マルチエージェント強化学習におけるハンガリアン法による報酬割当

EP20105 三島広暉　　指導教授：山下隆義

1. はじめに

実世界には，商品運搬におけるロボット群の制御のように複数のエージェントが協力し合う協調タスクが存在する．これらの協調タスクを解く手法としてマルチエージェント強化学習がある．本手法は，他エージェントの情報を考慮して自身の行動を決定する方策を学習する．このとき，各エージェントが目的地に移動するタスクの場合，どのエージェントがどの目的地に向けて行動するか決める必要がある．この割り当てが正しくないと高い報酬を得られない．そこで本研究では，ハンガリアン法 [1] を用いた割り当てに基づく報酬算出法を提案する．ハンガリアン法を用いることで報酬算出時のエージェントの割り当てを求めることができ，報酬割当問題を解決可能となる．Simple Spread 環境を用いたハンガリアン法の有無による評価から，本手法の有効性を調査する．

2. マルチエージェント強化学習における問題点

マルチエージェント強化学習とは，複数のエージェントの方策を同時に学習する手法である．各エージェントが独自の方策に従い行動を決定し，決定した行動を全エージェント共有の評価ネットワークにより評価することにより，他エージェントを考慮した協調行動を獲得できる．一方で，エージェント数が増えると学習時の報酬割当が困難となり，選択した行動に対する正しい評価ができない．複数のエージェントが異なるゴール（ランドマーク）を目指すタスクの場合，ランドマークを基準としてエージェントを割り当てるのが一般的である．しかし，この割り当てでは複数のランドマークに対して 1 つのエージェントが割り当てられることがある．そのため，報酬算出時に正しい割り当てを行うことは最適な協調行動の獲得に重要である．

3. 提案手法

学習時の報酬割当問題にハンガリアン法を導入した報酬算出法を提案する．ハンガリアン法による割り当ての手順を図 1 に示す．ここで，各色の丸はランドマーク，数字はエージェントを示す．各 Step を以下に示す．

- Step 1： ランドマークに基づくコスト関数からコスト行列を作成.
- Step 2： 行方向および列方向ごとの最小値を各要素から減算
- Step 3： 行方向および列方向ごとに要素が 0 であるペアを割り当て候補とし，割り当て候補から完全マッチング可能か判定．全割り当てが可能な場合 Step5 へ遷移．ここで完全マッチングとは，割り当て先が重複しない割り当てのことである．
- Step 4： 割り当て候補以外の最小値を用いた加減算による割り当て処理後，Step3 へ遷移.
- Step 5： 割り当てに基づき報酬計算

報酬算出時にハンガリアン法による割り当てを行うことで，エージェントの割り当てを決定し報酬を算出する．コスト関数にはエージェントとランドマーク間のユークリッド距離を用いる．図 1 より，エージェントとランドマークが 1 対 1 対応した割り当てが求められ，環境内の全てのエージェントとランドマークの情報を用いて報酬算出可能であることがわかる．

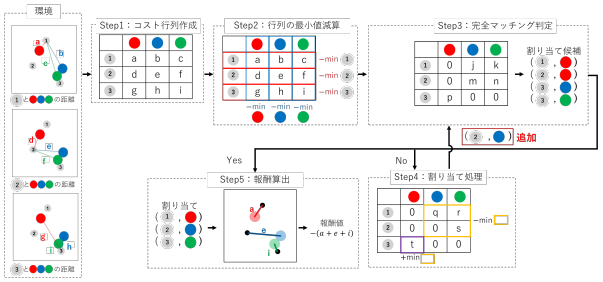

図 1: ハンガリアン法による割り当て手順

4. 評価実験

本報酬算出法の有効性を確認するため，Simple Spread を用いて評価する．

4.1. 実験環境

Simple Spread は複数のエージェントとランドマークをランダムな位置に生成し，各エージェントが衝突を避けながらランドマークへの到達を目指すタスクである．Simple Spread における報酬設計は，ランドマークと各エージェント間の最小ユークリッド距離の負数，エージェント同士の衝突に対する負数 (−10) の和である．ここで，最小ユークリッド距離はランドマークを基準とし算出する．

4.2. 実験概要

本実験では，学習アルゴリズムとネットワーク構造は板谷らの手法 [2] を用い，報酬算出時のハンガリアン法の有無による比較を行う．比較対象は，ランドマーク基準による報酬（w/o hungarian）と提案手法であるハンガリアン法に基づく報酬 (w/ hunagarian) である．学習条件は，エージェント及びランドマーク数を 3 個から 5 個までランダムに変え，1 エピソード間の step 数を 40 とする．エージェント数が 3 または 4 では 500,000 エピソード，エージェント数 5 では 800,000 エピソード学習する．エピソード終了時における各ランドマークとエージェント間の平均ユークリッド距離により評価する．平均ユークリッド距離の値が低いほどエージェントがランドマークに到達できていることを意味する．

4.3. 実験結果

各報酬設計で学習したモデルを用いて 50 エピソード分を評価した．エピソード終了時における平均ユークリッド距離を表 1 に示す．ここで，表 1 の値はシード値を 5 回変更して平均をとったものである．表 1 より，エージェント数 4 と 5 ではハンガリアン法を用いると，ランドマークにより近づけたことがわかる．

表 1: 50 エピソード評価時の平均ユークリッド距離

ハンガリアン法	エージェント数		
	3	4	5
w/o	**0.64**	0.97	1.07
w/	0.73	**0.77**	**0.99**

エージェント数 4 のエピソード終了時における定性的比較を図 2 に示す．図 2 より提案手法は，エージェントがランドマークにより近づいていることがわかる．これらの結果から，エージェント数 4 と 5 の環境はエージェント数 3 の環境と比べ，エージェントとランドマークの配置が複雑となり，報酬算出時の割当が難しくなる．そのため，ハンガリアン法による割り当てが有効であったと考えられる．

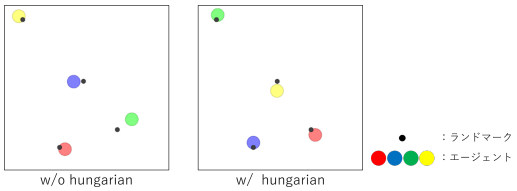

図 2: ハンガリアン法の有無における定性的比較

5. おわりに

本研究では，マルチエージェント強化学習における学習時の報酬割当問題に対しハンガリアン法に基づく報酬算出法を提案した．Simple Spread による評価から，タスクが複雑であるほど，ハンガリアン法が有効であることを確認した．今後は，他環境での本手法による有効性を検証する．

参考文献

[1] H. W. Kuhn, “The Hungarian method for the assignment problem”, Naval Research Logistics Quarterly, 1955.

[2] H. Itaya, *et al.*, “Visual Explanation for Cooperative Behavior in Multi-Agent Reinforcement Learning”, In IJCNN, 2023.