

[研究助成 (A)]

言語指示による群ロボット協調輸送のための深層強化学習技術の開発

2251015

研究代表者
(助成金受領者)奈良先端科学技術大学院大学
先端科学技術研究科

准教授

柴 田 一 騎

[研究の目的]

群ロボット協調輸送は複数の荷物を同時に運べることや、重いものも協調して運べるため、システムとして実現できれば効率が良いと考えられる。現行技術は、荷物やゴールの位置を与えて群ロボットを制御する。しかし、実運用を考えると、この方法は現実的ではない。例として、人が複数台のロボットに指示を与え、荷物をゴールへ移動することを考える。この場合、人は荷物やゴールの正確な位置を知ることとはできないため、現行技術を適用できない。人の使いやすさを考えると、「荷物をゴールへ運んで」のような言語指示に基づくシステムが望ましい。

本研究の目的は、言語指示に基づく群ロボット協調輸送のための学習技術の確立である。この技術を確立できれば、物流、建設、農業などの多岐にわたって人が直感的に群ロボットを操作可能となり、人と調和した群ロボットシステムの創出につながることが期待される。

[研究の内容]

本研究では、図1のように各ロボットが搭載カメラの観測に基づいて言語指示を解釈する、視覚・言語誘導型の群ロボット協調輸送に取り組んだ。この分散的な設定における主要な課題は、ロボット間の知覚の不整合である。すなわち、異なるロボットが同一の指示や環境に対して、異なる荷物やゴールを対応付けてしまう問

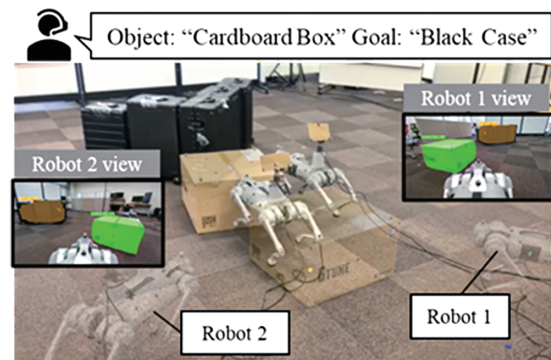


図1 言語指示による群ロボット協調輸送

題である。このような不整合はロボット間の協調行動を損ない、輸送性能を低下させる。特に、この問題は、各ロボットが自身の知覚および意思決定モジュールを持つ群ロボットシステムにおいて顕著となる。

この問題を緩和するために、本研究ではリーダー・フォロワー構造を採用する。この構造では、一方のロボットがリーダー、もう一方がフォロワーとして動作する。タスクに関連する指示の解釈をリーダーに集約することで、ロボット間の知覚不整合を軽減するとともに、タスクの解釈を単一のロボットにのみ要求すればよいため、実装コストの削減にもつながる。

本研究では、リーダー・フォロワー構造を獲得するため、群ロボットの学習 AI であるマルチエージェント強化学習 (MARL) を採用する。MARL は、群ロボットが試行錯誤で経験データを収集し、学習によって他者との協調行動ルールを自動獲得できる。

しかしながら、リーダー・フォロワー構造は

試行錯誤による学習から自発的に獲得されることは稀である。独立かつ対称な方策構造を前提とする従来の MARL は不安定な振る舞いに収束しやすく、一貫したリーダー・フォロワー構造を形成できない。

本研究では、一貫したリーダー・フォロワー構造を学習により獲得するための方法論の検討を進めた。具体的な実現方法として、以下の3つの要素によって構成される方法論を導出した。

- (1) 非対称な方策モデル：リーダーにはタスクの情報を与える一方で、フォロワーには局所観測のみを与える非対称な方策モデルを導入する。
- (2) 相互情報量に基づく目的関数：フォロワーが自身の局所観測からリーダーの行動を予測するように相互情報量に基づく目的関数を導入する。
- (3) 視覚・言語誘導型の群ロボット学習の枠組み：画像言語モデル (VLM) の推論時間に起因する訓練時間の増大を回避するため、訓練時には VLM を用いず、実行時のみ VLM を用いる学習の枠組みを導入する。

まず、一貫したリーダー・フォロワー構造を促進するために、図2に示す非対称な方策モデルを設計した。具体的には、リーダーとフォロワーの観測入力に非対称性を導入し、役割分担を誘導する。リーダーは、ゴールおよび自身の局所観測（荷物や他のロボットを含む）を入力とし、方策 π に基づいて行動を決定する。これ

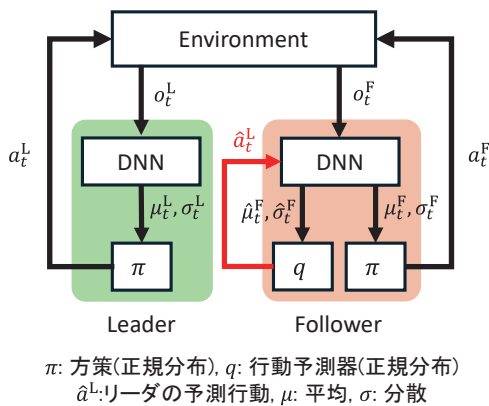


図2 非対称な方策モデル

により、リーダーはタスクに関する情報を利用して意思決定を行う。

フォロワーはゴール情報を用いず、自身の局所観測のみを入力として方策 π に基づいて行動を決定する。さらにフォロワーは、リーダーの行動を予測する補助関数 q を内部に持つ。この設計により、フォロワーは局所観測からリーダーの行動を予測しつつ行動し、一貫したリーダー・フォロワー構造が形成される。

次に、フォロワーが自身の局所観測からリーダーの行動を予測するように相互情報量に基づく目的関数を導入した。非対称な方策モデルでは、フォロワーは目標位置の情報にアクセスできないため、一貫した協調行動を学習できない可能性がある。

そこで、フォロワーの観測 o_t^F とリーダーの行動 a_t^L との間の相互情報量 $I(a_t^L; o_t^F) = H(a_t^L) - H(a_t^L | o_t^F)$ を最大化することで、フォロワーがリーダーの行動を予測するように学習を促す。リーダーとフォロワーの方策は同時に学習されるため、リーダーの行動エントロピー $H(a_t^L)$ は学習中に変化する。そのため、相互情報量を厳密に最適化する代わりに、CE 損失の更新時には $H(a_t^L)$ を一定とみなし、条件付きエントロピー $H(a_t^L | o_t^F)$ に基づく近似目的関数を最適化する。しかし、この項には未知の条件付き分布 $p(a_t^L | o_t^F)$ が含まれるため、解析的に解くことは困難である。

そこで、変分推論を用いてリーダーの行動分布を近似する。任意の近似分布 $q(a_t^L | o_t^F)$ に対して、KL ダイバージェンスの非負性

$$D_{KL}(p(a_t^L | o_t^F) \parallel q(a_t^L | o_t^F)) \geq 0 \quad (1)$$

が成立する。ここで、近似分布 $q(a_t^L | o_t^F)$ は、フォロワーのディープニューラルネットワーク (DNN) の出力によりガウス分布としてパラメータ化される。式 (1) を展開すると、条件付きエントロピーは

$$H(a_t^L | o_t^F) \leq -\mathbb{E}[\log q(a_t^L | o_t^F)] \quad (2)$$

となる。したがって、式 (2) の上界を最小化することは、対数尤度を最大化することに対応する。最終的に、目的関数は以下となる：

$$J = J_{\text{RL}} + \lambda \mathbb{E}[\log q(a_t^L | o_t^F)] \quad (3)$$

ここで、 J_{RL} は元の MARL の目的関数、 $\lambda > 0$ は重み係数である。式 (3) の目的関数により、タスク性能の向上と、フォロワーによるリーダーの行動予測を両立するように学習される。

最後に、視覚・言語誘導型の群ロボット学習の枠組みを導入した。VLM を用いた MARL では、多様な荷物およびゴールに対して方策を学習すると、訓練時間が大幅に増大する。また、訓練ループに VLM 推論を含めると、計算コストも増加する。

この問題に対処するため、本研究では VLM-free training and VLM-based execution の枠組みを採用した。具体的には、訓練時には物体およびゴールの位置から構成される状態ベクトルを入力とし、VLM を用いずに方策を学習する。一方で、実行時には VLM により推定された物体およびゴールの位置を入力としてロボットの方策を実行する。

各ロボットは図 3 のように搭載カメラの RGB 画像とテキスト情報から VLM を用いて、物体およびゴールの位置を推定する。具体的には、CLIPSeg によるテキストと画像領域の対応付けにより物体の 2 次元位置を推定し、推定位置、深度情報およびカメラの内部・外部パラメータを用いて 3 次元座標へ変換する。同様の処理を

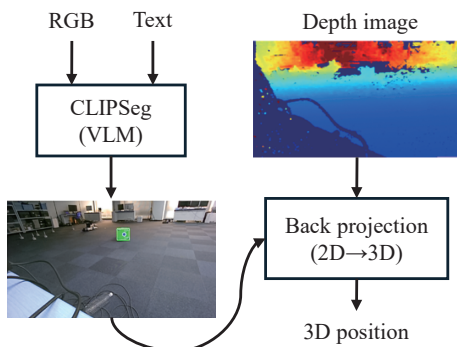


図 3 VLM による推論

ゴールにも適用し、得られた 3 次元座標をリーダー・フォロワー方策の入力として用いる。

[研究の成果]

提案手法の有効性を、シミュレーションおよび実機実験により評価した。ベースラインとして、MARL の手法の一つである MAPPO と比較した。

まず、フォロワーにゴール情報を与えない条件における協調行動の成立性を確認するため、単一ゴール環境において評価を行った。単一ゴール環境では MAPPO でも安定してタスクが達成可能であるため、本実験ではフォロワーがゴール情報を持たない場合の協調行動の成立性を確認することを目的とする。

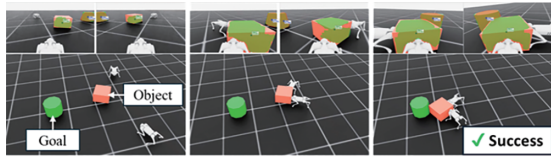
定量評価の結果を表 1 に示す。提案手法は、両ロボットにゴール情報を与えた MAPPO と同程度の成功率を達成した。また、タスク終了時の物体とゴール間距離においても同程度の性能を示した。これにより、フォロワーがゴール情報を持たない場合でも、ゴール情報を有する場合と同程度の性能で協調輸送が可能であることを確認した。さらに、フォロワーによるリーダー行動の予測の効果を確認するため、相互情報量に基づく目的関数を除いたアブレーション (Ours w/o MI) と比較した。その結果、提案手法は Ours w/o MI を上回り、目的関数の有効性を確認した。

ロボットの初期位置とゴールが同一の条件下でのシミュレーションの一例を図 4 に示す。図 4 より、MAPPO と提案手法はいずれも荷物をゴールまで輸送できていることが確認できる。

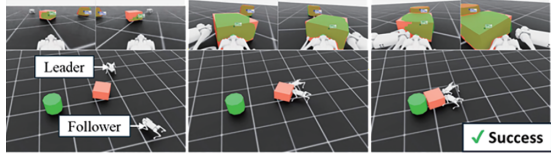
以上より、提案手法はフォロワーにゴール情

表 1 単一ゴール環境での性能比較

Method	Success rate	Distance [m]
MAPPO	0.71 ± 0.21	1.02 ± 0.26
Ours w/o MI	0.55 ± 0.31	0.99 ± 0.28
Ours	0.76 ± 0.17	0.80 ± 0.17

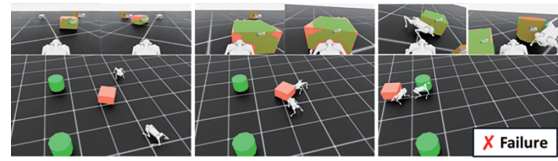


(a) MAPPO



(b) Ours

図4 単一ゴール環境でのシミュレーションの一例



(a) MAPPO



(b) Ours

図5 複数ゴール環境でのシミュレーションの一例

表2 複数ゴール環境での性能比較

Method	Success rate	Distance [m]
MAPPO	0.29 ± 0.08	1.31 ± 0.12
Ours w/o MI	0.34 ± 0.14	1.19 ± 0.32
Ours	0.76 ± 0.10	0.82 ± 0.09

報を与えない条件においても、一貫した協調行動を維持できることを確認した。

次に、知覚の不整合が生じる条件下での提案手法の有効性を確認するため、同一のゴールランドマークが2個存在する環境で評価を行った。

定量評価の結果を表2に示す。提案手法は、MAPPOやOurs w/o MIと比較して高い成功率および低い目標位置誤差を達成した。これは、同一のゴールランドマークが複数存在する場合でも、ロボット間で一貫した協調行動が可能であることを示している。

ロボットの初期位置とゴールが同一の条件下でのシミュレーションの一例を図5に示す。図5より、MAPPOでは2台のロボットが異なるゴールランドマークを推定し、荷物をゴールへ輸送できなかった。一方で、提案手法では一貫した協調行動により、荷物をゴールへ輸送できていることが確認できる。

以上より、複数ゴール環境において知覚の不整合が生じる条件下でも、一貫した協調行動が可能であることを確認した。

最後に、提案手法の実環境における有効性お



図6 5つの実環境シナリオ

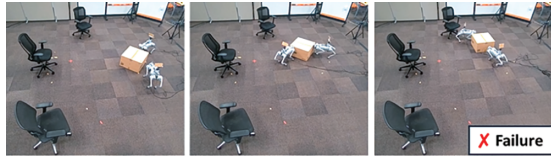
よび汎化性を検証するため、2台の四脚ロボット Unitree Go1 を用いた実験を行った。各ロボットは Intel RealSense D455 を搭載し、ロボット間の相対位置をモーションキャプチャにより計測した。ロボット間の相対位置は視覚情報から推定できるが、荷物を輸送しつつ他ロボットを視野内に保つための視点制御が必要となる。本実験ではこの問題を切り分けるために、他のロボットの位置をモーションキャプチャにより取得した。荷物は形状・質量・摩擦係数が固定された単一の物体を用いた。

実験では、異なるゴールランドマークを設定した5つのシナリオを用意し、各シナリオにつき初期条件を変えて5試行ずつ評価を行った。

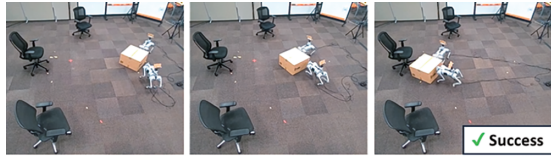
定量評価の結果を表3に示す。知覚の不整合が生じにくい Different-Height-Boxes および Single-Fridge では、MAPPO と提案手法のい

表 3 実環境における成功率

Task Scenario	MAPPO	Ours
Multi-Identical-Chairs	2/5	5/5
Multi-Colored-Chairs	1/5	5/5
Different-Height-Boxes	3/5	5/5
Single-Fridge	3/5	4/5
Clustered-Cases	2/5	5/5



(a) MAPPO



(b) Ours

図 7 Multi-Identical-Chairs での実機実験の一例

れも高い成功率を示した。一方で、複数の類似物体が存在し知覚の不整合が生じやすいシナリオでは、提案手法のみが高い成功率を達成した。

同一の初期条件下における Multi-Identical-Chairs の実機実験の一例を図 7 に示す。図 7 より、MAPPO では荷物をゴールへ輸送できないのに対し、提案手法では荷物をゴールへ輸送できた。なお、他のシナリオの実験については、成果の URL を参照されたい。

以上より、提案手法は実環境においても有効であり、異なるゴールに対する汎化性を確認した。

[まとめ]

本研究では、分散知覚下における視覚・言語情報に基づくマルチロボット協調輸送のための学習手法を提案した。提案手法では、非対称な方策設計および相互情報量に基づく学習により、リーダー・フォロワーの役割分担を安定的に獲得することを可能とした。シミュレーションの結果、提案手法は知覚の不整合が生じる条件下においても高い輸送性能を達成した。また、実機実験においても複数シナリオで安定した協調輸送を実現し、異なるゴールランドマークに対する汎化性を有することを確認した。

一方で、本研究には以下のような制約がある。

- (1) 単一の物体条件で評価しており、形状や物性の変化に対する性能低下の可能性がある。
- (2) 他のロボットの位置をモーションキャプチャにより取得しており、完全分散化には至っていない。
- (3) ロボット数の増加に対するスケーラビリティが未検証である。

今後は、多様な物体条件に対する汎化性の向上、視覚に基づく推定と視点制御の統合、およびロボット数に依存しない分散型強化学習手法との統合によるスケーラビリティの検証を進める予定である。

[成果の発表、論文など]

J. Y. Despature, K. Shibata, T. Matsubara, CoLF: Learning Consistent Leader-Follower Policies for Vision-Language-Guided Multi-Robot Cooperative Transport, arXiv preprint arXiv:2602.07776, 2026 (国際論文誌に投稿中)

実験動画 : <https://www.youtube.com/watch?v=CrP99ikjCHo>

[謝辞]

本研究のご支援を頂きました公益財団法人立石科学技術振興財団に厚く御礼申し上げます。