

# 量子近似优化算法 QAOA 中的最优初始化参数

## 摘要

在量子近似优化算法 (QAOA) 中, 变分量子线路初始参数的选择会极大影响迭代优化的效率。本赛题要求选手给出一个算法, 对任意给定 QAOA 线路给出一组初始化参数, 使得初始状态下目标函数尽可能小。本文以一份预制的最优参数表为基础, 对其进行多策略的微调训练, 可进一步降低标准 QAOA 线路初始状态下的期望值。该方法训练所得的预制表可进一步通过线性插值的方法, 外推到训练集中未见过的其他线路深度的情况。实验表明, 我们所提出的方法最终得分 30641.0100, 超过基线方法 22.56%。

**关键字:** 量子近似优化算法 QAOA 参数初始化 最优化问题

## 一、问题背景

量子近似优化算法 (QAOA) 是含噪中等规模量子计算机时代最具有应用潜力的变分量子算法之一, 可用于高效地近似求解诸如最大割 (MaxCut) 与最小割 (MinCut)、二次无约束二元优化问题 (QUBO)、Ising-模型等被证明为 NP-Hard 的离散组合优化问题, 或用于线性系统求解 (Linear-system Solver) 的加速。QAOA 通过迭代优化一个变分量子线路, 最小化系统哈密顿量的期望, 来近似模拟量子绝热演化过程, 制备含有目标解信息的末态。其中变分量子线路初始参数的选择会极大影响之后迭代优化过程的效率, 乃至最终收敛到解的解的质量。因此, 线路初参选择是本领域中一个重要的讨论课题。

## 二、问题分析

本赛题要求选手给出一个变分线路参数初始化方案, 对于任意给定的带权重高阶 Ising 模型所对应的标准 p-层 QAOA 线路, 如图 1 所示, 给出线路的初始参数  $\gamma_p$  和  $\beta_p$ , 使得线路在迭代优化之前的初始状态下的目标函数尽可能小。

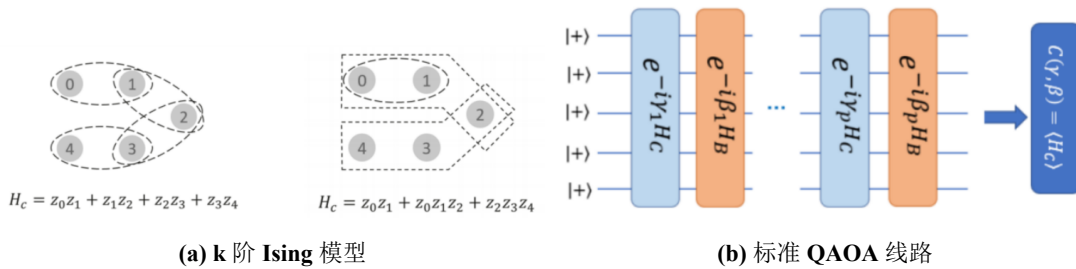


图 1 Ising 模型及其对应的 QAOA 线路

在目前相关工作中，对于此初参选择问题的主要方案是参数转移，即依据类比逻辑和损失函数地形图的拓扑相似性，将已知问题设置的最优参数经过一定的规则放缩从而转移到新的问题设置情况。例如，文献 [1] 提供了一张预制表，给出了在图的度数  $D \rightarrow \inf$  情况下无权图的 MaxCut 问题所对应 QAOA 线路层数从  $p=1$  到  $p=17$  的最优线路参数  $\gamma^{\inf}$  和  $\beta^{\inf}$ 。文献 [2] 提出公式 1 来对此预制表中的  $\gamma^{\inf}$  参数进行放缩变换得到  $\gamma^*$ ，以迁移应用到带权图 MaxCut 问题的情况。文献 [3] 进一步将其改进为公式 2，优化了初参质量，极大地降低后续线路优化所需的迭代次数。

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

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

### 三、解决方案

我们的解决方案主要包含三部分：为上述  $\gamma$ -迁移变换公式引入重缩放因子、对预制表进行多策略的微调训练、线性插值外推方案。

#### 3.1 重缩放因子

观察到公式 1 和公式 2 的分子  $\gamma^{\inf}$  和  $\gamma^{\text{median}}$  是相近的，核心差异主要在分母不同：文献 [2] 使用了  $L_1$  距离，而文献 [3] 使用了  $L_2$  距离。我们没有尝试别的闵可夫斯基  $L_p$  距离，但受此启发，我们在公式 2 的基础上，引入一个乘性的全局常量因子 **rescaler** 来作整体调控，并且使用  $\gamma^{\inf}$  来替换  $\gamma^{\text{median}}$ ，最终如公式 3 所示。

$$\gamma^* = \text{rescaler} * \frac{\gamma^{\inf}}{\sqrt{\frac{1}{|E|} \sum_{\{u,v\} \in E} w_{uv}^2}} \arctan\left(\frac{1}{\sqrt{D-1}}\right) \quad (3)$$

#### 3.2 多策略微调预制表

观察到文献 [2] 和文献 [3] 都只对  $\gamma$  作了迁移变化，对于  $\beta$  参数仍然直接沿用了文献 [1] 预制表中的  $\beta^{\inf}$ ，这或许是一个可以优化的点。另一方面，文献 [1] 给出的最优初参在  $p=4,8$  的情况下可如图 2a 所示：左半递增部分是  $\gamma^{\inf}$ ，右半递减部分是  $\beta^{\inf}$  参数。这些变化曲线的趋势和我们所见在一些量子绝热演化应用 [4] 中所设计的哈密顿量调度函数（如图 2b所示）曲线有些不太一致，即不太满足  $\gamma$  曲线单增，且其导函数单减的性质，这可能也是一个可以改进的方面。

综合上述两点，我们采用微调训练的方式进一步优化预制表中的参数  $\gamma^{\inf}$  和  $\beta^{\inf}$ ，即使用各具体样本的最优参数来调整全局平均最优参数。最朴素的更新策略如公式 4 所

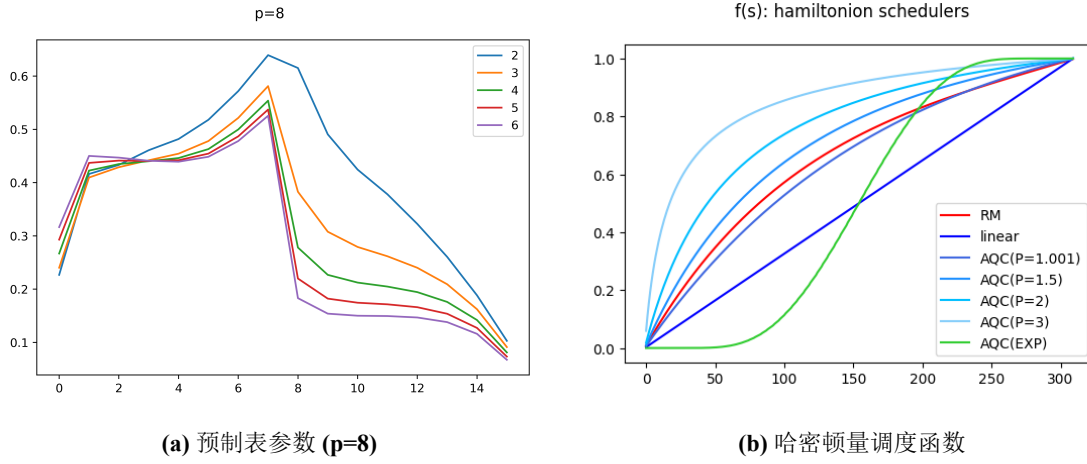


图 2 预制表参数与哈密顿量调度函数

示，其中  $\theta_i^*$  为第  $i$  次训练迭代后的全局平均最优参数，而  $\hat{\theta}_i$  为第  $i$  次训练时所用样本的最优参数， $\Delta x$  可理解为学习率。

$$\theta_{i+1}^* = (1 - \Delta x) * \theta_i^* + \Delta x * \hat{\theta}_i \quad (4)$$

为了获得更好的训练效果，我们进一步引入了学习率衰减、自适应损失、差分动量的更新策略，此时策略的更新公式变为 5。其中  $\Delta x$  依然为基准学习率； $w$  为学习率衰减因子，受到迭代次数的累计而减小； $\Delta E$  为各样本优化前后的能量值之差，来对数尺度地放缩学习率，使得能量值差越大的样本的个体参数才能对全局参数产生足够大影响；而  $\mu_i^*$  为第  $i$  此迭代时累计产生的最优参数的动量， $m$  为动量平衡因子，这变相实现了小批次平均的梯度下降，保证了训练过程的稳定。最后，观察到无权图和带权图在数据分布上的显著不一致，我们认为使用同一张预制表来应对两种情况可能会导致一些模型建模内在的矛盾。所以我们以图权值的方差是否超过特定阈值为判据，为近似无权图和带权图分别建立一套预制表，这就是无权-带权分离建模策略。

$$\begin{aligned}
 \Delta E &= E_{\text{before}} - E_{\text{after}} \\
 \tilde{\Delta x} &= \Delta x * \left( w_{\frac{\text{cur\_iter}}{\text{decay\_per\_iter}}} \right) * \log(1 + \Delta E) \\
 \mu_{i+1}^* &= (1 - m) * \mu_i^* + m * \hat{\theta}_i \\
 \theta_{i+1}^* &= (1 - lr) * \theta_i^* + \tilde{\Delta x} * \mu_{i+1}^*
 \end{aligned} \quad (5)$$

### 3.3 线性插值外推

预制表是针对特定层数  $p=4,8$  的 QAOA 线路进行建模的，当目前线路为其他层数时，可以使用如图 3 所示的线性插值外推来获得次优初参。以 6 层为例，首先将已建模的层数  $p=4$  (蓝色) 和  $p=8$  (橙色) 的最优参数进行长度对齐，然后求平均得到 fused (绿色)

参数，最后将其以首尾为锚点等比例线性拉伸为目标参数长度，保持轮廓一致性。须注意该方法不只对于内插值有效，小于 4 层或大于 8 层的操作流程也完全一致。

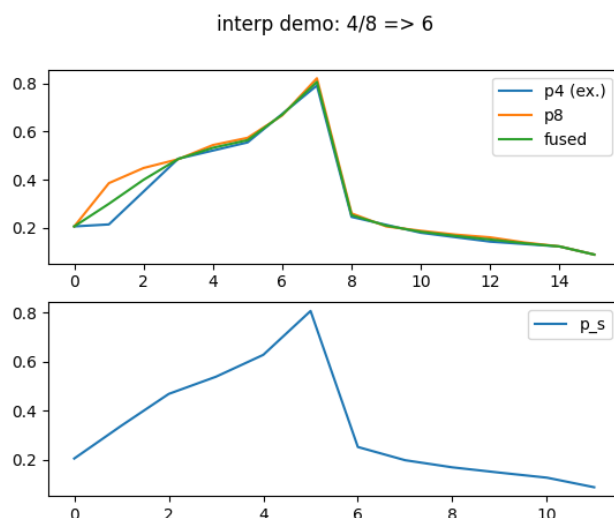


图 3 最优参数预制表的线性插值外推

## 四、结果

我们使用 Adam 优化器进行 QAOA 线路的优化，在赛题给出的数据集上训练每个样本 100 步 (fast 模式下仅训练 40 步)。其他超参设置为  $lr = 1e - 5$ ,  $\Delta x = 0.1$ ,  $w = 0.98$ ,  $\text{decay\_per\_iter} = 100$ ,  $m = 0.6$ 。实验结果如表 1 所示，简要介绍如下：

- 方法 #1 和 #2 为基线，其中 #2 为比赛模板代码给出的标准基线，而 #1 为其前作基础，可见距离函数的选择对性能影响巨大
- 方法 #3 在标准基线 #2 的基础上仅仅调整了 rescaler 就导致了 8.87% 的分数提升，说明了该参数对样本共性的重要影响
- 方法 #4 为朴素的优化-平均，即对所有样本各自求得最优参数后作简单平均。这引起了 12.04% 的分数提升，说明样本间差异也是重要的考虑因素
- 方法 #5 为朴素的微调，即对预制表按公式 4 进行迭代优化，它只比 #4 略好一些
- 方法 #6 ~ #10 为本文提出的多策略微调算法，即对预制表按公式 5 进行迭代优化
  - 方法 #6 ~ #10 均引入了损失值自适应策略，抑制恶劣样本产生的负面影响
  - 方法 #7 ~ #9 引入了学习率衰减策略，确保微调过程能呈现为收敛态势
  - 方法 #8 ~ #10 引入了差分动量策略，可使用更少的优化步数以节省时间
  - 方法 #10 在 #9 的基础上引入了均权-加权分离的策略，并在此配置下达到了我们的最高分 30641.0100，相对于标准基线提升了 22.56%

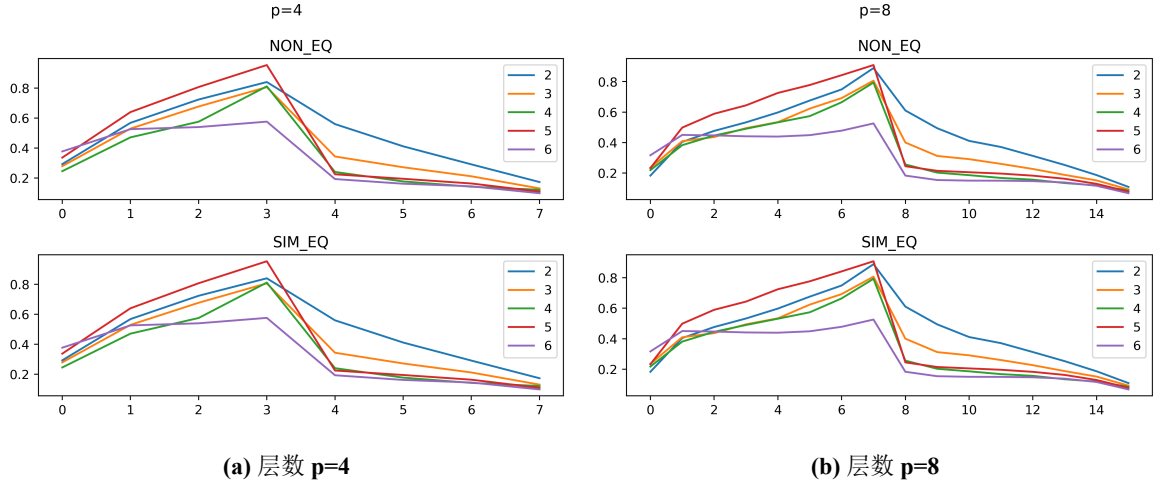


图 4 微调优化后的预制表参数

- 注意到方法 #9 和 #10 都重新调整了重缩放因子，是因为经过优化的新参数已经偏离了原始的位置，为了适应新的方差必须重新确定；而在 #10 中去掉了学习率衰减策略，是因为此时平均损失值已经很小了，再作放缩将无法推动学习

表 1 实验结果

#	local score	submit score	comment
#1	11730.14583	16627.4960	PT-WMC [2], reference
#2	16526.79871	<u>24999.9025</u>	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	<b>30641.0100</b>	ft-ada-moment-fast_ft-ex; rescaler=1.165

微调优化后的预制表参数如图 4 所示，可见左半  $\gamma$  部分已满足函数单增且导函数单减的性质，符合我们在量子绝热演化应用中的相应观察。NON\_EQ 和 SIM\_EQ 分别对应于带权图和近似无权图的预制表，在此图示中看不出显著的差异，但各数值上确有  $0.001 \sim 0.1$  量级的均值偏差。

## 五、总结及展望

针对 QAOA 线路最优初参选取问题, 本文提出了一种基于预制表进行多策略微调的方法, 进一步降低了线路初始损失值, 为后续迭代优化创造有利起点。提出了一种线性插值外推的方式, 可在保持参数数值比例一致的条件下进行内外插值, 以扩展训练优化后预制表的应用范围。

在比赛数据集上的进一步的实验表明, 我们所提方法能达到的最好表现, 仍然距离每个样本被优化到各自最优解这种完美情况仍有 10% 的差距, 这或许表明了此系列方法还有长足的改进空间, 亦或许暗示了这个问题本身的复杂, 即仍然存在某些各向异性的因素, 很难通过统计建模去总体性地刻画。未来的工作或将向着离群性样本, 奇异点和相变的方向讨论和发展。

## 参考文献

- [1] BASSO J, FARHI E, MARWAHA K, et al. The quantum approximate optimization algorithm at high depth for maxcut on large-girth regular graphs and the sherrington-kirkpatrick model[C/OL]//Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022. <https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.TQC.2022.7>. DOI: 10.4230/LIPIcs.TQC.2022.7.
- [2] SHAYDULIN R, LOTSHAW P C, LARSON J, et al. Parameter transfer for quantum approximate optimization of weighted maxcut[J/OL]. ACM Transactions on Quantum Computing, 2023, 4(3). <https://doi.org/10.1145/3584706>.
- [3] SURESHBABU S H, HERMAN D, SHAYDULIN R, et al. Parameter Setting in Quantum Approximate Optimization of Weighted Problems[J/OL]. Quantum, 2024, 8:1231. <https://doi.org/10.22331/q-2024-01-18-1231>.
- [4] AN D, LIN L. Quantum linear system solver based on time-optimal adiabatic quantum computing and quantum approximate optimization algorithm[J/OL]. ACM Transactions on Quantum Computing, 2022, 3(2):1–28. <http://dx.doi.org/10.1145/3498331>.

## 附录 A 主要代码

预制表的微调训练流程:

```
1 for iter in range(n_samples):
2     # 优化前的参数 & 能量
3     params, E_before = ..., rc()
4     # 优化线路!
5     optimize(circ)
6     # 优化后的参数 & 能量
7     tuned_params, E_after = ..., circ()
8     # 损失值自适应 + 学习率衰减
9     ΔE = E_before - E_after
10    lr = dx * (dx_decay ** (iter / dx_decay_every)) * log1p(ΔE)
11    # 更新动量状态 & 预制表参数
12    mu = (1 - m) * mu + m * tuned_params
13    theta = (1 - lr) * theta + lr * mu
```

预制表的查表方法:

```
1 # 计算图的度和阶数
2 D = ave_D(Jc_dict, Nq)
3 k = min(order(Jc_dict), 6)
4 # 以权值的方差判断无权-带权图
5 theta = theta_ex[SIM_EQ if std(Jc_dict) < THRESHOLD else
6                 NON_EQ]
7 # 训练时见过的配置, 直接使用
8 if p in [4, 8] and k in [2, 3, 4, 5]:
9     params = theta[p][k]
10 else: # 否则使用外插扩展
11     params = interp_expand(p, theta[4][k], theta[8][k])
12 gammas, betas = np.split(params, 2)
13 gammas = trans_gamma(gammas, D)
14 # 处理 rescaler 因子, 返回初参 gamma 和 beta
15 factor = rescale_factor(Jc_dict) * rescaler
16 return gammas * factor, betas
```