ARES-CORE と Dueling DQN の数理的比較

1　序論

Dueling DQN と ARES-CORE は, 共に Q 学習に基づいた強化学習アルゴリズムだが, 設計の違いによりそれぞれの強みが異なる. Dueling DQN は, 状態価値 $V(s)$ と行動アドバンテージ $A(s,a)$ の分離により, 効率的な行動選択を可能にしている. 一方, ARES-CORE は動的な解像度適応を組み込むことで, より効率的な状態空間の探索と迅速な収束を目指す.
この数理的解析は, ARES-CORE の理論的な優位性を明確に示し, Dueling DQN と比較した際の特性を詳述するものである.

2　Dueling DQN の数理モデル

Dueling DQN の Q 値は, 状態価値 $V(s)$ と行動アドバンテージ $A(s,a)$ を用いて次のように定義される.:

$$Q(s,a;\theta,\alpha,\beta) = V(s;\alpha,\theta) + \left(A(s,a;\beta,\theta) - \frac{1}{|\mathcal{A}|} \sum_{a'} A(s,a';\beta,\theta) \right)$$

ここで,

- $V(s)$ は状態価値関数で, 状態 s における価値を評価する.
- $A(s,a)$ は行動アドバンテージで, 状態 s で特定の行動 a を取ることの優位性を示す.
- θ, α, β は畳み込みネットワークのパラメータである.

Dueling DQN はこのアドバンテージ項 $A(s,a)$ により, 効率的な行動選択を行うが, 状態の観測が固定解像度に依存しているため, 高次元の状態空間では探索が非効率になる場合がある.

3　ARES-CORE の数理モデル

ARES-CORE では, 解像度適応機能を組み込み, 探索段階での解像度を動的に変更することにより, 広域探索と詳細探索のバランスを取っている. このモデルでは, 状態 s の解像度レベル ρ に依存する Q 値が次のように定義される.:

$$Q_\rho(s,a;\theta,\alpha,\beta) = V_\rho(s;\alpha,\theta) + \left(A_\rho(s,a;\beta,\theta) - \frac{1}{|\mathcal{A}|} \sum_{a'} A_\rho(s,a';\beta,\theta) \right)$$

ここで, 解像度レベル ρ は, エピソードの進行に伴って次のように動的に変化する.:

$$\rho = \min(\rho_{\max}, \rho_{\text{init}} + k \cdot t)$$

- ρ_{init} は初期解像度レベル.
- $\rho_{\max}$ は最大解像度レベル.
- k は解像度の上昇速度を制御する係数.
- t はエピソード内のステップ数.

この動的解像度により, ARES-CORE はエピソードの初期には低解像度で広範囲を探索し, 徐々に解像度を上げて詳細な探索に移行する.

4　生涯獲得報酬の比較

生涯獲得報酬 G は, 各エージェントが得る総報酬の期待値を表し, 次のように定義される.

Dueling DQN の場合

$$G_{\text{Dueling DQN}} = \sum_{t=0}^{T} \gamma^t R(s_t, a_t)$$

ARES-CORE の場合

$$G_{\text{ARES-CORE}} = \sum_{t=0}^{T} \gamma^t R(s_t, a_t, \rho_t)$$

ここで, ρ_t は時刻 t における解像度レベルであり, 解像度の動的適応により, Dueling DQN に比べて非最適行動が減少し, より高い生涯獲得報酬が期待される.

5　収束速度の理論的解析

収束速度 $C(\theta)$ は, エピソード数 T として定義され, 次の条件を満たす最小のエピソード数を意味する.:

$$C(\theta) = \inf\{T \in \mathbb{N} \mid \|\theta_T - \theta^*\| < \epsilon\}$$

解像度の影響　ARES-CORE における動的解像度の導入により, エピソード数 T に依存する収束が加速するため, 以下の関係が予測される.:

$$\mathbb{E}_{\pi_{\text{adaptive}}}\left[\|\theta_T - \theta^*\|\right] < \mathbb{E}_{\pi_{\text{Dueling DQN}}}\left[\|\theta_T - \theta^*\|\right]$$

これにより, ARES-CORE は Dueling DQN よりも速く収束しやすい構造を持つ.

6　行動回数の急減とメモリ効率

ARES-CORE では, 解像度の動的調整により, 探索に要する行動回数が減少し, 必要なメモリ消費も抑えられる. これにより, メモリ効率も向上する. 行動回数の急減は以下の期待値で定義できる.

Dueling DQN の場合

$$\mathbb{E}_{\text{Dueling DQN}}[A(s,a)]$$

ARES-CORE の場合

$$\mathbb{E}_{\text{ARES-CORE}}[A(s,a;\rho)] = \frac{1}{\rho}\mathbb{E}_{\text{Dueling DQN}}[A(s,a)]$$

解像度 ρ によって行動回数が減少するため, 行動選択にかかるコストも低くなり, メモリ効率が向上することが期待される.

7　結論

ここまでの数学的な解析から, ARES-CORE は Dueling DQN に対して以下の利点を持つと結論づけられる.:

1. **収束速度の向上**：動的解像度の導入により, 広域探索から詳細探索へと効率的に移行し, 迅速な収束を達成するものと考えられる.

2. **生涯獲得報酬の増加**：動的解像度により非最適行動が抑制され, より高い報酬が期待される.
3. **行動回数の急減**：解像度に応じて行動選択の効率が向上し, メモリ効率の改善が期待される.