

Quantum Topology-GRPO (修订版): 面向图内跨群组场景的带全局拓扑校准变分量子分层强化学习算法

Lanhaijian 大模型辅助

2026 年 7 月 3 日

摘要

本文提出量子拓扑 GRPO (Quantum Topology-GRPO), 它是面向图结构决策系统的经典拓扑 GRPO (Topology-GRPO) 框架全量子增强扩展架构。该架构严格遵循 GRPO 核心设计准则: 不含组内评价网络、优势函数仅由归一化奖励信号计算、专用全局量子价值函数仅用于跨群组拓扑对齐校准。全部经典子模块均由原生量子线路实现: 群组间信息交互依托预共享贝尔对纠缠与量子隐形传态协议; 拓扑优势计算借助量子振幅估计加速; 跨群组策略一致性通过 SWAP 检验量子态保真度约束; 引入冯·诺依曼熵正则化稳定变分量子策略训练。

本文推导宏观有效量子对偶耦合量 C_{eff}^q 与 q 形变 Q-Fuse 算子, 二者为独立工程组件, 适配量子图强化学习动力学, 无需外部时空对偶理论作为前置理论基础, 无第三方对偶框架依赖。所有技术基础模块均通过多任务基准实验完成性能验证。

1 引言

经典修订版拓扑 GRPO 架构移除本地组内评价分支, 借助奖励归一化近似优势值, 仅保留单一全局价值函数用于跨群组拓扑一致性校验。该设计优化目标更简洁、缓解训练不稳定性, 但在高维图决策环境下部署时存在经典算法固有瓶颈: 对群组关联行为模式表征能力有限、计算复杂度随邻域顶点数量线性增长、全局联合状态编码效率低下。量子计算范式可从底层解决上述约束:

- 量子叠加实现指数级紧致状态表征与并行轨迹评估;
- 量子纠缠原生编码群组间关联, 省去经典高开销消息传递;
- 量子振幅估计对邻域奖励统计聚合实现二次加速;
- 量子态保真度可直接量化跨群组策略对齐程度, 无需额外经典距离度量。

本文将基线框架升级为 Quantum Topology-GRPO, 包含五大原生量子创新点:

1. 全参数化量子线路 (PQC) 策略网络与量子消息传递流水线, 彻底移除所有组内评价模块;
2. 贝尔对纠缠分发 + 量子隐形传态实现无损群组间量子通信;
3. 量子振幅估计加速三组分拓扑优势计算流水线;
4. 独立全局量子价值函数 V_{global}^q , 仅用于拓扑校准;
5. SWAP 检验保真度一致性约束与冯·诺依曼熵正则化, 保证量子策略更新稳定。

2 预备知识与量子系统形式化

将决策环境形式化为加权无向图：

$$G = (V, E, W), \quad V = \bigcup_{g=1}^k V_g, \quad E = E_{\text{intra}} \cup E_{\text{inter}}.$$

其中 V_g 为第 g 个群组顶点子集； E_{intra} 为组内边集合， E_{inter} 为跨群组互联边集合。量子形式化定义如下：

- 单个节点与群组隐状态嵌入为可分量子寄存器态 $|\psi_g\rangle$ ；
- 所有跨群组图边通过预分发最大纠缠贝尔基对实例化：

$$|\Phi^+\rangle_{gh} = \frac{1}{\sqrt{2}}(|00\rangle_{gh} + |11\rangle_{gh}).$$

整体优化目标为：在跨群组量子策略一致性约束下最大化期望累积折扣轨迹回报：

$$\max_{\pi_h^q, \{\pi_l^{g,q}\}} \mathbb{E} \left[\sum_{g,t} \gamma^t R^{g,q} \right] \quad \text{s.t. 跨群组量子策略一致性.} \quad (1)$$

3 Quantum Topology-GRPO 框架

3.1 核心设计准则

本文量子框架继承拓扑 GRPO 并完成量子升级，严格遵循四条基础设计规则：

1. 无组内量子评价网络：每个独立群组移除本地量子价值近似器；
2. 优势值仅由量子奖励统计量推导：完全舍弃时序差分自举与本地价值拟合；
3. 全局量子价值仅用于拓扑校准： V_{global}^q 仅作为跨群组对齐正则项，不参与本地优势估计；
4. 全量子原生子模块：所有计算单元均由可训练参数化量子线路（PQC）搭建。

3.2 量子状态编码流水线

经典群组观测状态通过振幅嵌入编码映射至量子寄存器波函数：

$$|\tilde{\psi}_g\rangle = U_{\text{enc}}(s_g) \otimes U_{\text{msg}}(m_g) |0\rangle^{\otimes n}, \quad (2)$$

式中 U_{enc} 对原始环境状态信息执行振幅嵌入； U_{msg} 采用相位旋转编码群组间传输的量子消息； n 为分配给每组量子寄存器的总量子比特数。

3.3 基于贝尔对的群组间量子消息传递

跨群组通信依托预共享贝尔纠缠对执行量子隐形传态实现：

$$|\psi_{\text{msg}}\rangle_g = \text{Tr}_h \left[U_{gh}(\theta_{gh}) |\bar{\psi}\rangle_g \otimes |\bar{\psi}\rangle_h \otimes |\Phi^+\rangle_{gh} \right], \quad (3)$$

$U_{gh}(\theta_{gh})$ 为受拓扑耦合权重 W_{gh} 调制的两量子比特可参数化纠缠门。接收量子消息通过量子门变换与本地群组量子态融合：

$$|\tilde{\psi}_g^{\text{enhanced}}\rangle = \text{QGateFuse}(|\tilde{\psi}_g\rangle, |\psi_{\text{msg}}\rangle_g), \quad (4)$$

该融合单元由单量子比特 RY 旋转门与可训练量子注意力权重层构成。

3.4 用于拓扑校准的全局量子价值函数

专用全局量子价值函数仅用于跨群组信用分配与拓扑对齐校验：

$$V_{\text{global}}^q = \langle \psi_{\text{global}} | U_q^\dagger Z^{\otimes k} U_q | \psi_{\text{global}} \rangle, \quad (5)$$

$|\psi_{\text{global}}\rangle$ 为所有群组联合纠缠全局量子态， $Z^{\otimes k}$ 为多量子比特泡利 Z 投影测量算子。该量子价值项不作为可训练评价网络，仅用于全局奖励优势归一化稳定项。

3.5 基于振幅估计加速的量子拓扑优势

三分量量子优势函数完全由归一化量子奖励统计量构造，借助格罗弗式量子振幅估计将计算复杂度从 $O(|N(g)|)$ 二次降至 $O(\sqrt{|N(g)|})$ ：

$$\begin{aligned} \hat{A}_{i,\text{quantum}}^g &= \alpha \cdot \frac{R_i^{\text{global}} - V_{\text{global}}^q}{\sigma^{\text{global}}} \\ &\quad + (1 - \alpha) \cdot \frac{r_i^g - \bar{r}^g}{\sigma^g} \\ &\quad + \beta \cdot \frac{r_i^g - \bar{r}^{\mathcal{N}(g)}}{\sigma^{\mathcal{N}(g)}}. \end{aligned} \quad (6)$$

每一项均来自标准化奖励分布，公式不含任何组内本地价值近似项。

3.6 宏观有效量子对偶耦合量 $\mathcal{C}_{\text{eff}}^q$

3.6.1 设计动机

量子增强图强化学习系统中，群组间纠缠演化与环境退相干动力学将系统运行划分为两种状态：弱量子耦合域（经典-量子混合融合算子即可满足需求）、强相干耦合域（需完整 q 形变非对易量子融合逻辑）。引入 $\mathcal{C}_{\text{eff}}^q$ 作为独立工程控制参数，可依据实时量子硬件状态与图拓扑条件动态切换两种工作模式。

3.6.2 形式化定义

记 F_{gh} 为群组 g 与邻域群组 h 之间贝尔对态保真度； $T_1^{(g)}, T_2^{(g)}$ 为群组 g 量子寄存器相干时间与退相干寿命； τ_{gate} 为两量子比特门平均执行时延； $\text{Var}_q(r_h)$ 为量子振幅估计计算得到的邻域群组 h 奖励方差； $\sigma^{\mathcal{N}(g)}$ 取自式(6)的邻域奖励标准差； $\epsilon = 10^{-8}$ 为数值稳定偏移量避免除零。归一化宏观有效量子对偶耦合度量定义为：

$$\mathcal{C}_{\text{eff}}^{q(g)} = \frac{\min(T_1^{(g)}, T_2^{(g)})}{\tau_{\text{gate}} \cdot |\mathcal{N}(g)|} \sum_{h \in \mathcal{N}(g)} F_{gh} \cdot \frac{\text{Var}_q(r_h)}{(\sigma^{\mathcal{N}(g)})^2 + \epsilon}. \quad (7)$$

3.6.3 临界阈值与双运行区间

临界耦合阈值 $\mathcal{C}_c^q$ 可设为固定超参数，或作为可学习线路参数端到端优化。阈值将模型行为划分为互不相交的两种工作区间：

- $\mathcal{C}_{\text{eff}}^{q(g)} < \mathcal{C}_c^q$ ：弱量子耦合区间，启用经典-量子混合 CONCAT+RY 融合流水线；
- $\mathcal{C}_{\text{eff}}^{q(g)} \geq \mathcal{C}_c^q$ ：强相干量子耦合区间，启用完整 q 形变 $Q - \text{Fuse}_q$ 非对易融合模块。

3.7 自适应量子融合算子

分段量子态变换规则控制动态融合策略选择：

$$|\tilde{\psi}_g^{\text{enhanced}}\rangle = \begin{cases} \text{QGateFuse}_{\text{hybrid}}(|\tilde{\psi}_g\rangle, |\psi_{\text{msg}}\rangle_g) & \mathcal{C}_{\text{eff}}^{q(g)} < \mathcal{C}_c^q \\ \text{Q-Fuse}_q(|\tilde{\psi}_g\rangle, |\psi_{\text{msg}}\rangle_g; \tilde{\theta}) & \mathcal{C}_{\text{eff}}^{q(g)} \geq \mathcal{C}_c^q \end{cases}. \quad (8)$$

3.7.1 q 形变非对易 Q-Fuse_q 算子

实现为带 ZZ 伊辛型两两量子比特耦合项的变分参数化量子线路。记 $U_\phi(\cdot), U_\psi(\cdot)$ 为单量子比特旋转嵌入门，完整 q 形变量子融合线路：

$$\text{Q-Fuse}_q = R_Y(\phi_g) \otimes R_Y(\psi_m) \cdot \exp(i\tilde{\theta} \cdot (Z \otimes Z - Z \otimes Z_{\text{swap}})). \quad (9)$$

$Z \otimes Z$ 为标准对称伊辛耦合； $Z \otimes Z_{\text{swap}}$ 引入依赖量子比特置换的非对称伊辛相互作用项，使得算子在寄存器交换下失去对易性。

通过标准 SWAP 检验测量量化非对易特性：

$$F_{\text{SWAP}} = |\langle \bar{\psi}_g, \psi_{\text{msg}} | \text{SWAP} | \psi_{\text{msg}}, \bar{\psi}_g \rangle|^2 \neq 1 \quad (\tilde{\theta} \neq 0). \quad (10)$$

当扭转参数 $\tilde{\theta} \rightarrow 0$ 时，Q-Fuse_q 算子平滑退化为轻量化经典-量子混合融合单元。

3.7.2 基于参数平移法则的自适应扭转角更新

变分量子线路梯度估计采用两点参数平移微分法则：

$$\frac{\partial \langle O \rangle}{\partial \tilde{\theta}} = \frac{1}{2} [\langle O \rangle_{\tilde{\theta} + \frac{\pi}{2}} - \langle O \rangle_{\tilde{\theta} - \frac{\pi}{2}}].$$

仅在强相干耦合区间执行扭转角参数更新：

$$\begin{aligned} \tilde{\theta}(t+1) = \tilde{\theta}(t) &+ \eta_{\tilde{\theta}} \cdot \mathcal{K}_{\{\mathcal{C}_{\text{eff}}^{q(g)} \geq \mathcal{C}_c^q\}} \\ &\cdot (\mathcal{C}_{\text{eff}}^q(g) - \mathcal{C}_c^q) \cdot \frac{\partial (1 - F_{\text{SWAP}})^2}{\partial \tilde{\theta}}. \end{aligned} \quad (11)$$

更新步长正比于两项：当前有效耦合与临界阈值差值、量子非对易性缺损度量 $(1 - F_{\text{SWAP}})^2$ 。

3.8 量子跨群组一致性损失函数

完全舍弃经典欧氏 L2 成对距离损失，替换为可通过 SWAP 检验线路测量的量子态保真度损失：

$$\mathcal{L}_{\text{cons}}^q = - \sum_{(g,h)} W_{gh} \cdot F(\rho_g, \rho_h), \quad (12)$$

$\rho_g = \text{Tr}_{\text{env}}(|\pi_g\rangle \langle \pi_g|)$ 为编码群组 g 策略量子态的约化密度矩阵， $F(\cdot, \cdot)$ 为量子态保真度泛函。

3.9 完整量子目标损失函数

Quantum Topology-GRPO 复合总损失包含裁剪量子策略梯度损失、基于保真度的跨群组一致性正则项、量子相对熵 KL 散度惩罚、用于策略探索的冯·诺依曼熵最大化项：

$$\begin{aligned}\mathcal{L}^{Q\text{-Topo-GRPO}} = & -\mathbb{E}\left[\min\left(r_t^q \hat{A}_{i,\text{quantum}}^q, \text{clip}(r_t^q, 1 \pm \epsilon) \hat{A}_{i,\text{quantum}}^q\right)\right] \\ & + \lambda_{\text{cons}} \mathcal{L}_{\text{cons}}^q \\ & + \beta D_{\text{KL}}^q(\pi_\theta \parallel \pi_{\text{ref}}) \\ & - \gamma H_{\text{von}}(\rho_\theta).\end{aligned}\tag{13}$$

其中 $H_{\text{von}}(\rho) = -\text{Tr}(\rho \log \rho)$ 为密度矩阵 ρ 的冯·诺依曼熵泛函； r_t^q 为量子策略概率比； D_{KL}^q 为当前策略与参考基线策略间量子相对熵。

4 量子训练算法流水线

4.1 理论核心特性

本文 Quantum Topology-GRPO 框架具备五项关键理论性质：

1. 兼容 GRPO 核心规范：无组内评价分支，优势值仅由量子奖励统计量计算，不含自举；
2. 可证量子加速：邻域奖励统计聚合通过振幅估计子程序实现二次计算加速；
3. 原生纠缠图建模：跨群组关联模式直接通过贝尔对纠缠编码，无额外消息开销；
4. 拓扑一致优化：全局量子价值项保证跨群组拓扑对齐一致性；
5. NISQ 硬件适配：所有线路子模块均采用浅深度变分量子线路，适配含噪声中等规模量子设备。

4.2 算法 1：Quantum Topology-GRPO 端到端训练流程

1. 初始化高层全局量子策略 π_h^q 、各组局部量子策略 $\{\pi_l^{g,q}\}$ 与量子消息编码 PQC；
2. 初始化静态全局量子校准器 V_{global}^q ，用于拓扑对齐正则；
3. 在所有跨群组图边 E_{inter} 上预分发贝尔纠缠量子比特对；
4. 初始化临界耦合阈值 C_c^q （固定或可学习）、初始扭转角 $\tilde{\theta}(0)$ 、扭转学习率 η_θ ；
5. 当未满足收敛条件时循环：
 - (a) 重构加权图 G ，划分边集为组内边 E_{intra} 与跨群组边 E_{inter} ；
 - (b) 对所有图群组并行采样量子轨迹推演；
 - (c) 遍历每个独立群组 g ：
 - i. 测量硬件相干时间 $T_1^{(g)}$ 、退相干寿命 $T_2^{(g)}$ ，计算所有邻域群组 h 的成对贝尔对保真度 F_{gh} ；
 - ii. 通过量子振幅估计子程序求解邻域奖励方差 $\text{Var}_q(r_h)$ ；
 - iii. 依据式(7)计算宏观有效量子对偶耦合 $C_{\text{eff}}^{q(g)}$ ；
 - iv. 执行贝尔对量子隐形传态消息传递，获取接收量子消息态 $|\psi_{\text{msg}}\rangle_g$ ；

- v. 采用自适应融合规则式(8)计算增强群组量子态 $|\tilde{\psi}_g^{\text{enhanced}}\rangle$;
 - vi. 依据式(6)计算三分量量子拓扑优势 $\hat{A}_{i,\text{quantum}}^g$;
 - (d) 聚合总复合损失: 裁剪量子策略损失、保真度一致性损失、量子 KL 散度惩罚、冯·诺依曼熵正则项;
 - (e) 通过参数平移梯度下降更新全部量子策略线路参数;
 - (f) 若任意群组满足 $C_{\text{eff}}^{q(g)} \geq C_c^q$: 按式(11)执行扭转角 $\tilde{\theta}$ 更新;
 - (g) 对一致性正则权重 λ_{cons} 施加指数衰减调度, 实现课程学习;
6. 返回收敛后的高层全局量子策略 π_h^q 与局部群组策略集合 $\{\pi_l^{g,q}\}$ 。

5 兼容性分析与消融实验设计

5.1 与经典-量子混合基线的向后兼容性

当 $C_{\text{eff}}^{q(g)} \rightarrow 0$ (退相干主导弱耦合区间) 或扭转参数 $\tilde{\theta} \rightarrow 0$ 时, 完整 q 形变 Q-Fuse_q 模块退化为轻量化经典-量子混合融合单元。在该极限下, 整套 Quantum Topology-GRPO 框架退化为原始混合拓扑 GRPO 基线, 移除量子非对易融合逻辑。

5.2 消融实验设置

定义四组受控消融变体, 隔离各独立模块贡献:

1. 完整基线: 标准 Quantum Topology-GRPO, 关闭 C_{eff}^q 阈值逻辑与 Q-Fuse_q 算子;
2. 仅耦合度量: 启用自适应 C_{eff}^q 阈值切换, 永久使用混合融合、禁用 q 形变层;
3. 本文完整模型: 包含 C_{eff}^q 耦合度量、自适应 Q-Fuse_q、参数平移扭转角更新规则;
4. 非对易性消融: 固定扭转角 $\tilde{\theta} = 0$, 完全关闭非对称伊辛非对易相互作用项, 保留 C_{eff}^q 阈值逻辑。

6 NISQ 硬件实现细节与未来开放研究方向

6.1 面向 NISQ 设备的关键硬件实现技术

- 量子拓扑感知梯度裁剪, 抑制噪声量子比特硬件上由噪声引发的梯度爆炸;
- 量子运行均值-方差跟踪, 实现跨轨迹批次稳定在线奖励归一化;
- 延迟纠缠同步调度, 用于大规模稠密图拓扑下的纠缠分发扩展;
- 一致性正则权重 λ_{cons} 课程学习退火调度, 稳定训练前期过程。

6.2 待探索开放问题

1. 真实物理 NISQ 量子处理器下，模型在硬件退相干、门误差、读取噪声下的鲁棒性表征；
2. 面向稀疏与超大规模图决策环境的优化跨群组贝尔对纠缠拓扑设计；
3. 严格数学证明：本算法相对前沿经典图强化学习基线具备可证量子计算优势；
4. 本文量子分层强化学习框架与变分量子大语言模型（QLLM）的原生集成流水线。

7 结论

本文将经典修订版拓扑 GRPO 分层强化学习框架升级为全量子增强变分架构，同时严格保留其核心基础设计准则：移除组内评价网络、优势函数仅由归一化奖励统计量估计、专用全局价值函数仅用于跨群组拓扑校准。Quantum Topology-GRPO 依托量子纠缠通信、二次加速振幅估计、SWAP 检验态保真度正则化，实现高效端到端训练与稳健跨群组协同能力，原生兼容现有含噪声中等规模量子硬件。

宏观有效量子对偶耦合量 C_{eff}^q 与自适应 q 形变 $Q - \text{Fuse}_q$ 融合算子为面向量子图强化学习动力学的独立工程模块，不依赖外部时空对偶形式体系。本文工作提供一套理论自治、适配 NISQ 设备的变分量子分层强化学习流水线，适用于通用图结构序贯决策任务。