

Topology-GRPO (修订版): 基于宏观有效对偶耦合的自适应 q 形变消息融合

Lanhaijian 大模型辅助

摘要

本文针对 Topology-GRPO 框架完成修订, 引入由宏观有效对偶耦合控制的自适应消息融合机制。该图分组策略优化算法可根据组间相干耦合强度, 在经典可交换聚合与 q 形变非交换融合之间动态切换。所有模块均为独立建模假设, 且通过下游任务指标完成有效性验证。

1 修订版 Topology-GRPO 框架

1.1 三分量拓扑优势

三分量拓扑优势仅通过奖励归一化计算, 不含组内评价网络, 严格遵循分组相对策略优化 (GRPO) 核心准则。对任意分组 g , 拓扑优势定义为:

$$\begin{aligned}\hat{A}_i^g = & \alpha \cdot \frac{R_i^{global} - \bar{R}^{global}}{\sigma^{global}} \\ & + (1 - \alpha) \cdot \frac{r_i^g - \bar{r}^g}{\sigma^g} \\ & + \beta \cdot \frac{r_i^g - \bar{r}^{\mathcal{N}(g)}}{\sigma^{\mathcal{N}(g)}}.\end{aligned}$$

1.2 宏观有效对偶耦合与自适应消息融合

1.2.1 研究动机

图结构多分组系统的组间信息传播存在两种运行区间: 弱耦合场景下采用经典可交换特征融合; 强相干耦合场景会产生非交换交互效应。本文引入宏观有效对偶耦合 C_{eff} 区分两类区间, 并自适应切换消息融合策略。

1.2.2 宏观有效对偶耦合 $C_{eff}^{(g)}$ 定义

记 $\tau_{coh}^{(g)}$ 为分组 g 的局部策略相干步数, $\tau_{delay}^{(g)}$ 为组间平均消息延迟步数, W_{gh} 为分组 g 与邻域分组 h 的拓扑耦合权重, $Var(r_h)$ 为邻域分组 h 的奖励方差; $\sigma^{\mathcal{N}(g)}$ 为式 (1) 中邻域集合奖励标准差, $\epsilon = 10^{-8}$ 为数值稳定项, 避免除零。归一化宏观有效对偶耦合表达式为:

$$C_{eff}^{(g)} = \frac{\tau_{coh}^{(g)}}{\tau_{delay}^{(g)}} \cdot \frac{1}{|\mathcal{N}(g)|} \sum_{h \in \mathcal{N}(g)} W_{gh} \cdot \frac{Var(r_h)}{(\sigma^{\mathcal{N}(g)})^2 + \epsilon}.$$

1.2.3 临界阈值与运行区间划分

c_c 为临界耦合阈值, 可设为固定超参, 也可在训练中作为可学习参数优化。以 c_c 划分两种工作区间:

- $\mathcal{C}_{eff}^{(g)} < c_c$: 弱耦合区间, 采用经典拼接 + 多层感知机 (Concat+MLP) 消息融合;
- $\mathcal{C}_{eff} \geq c_c$: 强相干耦合区间, 启用 q 形变 Q-Fuse 融合模块。

1.2.4 自适应融合算子

设 s_g 为分组 g 的状态, m_g 为聚合后的组间消息, 自适应融合规则为:

$$\tilde{s}_g = \begin{cases} \text{Concat}(s_g, \sigma(\text{MLP}[s_g; m_g])) \odot m_g & \mathcal{C}_{eff}^{(g)} < \mathcal{C}_c \\ Q - \text{Fuse}(s_g, m_g; \tilde{\theta}) & \mathcal{C}_{eff}^{(g)} \geq \mathcal{C}_c \end{cases}$$

1.2.5 Q-Fuse: q 形变非交换融合模块

设 $\phi(\cdot) = W_\phi \cdot + b_\phi$, $\psi(\cdot) = W_\psi \cdot + b_\psi$ 为两组可学习线性变换, q 形变融合算子定义为:

$$Q - \text{Fuse}(s_g, m_g; \tilde{\theta}) = W_1 s_g + W_2 m_g + \cos(2\pi\tilde{\theta}) \cdot (\phi(s_g) \odot \psi(m_g) - \phi(m_g) \odot \psi(s_g)).$$

交换输入参数可直接验证其非交换特性:

$$Q - \text{Fuse}(m_g, s_g) - Q - \text{Fuse}(s_g, m_g) = 2\cos(2\pi\tilde{\theta}) (\phi(m_g) \odot \psi(s_g) - \phi(s_g) \odot \psi(m_g)) \neq 0.$$

当 $\tilde{\theta} \rightarrow 0$ 或系统落入弱耦合区间时, 非交换项趋于消失, Q-Fuse 退化为经典线性融合。

1.2.6 扭转角 $\tilde{\theta}$ 自适应更新规则

扭转角 $\tilde{\theta}$ 独立于策略网络优化, 仅在强相干耦合区间激活更新:

$$\tilde{\theta}(t+1) = \tilde{\theta}(t) + \eta_\theta \cdot \mathbf{1}_{\{\mathcal{C}_{eff}^{(g)} \geq \mathcal{C}_c\}} \cdot (\mathcal{C}_{eff}^{(g)} - \mathcal{C}_c) \cdot \frac{\partial \|Q - \text{Fuse}(s_g, m_g) - Q - \text{Fuse}(m_g, s_g)\|^2}{\partial \tilde{\theta}}.$$

式中 η_θ 为扭转角学习率; 更新幅度与当前耦合强度距临界阈值的差值成正比。

1.3 整体目标损失函数

总损失继承原始 Topology-GRPO 形式, 包含裁剪策略损失、组间一致性损失与 KL 散度正则项:

$$\mathcal{L}_{\text{Topo-GRPO}} = -\mathbb{E} \left[\min \left(r_t(\theta) \hat{A}^g, \text{clip}(r_t, 1 \pm \epsilon) \hat{A}^g \right) \right] + \lambda_{\text{cons}} \mathcal{L}_{\text{cons}} + \beta D_{\text{KL}}(\pi_\theta \| \pi_{\text{ref}}).$$

2 训练算法流程

修订版 Topology-GRPO 完整训练流程见算法 1。[htbp] 修订版 Topology-GRPO 训练流程

1. 初始化高层策略 π_h 、各组分策略 $\{\pi^{(g)}\}$ 与消息编码器;
2. 初始化全局价值网络 V_{global} , 用于拓扑奖励校准;
3. 初始化超参数: $\alpha, \beta, \epsilon, \lambda_{\text{cons}}$;
4. 初始化临界阈值 c_c (固定/可学习)、初始扭转角 $\tilde{\theta}(0)$ 、扭转角学习率 η_θ ;
5. **while** 训练未收敛 **do**

- (a) 构建组内边集合 E_{intra} 、组间拓扑边集合 E_{inter} ;
- (b) 并行采集所有分组策略交互轨迹;
- (c) **for** 每个分组 g **do**
 - i. 计算 $\tau_{coh}^{(g)}$ 、 $\tau_{delay}^{(g)}$ 、邻域奖励方差 $Var(r_h)$ 、邻域奖励标准差 $\sigma^{\mathcal{N}(g)}$;
 - ii. 代入式 (2) 计算宏观有效对偶耦合 $\mathcal{C}_{eff}^{(g)}$;
 - iii. 执行图消息传递, 得到聚合组间消息 m_g ;
 - iv. 代入式 (3) 计算自适应融合状态 $\tilde{s}_g$;
 - v. 代入式 (1) 计算三分量拓扑优势 $\hat{A}^g$;
- (d) 计算裁剪策略损失、组间一致性损失 $\mathcal{L}_{cons}$ 、KL 散度正则项;
- (e) 最小化总损失 $\mathcal{L}_{Topo-GRPO}$, 更新高层策略 π_h 与分组策略 $\{\pi^{(g)}\}$;
- (f) **if** 存在分组满足 $\mathcal{C}_{eff}^{(g)} \geq c_c$ **then**
 - i. 代入式 (6) 更新扭转角 $\tilde{\theta}$;
- (g) 训练过程中逐步衰减一致性损失权重 λ_{cons} ;
- 6. **end while**
- 7. 返回优化后的高层策略 π_h 与分组策略 $\{\pi^{(g)}\}$ 。

3 兼容性与消融实验分析

3.1 向后兼容性

本文提出的自适应耦合判别机制与 Q-Fuse 模块完全兼容原始 Topology-GRPO 框架。当 $\mathcal{C}_{eff}^{(g)} \rightarrow 0$ 或 $\tilde{\theta} \rightarrow 0$ 时, 非交换交互项被抑制, 整体框架退化为原始可交换版本 Topology-GRPO, 原有全部实验配置可直接复用。

3.2 消融实验设计

为定量评估各新增模块独立贡献, 设置如下消融对照组:

1. 基线模型: 原始 Topology-GRPO (移除 $\mathcal{C}_{eff}$ 判别与 Q-Fuse 模块);
2. 仅耦合判别模块: 启用 $\mathcal{C}_{eff}$ 阈值切换, 但全程使用经典 Concat+MLP 融合;
3. 完整模型: $\mathcal{C}_{eff}$ 自适应区间切换 + Q-Fuse 融合 + 解耦扭转角 $\tilde{\theta}$ 优化;
4. 移除非交换项: 强制 $\cos(2\pi\tilde{\theta}) \equiv 0$, 屏蔽不对称交互项;
5. 阈值对比实验: 固定临界阈值 c_c vs 可学习自适应阈值 c_c 。

声明: 虹桥大学实体尚未正式设立, 虹桥大学科技学术网络已投入运营, 作为本研究工作官方署名单位。