量子近似优化算法中的最优初始化参数

队名: Quiscus

问题描述

- 给定带权高阶 Ising 模型，求其对应的标准 p-层 QAOA 线路最优~~初始化~~参数
  - 建模形式：$F(graph, p) \rightarrow [\gamma, \beta]$，可理解为回归问题
  - 不允许显式迭代 QAOA$(graph, p)$，可理解为一步蒸馏近似问题

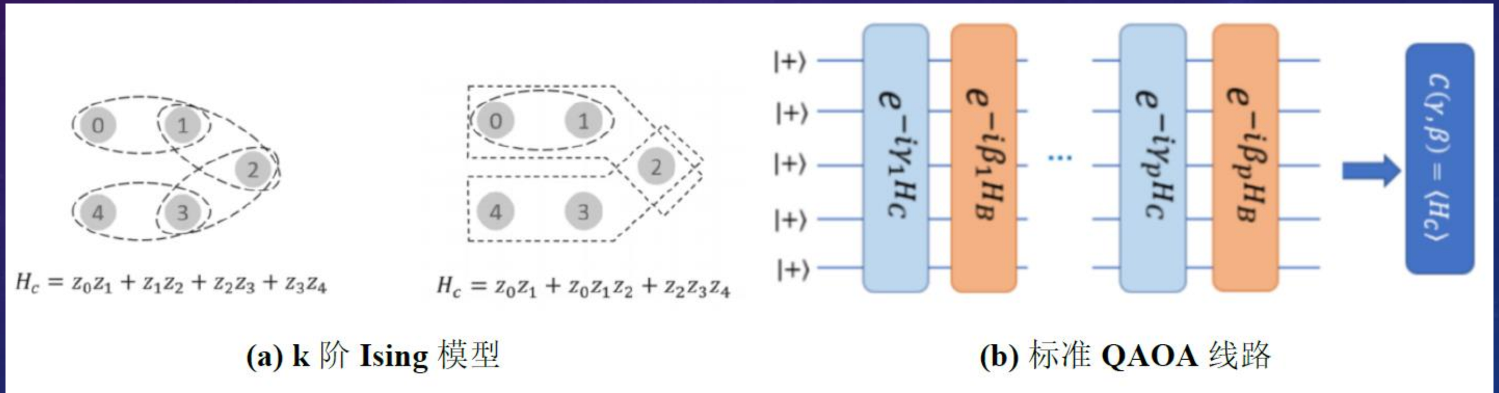
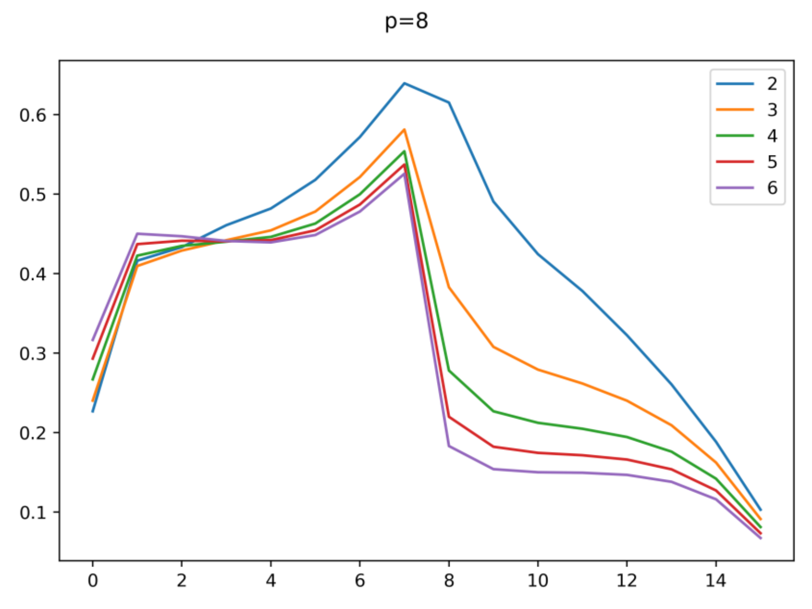


图 1　**Ising** 模型及其对应的 **QAOA** 线路

解决方案

- 直接建模
  - $F(graph, p) \to [\gamma, \beta]$
- 参数迁移
  - $G(graph) \to [D, k, \dots]$ // 图 graph 的度数 D 和阶数 k
  - $H(D, k, \dots, p, \gamma^*, \beta^*) \to [\gamma, \beta]$ // 参考样例的最优参数 γ^* 和 β^*

  - 参考样例从何而来？预制表！
  - 原理近似于kNN回归



arXiv:2110.14206 给出的最优参数预制表

重缩放因子

- 文献 arXiv:2201.11785 和 arXiv:2305.15201 中的 γ 迁移公式
  - 仅分母的距离度量不同

$$\gamma^* = \frac{\gamma^{\text{inf}}}{\frac{1}{|E|}\sum_{\{u,v\}\in E}|w_{uv}|}\arctan\left(\frac{1}{\sqrt{D-1}}\right)$$

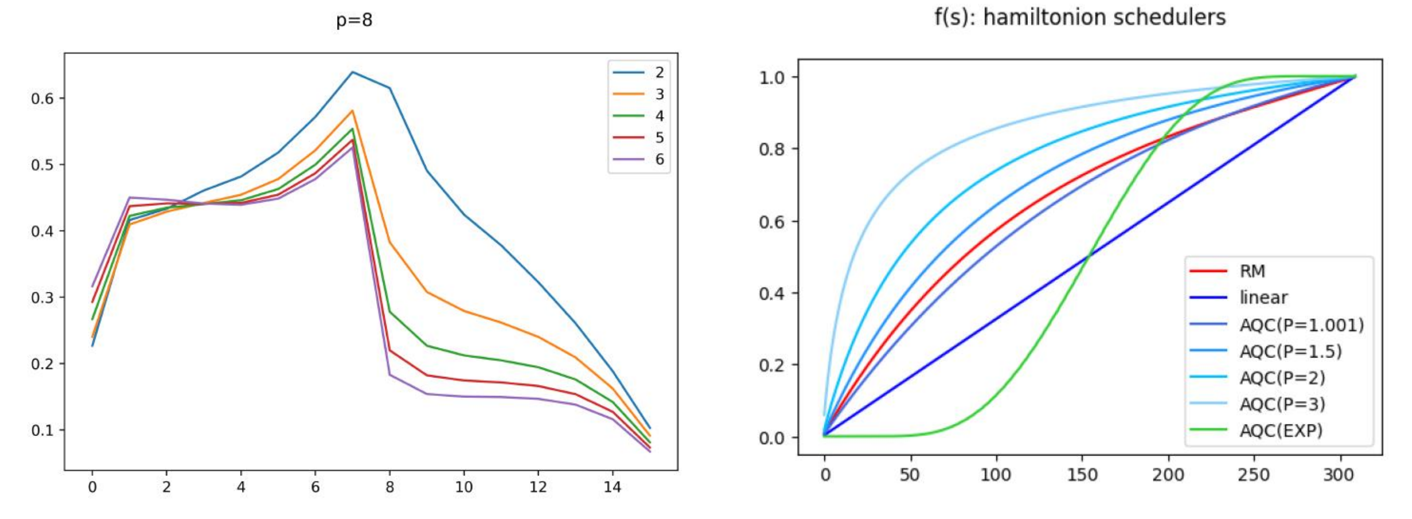
$$\gamma^* = \frac{\gamma^{\text{median}}}{\sqrt{\frac{1}{|E|}\sum_{\{u,v\}\in E} w_{uv}^2}}\arctan\left(\frac{1}{\sqrt{D-1}}\right)$$

- 我们使用的 γ 迁移公式
  - 引入rescaler

$$\gamma^* = \text{rescaler} * \frac{\gamma^{\text{inf}}}{\sqrt{\frac{1}{|E|}\sum_{\{u,v\}\in E} w_{uv}^2}}\arctan\left(\frac{1}{\sqrt{D-1}}\right)$$

这个对...对吗?

- 绝热演化参数 β 相关于初态哈密顿量，γ 相关于末态哈密顿量
  - 经验设计上 γ 应满足函数单增，且导函数单减的性质



(a) 预制表参数 **(p=8)**　　**(b)** 哈密顿量调度函数

图 **2**　预制表参数与哈密顿量调度函数

不对，所以多策略微调预制表

- 朴素的更新策略
- 改进的更新策略
  - 学习率衰减：保证学习呈现收敛趋势
  - 自适应损失：动态样例权重
  - 动量：变相加大epoch
  - 均权-加权分离：观察到uniform情况下的衰退

$$\theta_{i+1}^* = (1 - \Delta x) * \theta_i^* + \Delta x * \hat{\theta}_i$$

$$\Delta E = E_{\text{before}} - E_{\text{after}}$$

$$\tilde{\Delta x} = \Delta x * \left(w^{\frac{\text{cur\_iter}}{\text{decay\_per\_iter}}}\right) * \log(1 + \Delta E)$$

$$\mu_{i+1}^* = (1 - m) * \mu_i^* + m * \hat{\theta}_i$$

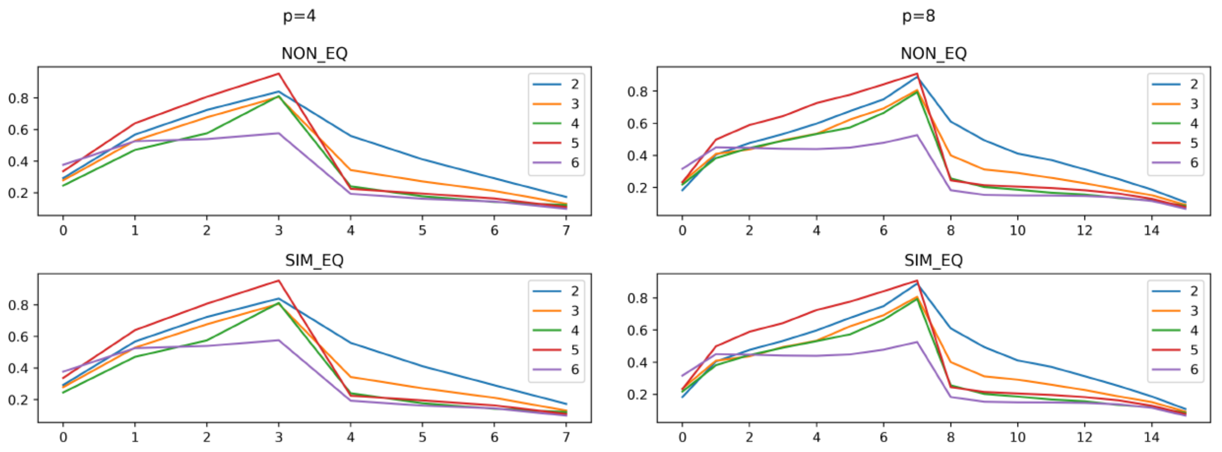
$$\theta_{i+1}^* = (1 - \tilde{\Delta x}) * \theta_i^* + \tilde{\Delta x} * \mu_{i+1}^*$$

```
### ft-ada-decay_T=9400
# [prop] edge density (0, -7, -35 of each; 边越密集增益越大)
#   avg: [-127.6368, -164.26228, -158.56534]
#   std: [73.07497, 92.10416, 103.769066]
# [k] order of ising-model (0, -2, -31, -7 of each; 阶数越高增益越大)
#   avg: [-68.5941, -127.13828, -169.03491, -235.85196]
#   std: [30.184612, 50.04047, 77.59191, 97.498344]
# [coef] edge weight probdist (-22, +2, -9 of each; 分布为uni的样例有退化, TODO: 可能要按不同分布分别建模)
#   avg: [-111.96218, -112.613014, -22[illegible].8892]
#   std: [53.582024, 56.71622, 103.32686]
# [r] sample index (-10 on avg)
#   avg: [-149.15585, -150.24004, -150.1911, -150.52188, -152.05824, -148.99612, -148.21541, -150.36801, -150.84392, -150.95749]
#   std: [91.202065, 94.491295, 91.501434, 90.81646, 90.03252, 91.7499, 92.028015, 93.54402, 92.99735, 91.050766]
# [p] circuit depth (-7, -12 for each)
#   avg: [-133.74776, -166.56184]
#   std: [79.67872, 100.124405]
```

实验结果

表 1　实验结果

#	local score	submit score	comment
#1	11730.14583	16627.4960	PT-WMC [2] , reference
#2	16526.79871	24999.9025	PS-WP [3] , baseline
#3	17816.62534	27217.3001	baseline; rescaler=1.275
#4	18516.17254	-	opt-avg
#5	18535.15152	27998.4487	ft; rescaler=1.275
#6	20489.27983	29920.0329	ft-ada; rescaler=1.275
#7	20948.34952	30332.7544	ft-ada-decay; rescaler=1.275
#8	20381.14635	29800.8173	ft-ada-decay-moment-fast; rescaler=1.275
#9	21033.52542	30264.7276	ft-ada-decay-moment-fast_ft; rescaler=1.165
#10	21432.33415	**30641.0100**	ft-ada-moment-fast_ft-ex; rescaler=1.165



(a) 层数 **p=4**　(b) 层数 **p=8**

图 4　微调优化后的预制表参数

总结：我们的工作

- 重缩放因子
- 多策略预制表微调
    - 学习率衰减
    - 自适应损失
    - 动量
    - 均权-加权分离
- 线性插值外推

- ⭐结果：30641.0100 分

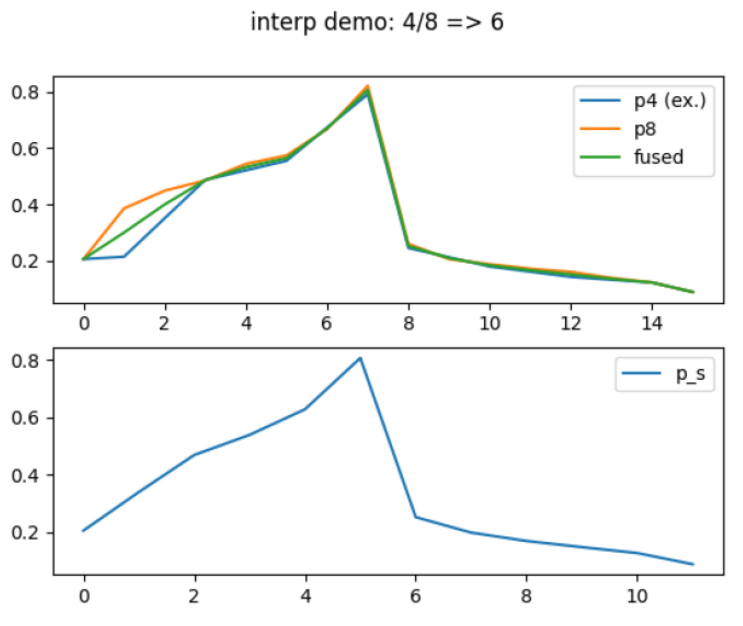


图 **3** 最优参数预制表的线性插值外推

谢谢观看

量子近似优化算法中的最优初始化参数

队名: Quiscus