$O\left[\epsilon\sqrt{s_j^{(k)}\left(1-s_j^{(k)}\right)}\right]$

[1] Quantum Random Access Memory, V.Giovannetti, S. Lloyd, and L. Maccone, Phys. Rev. Lett. 100, 160501 (2008). [2] G. Brassard, P. Høyer, M. Mosca, and A. Tapp, *Quantum Amplitude Amplification and Estimation*, Contemporary Mathematics Series, Millenium Vol. 305 (AMS, New York, 2002).

复杂度分析

(3) 利用幅度放大搜索$s_j^{(k)} = \sin^2\left(\theta_j^{(k)}\right) = \sin^2\left(\pi - \theta_j^{(k)}\right) \geq s_m$ (阈值)的项

$$\sum_{s_j^{(k)} \geq s_m} \left[e^{i\theta_j^{(k)}} |2\theta_j^{(k)}\rangle \left|\phi_{j+}^{(k)}\right\rangle - e^{-i\theta_j^{(k)}} \left|2\pi - 2\theta_j^{(k)}\right\rangle \left|\phi_{j-}^{(k)}\right\rangle \right] \left|C_j^{(k)}\right\rangle / \sqrt{2M_f^{(k)}}$$

- 迭代次数：$O\left(\sqrt{M_c^{(k)}/M_f^{(k)}}\right)$

(4) 测量第1、3寄存器获得频繁k-项集及其支持度（每次测得1个）

- 测量次数：$O\left(M_f^{(k)}\right)$

$T = 2^t = O(1/\epsilon)$

总体复杂度：

(基础黑盒O的)查询复杂度：$O\left(\frac{k\sqrt{M_c^{(k)} M_f^{(k)}}}{\epsilon}\right)$

额外门复杂度：$O\left[\sqrt{M_c^{(k)} M_f^{(k)}} \left(\log\frac{1}{\epsilon} + \frac{\log N}{\epsilon} + k \log M_c^{(k)}\right)\right]$

与经典算法相比较

- 量子算法：查询复杂度为 $O\left(\frac{k\sqrt{M_c^{(k)}M_f^{(k)}}}{\epsilon}\right)$

- 经典基于抽样的Apriori算法：对每个候选k-项集$C_j^{(k)}$的支持度$s_j^{(k)}$执行采样估计。根据二项分布特点，采样$O\left(\frac{1}{\epsilon^2}\right)$次可以使得$s_j^{(k)}$估计误差为$O\left[\epsilon\sqrt{s_j^{(k)}\left(1-s_j^{(k)}\right)}\right]$（为了便于比较，该误差取与量子算法相同的值），每次采样需要执行k次基础黑盒O判断交易是否包含$C_j^{(k)}$。因此基于采样算法总查询复杂度为 $O\left(\frac{kM_c^{(k)}}{\epsilon^2}\right)$
  - 当$M_f^{(k)} \approx M_c^{(k)}$, 关于$\epsilon$具有平方加速
  - 当$M_f^{(k)} \ll M_c^{(k)}$, 关于$M_c^{(k)}$和ϵ均具有平方加速

$M_c^{(k)}$：候选k-项集数目
$M_f^{(k)}$：频繁k-项集数目
ϵ：项集支持度估计误差

- 确定的Apriori算法（前面两个都是非确定的）：直接计算每个$s_j^{(k)}$，扫描整个数据库每个交易，对每个交易执行k次基础黑盒判断该交易是否包含每个候选k-项集$C_j^{(k)}$，总的查询复杂度为$O\left(kNM_c^{(k)}\right)$

N：交易记录总数

提纲

- 量子关联规则挖掘算法
  - ❑关联规则挖掘
  - ❑量子算法
  - ❑复杂度分析
  - ❑总结与展望

总结与展望

➢ 总结

- ❑ 提出了一个量子关联规则挖掘算法，与经典算法相比至少关于支持度估计误差具有平方加速
- ❑ 量子搜索、幅度放大、幅度估计中，带条件“尾巴”也可以做
- ❑ 量子并行幅度估计：可以（对不同的解条件）并行估计

➢ 展望

- ❑ 该量子算法基于经典著名关键规则挖掘算法—Apriori算法，而基于其他经典关联规则挖掘算法的量子算法未来值得研究
- ❑ 量子关联规则中的隐私保护问题[1]

[1] S. Ying, M. Ying, Y. Feng, Quantum Privacy-Preserving Data Analytics, arXiv:1702.04420

Email: gaof@bupt.edu.cn
Tel: 86 -10 -62283192
Web: www.bupt.edu.cn

谢 谢！