Topology-GRPO (Revised Version)
Adaptive q-Deformed Message Fusion via Macro Effective Duality Coupling

Lanhaijian
Global Personal Academic Platform
Hongqiao University Tech Academic Site

Declaration: The physical entity of Hongqiao University has not yet been formally established, but the *Hongqiao University Science and Technology Academic Network* is operational and serves as the official affiliation for this research work.

Abstract

This paper revises the Topology-GRPO framework by introducing an adaptive message fusion mechanism controlled by macro effective duality coupling. The graph-based group policy optimization dynamically switches between classical commutative aggregation and q-deformed non-commutative fusion according to inter-group coherent coupling strength. All technical components are independent modeling assumptions, which are validated by downstream task performance metrics.

1 Revised Topology-GRPO Framework

1.1 Three-Component Topological Advantage

The three-component topological advantage is calculated purely via reward normalization without intra-group critics, strictly adhering to the core principle of Group Relative Policy Optimization (GRPO). For each group g, the topological advantage is defined as:

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

1.2 Macro Effective Duality Coupling and Adaptive Message Fusion

Motivation Inter-group information propagation in graph-structured multi-group systems exhibits two distinct operating regimes: weak coupling with classical commutative feature fusion, and strong coherent coupling where non-commutative interaction emerges. We introduce the macro effective duality coupling $\mathcal{C}_{\text{eff}}$ to dynamically distinguish these two regimes and adaptively switch message fusion strategies.

Definition of $\mathcal{C}_{\text{eff}}^{(g)}$ Let $\tau_{\text{coh}}^{(g)}$ denote the local policy coherent steps of group g, $\tau_{\text{delay}}^{(g)}$ the average inter-group message delay steps, W_{gh} the topological coupling weight between group g and neighbor group h, and $\text{Var}(r_h)$ the reward variance of neighbor h. $\sigma^{\mathcal{N}(g)}$ is the reward standard deviation of the neighbor set from Eq. (1), and $\epsilon = 10^{-8}$ is a small numerical stabilizer to avoid division by zero. The normalized macro effective duality coupling is formulated as:

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

Critical Threshold and Operating Regimes $\mathcal{C}_c$ denotes the critical coupling threshold, which can be set as a fixed hyperparameter or optimized as a learnable parameter during training. Two operating regimes are divided by $\mathcal{C}_c$:

- $\mathcal{C}_{\text{eff}}^{(g)} < \mathcal{C}_c$: Weak coupling regime. Adopt classical Concat+MLP message fusion.
- $\mathcal{C}_{\text{eff}}^{(g)} \geq \mathcal{C}_c$: Strong coherent regime. Activate the q-deformed Q-Fuse fusion module.

Adaptive Fusion Operator Let s_g be the state of group g and m_g be the aggregated inter-group message. The adaptive fusion rule is given by:

$$\tilde{s}_g = \begin{cases} \text{Concat}\big(s_g, \sigma\big(\text{MLP}[s_g; m_g]\big) \odot m_g\big) & \mathcal{C}_{\text{eff}}^{(g)} < \mathcal{C}_c \\ \text{Q-Fuse}(s_g, m_g; \tilde{\theta}) & \mathcal{C}_{\text{eff}}^{(g)} \geq \mathcal{C}_c \end{cases} \tag{3}$$

Q-Fuse: q-Deformed Non-Commutative Fusion Let $\phi(\cdot) = W_\phi \cdot + b_\phi$ and $\psi(\cdot) = W_\psi \cdot + b_\psi$ be two learnable linear transformations. The q-deformed fusion operator is defined as:

$$\text{Q-Fuse}(s_g, m_g; \tilde{\theta}) = W_1 s_g + W_2 m_g + \cos(2\pi\tilde{\theta}) \cdot \big(\phi(s_g) \odot \psi(m_g) - \phi(m_g) \odot \psi(s_g)\big). \tag{4}$$

The non-commutative property can be verified directly by swapping input arguments:

$$\text{Q-Fuse}(m_g, s_g) - \text{Q-Fuse}(s_g, m_g) = 2\cos(2\pi\tilde{\theta})\big(\phi(m_g) \odot \psi(s_g) - \phi(s_g) \odot \psi(m_g)\big) \neq 0. \tag{5}$$

When $\tilde{\theta} \to 0$ or the system falls into the weak coupling regime, the non-commutative term vanishes, and Q-Fuse degenerates into classical linear fusion.

Adaptive Update for $\tilde{\theta}$ The twist angle $\tilde{\theta}$ is optimized independently of policy networks, and its update is only activated in the strong coherent regime:

$$\tilde{\theta}(t+1) = \tilde{\theta}(t) + \eta_\theta \cdot \mathbf{1}_{\{\mathcal{C}_{\text{eff}}^{(g)} \geq \mathcal{C}_c\}} \cdot \big(\mathcal{C}_{\text{eff}}^{(g)} - \mathcal{C}_c\big) \cdot \frac{\partial \|\text{Q-Fuse}(s_g, m_g) - \text{Q-Fuse}(m_g, s_g)\|^2}{\partial \tilde{\theta}}. \tag{6}$$

where η_θ is the learning rate for $\tilde{\theta}$. The update magnitude is proportional to the distance from the critical threshold $\mathcal{C}_c$.

1.3 Overall Objective Function

The total loss function inherits the original Topology-GRPO formulation, including clipped policy loss, inter-group consistency loss and KL divergence regularization:

$$\mathcal{L}^{\text{Topo-GRPO}} = -\mathbb{E}\left[\min\left(r_t(\theta)\hat{A}^g, \text{clip}\left(r_t, 1 \pm \epsilon\right)\hat{A}^g\right)\right] + \lambda_{\text{cons}}\mathcal{L}_{\text{cons}} + \beta D_{\text{KL}}\left(\pi_\theta \| \pi_{\text{ref}}\right). \tag{7}$$

2 Training Algorithm

The complete training pipeline for the updated Topology-GRPO is summarized in Algorithm 1.

Algorithm 1 Updated Revised Topology-GRPO Training Pipeline

Initialize high-level policy π_h, group policies $\{\pi_l^{(g)}\}$ and message encoders
Initialize global value network V_{global} for topological reward calibration
Initialize hyperparameters: $\alpha, \beta, \epsilon, \lambda_{\text{cons}}$
Initialize critical threshold $\mathcal{C}_c$ (fixed or learnable), initial twist angle $\tilde{\theta}^{(0)}$ and learning rate η_θ
while training not converged **do**
    Construct intra-group edges E_{intra} and inter-group topological edges E_{inter}
    Collect parallel interaction trajectories for all policy groups
    for each group g **do**
        Compute $\tau_{\text{coh}}^{(g)}, \tau_{\text{delay}}^{(g)}$, neighbor reward variance $\text{Var}(r_h)$ and $\sigma^{\mathcal{N}(g)}$
        Calculate macro effective duality coupling $\mathcal{C}_{\text{eff}}^{(g)}$ via Eq. (2)
        Perform graph message passing to obtain aggregated inter-group message m_g
        Compute adaptive fused state $\tilde{s}_g$ via Eq. (3)
        Calculate three-component topological advantage $\hat{A}^g$ via Eq. (1)
    end for
    Compute clipped policy loss, inter-group consistency loss $\mathcal{L}_{\text{cons}}$ and KL divergence regularization
    Update high-level policy π_h and group policies $\{\pi_l^{(g)}\}$ by minimizing $\mathcal{L}^{\text{Topo-GRPO}}$
    if $\exists\, g, \mathcal{C}_{\text{eff}}^{(g)} \geq \mathcal{C}_c$ **then**
        Update twist angle $\tilde{\theta}$ via Eq. (6)
    end if
    Gradually decay the consistency weight λ_{cons} during training
end while
return optimized high-level policy π_h and group policies $\{\pi_l^{(g)}\}$

3 Compatibility and Ablation Study

3.1 Backward Compatibility

The proposed adaptive coupling and Q-Fuse module is fully backward compatible with the original Topology-GRPO framework. When $\mathcal{C}_{\text{eff}}^{(g)} \to 0$ or $\tilde{\theta} \to 0$, the non-commutative interaction term is suppressed, and the entire framework degenerates to the original commutative Topology-GRPO. All existing experimental settings can be reused directly.

3.2 Ablation Experiment Design

To quantitatively evaluate the independent contribution of each proposed component, we design the following ablation groups:

1. Baseline: Original Topology-GRPO ($\mathcal{C}_{\text{eff}}$ and Q-Fuse modules removed)
2. $\mathcal{C}_{\text{eff}}$ only: Adaptive coupling threshold enabled, always use classical Concat+MLP fusion
3. Full model: $\mathcal{C}_{\text{eff}}$ adaptive switching + Q-Fuse + decoupled $\tilde{\theta}$ optimization

4. Ablate non-commutativity: Set $\cos(2\pi\tilde{\theta}) \equiv 0$ to disable asymmetric interaction terms
5. Fixed $\mathcal{C}_c$ vs. learnable $\mathcal{C}_c$: Compare fixed threshold and data-adaptive learned threshold settings