Parallel-MAT によるマルチエージェント強化学習の協調性能の改善と計算の効率化

EP21014　牛古直希　　指導教授：山下 隆義

1. はじめに

交通環境における信号制御や物流における配送ロボットの制御など，実世界には複数のエージェントが協力して取り組む協調タスクが存在する．このタスクに対し，複数のエージェントが共通の目標を達成するための協調行動を深層強化学習により獲得するマルチエージェント強化学習 (MARL) が注目されている，その代表的な手法に Multi-Agent Transoformer (MAT) [1] がある．MAT では，各エージェントの行動を逐次的に決定することで，前エージェントを考慮した協調行動を獲得する．しかし，MAT では各エージェントの行動を順番に決定するため，順番に依存した行動決定となる．そのためエージェント全体を考慮した意思決定ができず，計算効率も低い．そこで本研究では，並列意思決定および並列演算が可能である Parallel-MAT を提案し，高い協調性能を持った計算効率の高い MARL 手法の実現を目指す．

2. Multi-Agent Transformer

MAT は Transformer [2] の Encoder-Decoder 構造を用いることで，複数エージェントの意思決定をシーケンスモデルとして扱う MARL の手法である．MAT のモデル構造を図 1 に示す．MAT は Transformer の入力に対する長期依存関係を捉える特性を活用している．しかし，MAT によって獲得した協調行動は，エージェントの推論順が大きく影響し，特定のエージェント間で協調行動が不可能である．また，逐次的に推論を行うため，計算効率の低さも課題である．

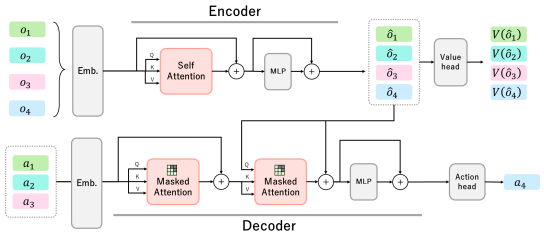

図 1：MAT のモデル構造

3. Parallel-MAT

MAT の問題点であるエージェントの推論順に対する依存性と計算効率の低さに着目する．この問題点に対し，Decoder に並列意思決定と並列演算を導入した Parallel-MAT を提案する．Parallel-MAT のモデル構造を図 2 に示す．

3.1. 並列意思決定

全エージェント間での協調行動を目的とし，Transformer Decoder において，自エージェントの状態を Query，他エージェントの状態を Key，Value とした Cross-Attention を計算する．これにより，全エージェントの並列的で相互に協調した意思決定が可能となる．

3.2. 並列演算

並列意思決定の Cross-Attention 計算に，対角成分をマスクする Masked-Attention を利用し，あるエージェントから見た他のエージェントの Attention を全て一度に計算可能とする．これにより意思決定に対する計算効率の向上を図る．

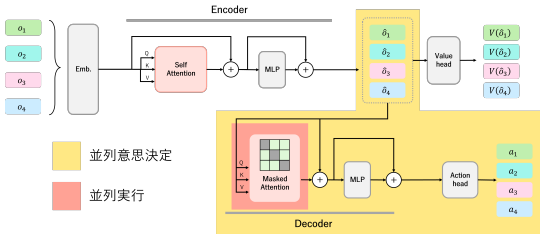

図 2：Parallel-MAT のモデル構造

4. 評価実験

Parallel-MAT の有用性を示すため，MuJoCo を用いた評価実験を行う．本実験では，MuJoCo の以下に示す 4 環境にて評価を行った．() はエージェント数を示す．

HalfCheetah (6). 制御対象はチーターの半身を模した 2 足ロボットで，目的は水平方向への移動である．

Ant (4). 制御対象は昆虫を模した 4 足ロボットで，目的は前進である．

HumanoidStandup (2). 制御対象は人体を模した 2 足ロボットで，目的は直立である．

Walker2D (6). 制御対象は 2 足歩行ロボットを制御し，目的は転倒せず水平方向への移動である．

4.1. 性能に対する評価

それぞれの環境において，学習済みモデルを用いた評価を 5 回行い，その平均報酬で比較する．平均報酬を表 1 に示す．

表 1：シナリオごとの平均報酬

シナリオ	MAT	Parallel-MAT
HalfCheetah	9321 ± 137	9529 ± 75
Ant	2114 ± 561	4632 ± 129
HumanoidStandup	130683 ± 2070	127193 ± 4175
Walker2D	378 ± 4	609 ± 40

HalfCheetah と HumanoidStandup では報酬値が同等，Ant と Walker2D では Parallel-MAT の報酬値の高い結果となった．報酬値が同等のシナリオは易度が低く，どちらのモデルを用いても最適な解法に辿り着くことができたと考えられる．提案手法の報酬値が高いシナリオでは，全エージェントに対して協調行動の行える Parallel-MAT の特性により多くの報酬が獲得できたと考えられる．

4.2. 推論速度に対する評価

MAT，並列意思決定のみを導入した Parallel-MAT，並列意思決定と並列演算を導入した Parallel-MAT の 3 モデルに対して 500step 分の推論時間を計測した．計測は 10 回行い平均値とする．平均推論時間を表 2 に示す．() 内は同エージェント数の MAT 比の計算時間の割合である．この結果から Parallel-MAT は MAT と比べ推論時間をおよそ 31%に短縮していることを確認した．

表 2：モデルごとの平均推論時間 [秒] と高速化率

並列意思決定	並列演算	エージェント数			
		5	10	50	100
		4.35 (100%)	8.22 (100%)	41.84 (100%)	86.12 (100%)
✓		2.57 (59.1%)	4.53 (55.1%)	21.11 (50.5%)	41.61 (48.3%)
✓	✓	1.94 (44.6%)	3.28 (39.9%)	13.69 (32.7%)	26.87 (31.2%)

5. おわりに

本研究では，並列実行型マルチエージェント強化学習手法である Parallel-MAT を提案した．評価実験では，提案手法による高い協調性を持った行動の獲得及び推論速度の改善が確認できた．今後は，他マルチエージェント環境での実験やモデル構造の改善に取り組む．

参考文献

[1] Wen *et al.*, “Multi-Agent Reinforcement Learning is a Sequence Modeling Problem”, *arXiv preprint arXiv:2205.14953*, 2022.

[2] Vaswani *et al.*, “Attention Is All You Need”, *NeurIPS, 2017.*