7W-01

安全なマルチエージェント強化学習によるドローン配送

加地　正拡†　　林　冬惠‡

岡山大学　工学部情報系学科†　　岡山大学　学術研究院環境生命自然科学学域‡

1.　はじめに

近年，ドローンの物流分野への応用についての研究が進められている．多数のドローンによる配送を想定する場合，衝突を回避するためには，各々を自由に飛行させるのではなく，飛行計画や航空管制を考える必要がある．[1]では，複数のドローンでの配送の経路計画問題が Drone Routing Problem（DRP）と定義されている．DRP は，限られた空路を衝突せずに，できるだけ効率的に移動することを目的とする．また，オンデマンド配送による秒単位の経路計画を想定するため，DRP をマルチエージェント強化学習によって解決することが検討されている．しかし，マルチエージェント強化学習では，学習中や学習済みモデルにおいて，エージェント間の衝突を防ぐことが困難であり，安全が保障されない．そのため，本研究では，マルチエージェント強化学習を用いる際に安全を確保するための設計について提案する．また，複数のマップでの実験によって，安全性と学習効果について提案手法の有効性を示す．

2.　Drone Routing Problem

複数のドローンが，限定された道を衝突せずにできるだけ効率よく移動することを考える問題が Drone Routing Problem (DRP)として，[1]にて定義されている．DRP ではドローンの移動可能な領域は図 1 のような無向グラフで表現される．ドローンはノードからノードへの移動を繰り返し，目的地のノードを目指す．DRP は，すべてのドローンの移動コストの総和を最小化することを目的としている．

本研究では，複数のドローンを複数のエージェントとみなし，DRP をマルチエージェント強化学習で解決することを考える．エージェントの移動時に−1，移動しない時に−10，ゴール時に 100，衝突時に−10 の報酬を与える．また，学習時にエピソードの上限ステップ数を設けている．

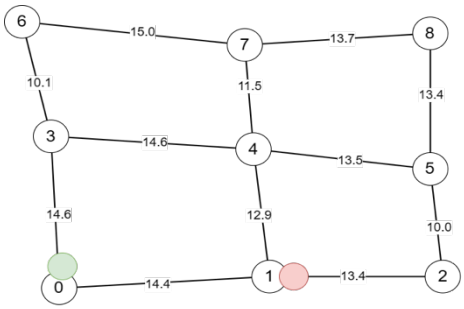

図 1：ドローンの移動可能な領域

DRP をマルチエージェント強化学習で解決しようとする場合，試行錯誤によって学習するため衝突が発生し，エージェント間の安全が保障されないという問題がある．そこで，本研究では安全なマルチエージェント強化学習の設計を提案する．

3.　安全のための設計

DRP において安全性を確保するために，状態表現への視野の追加と危険な行動の制限について提案する．視野の追加の目的は，衝突の危険性に関する潜在的な情報をエージェントに与えることである．危険な行動の制限の目的は，衝突が起こる原因となる行動を防ぐことである．

3.1.　視野の設計

状態表現とは，各エージェントの現在地，目的地を表しており，それに視野を追加する．視野は，現在地のノードと隣接するノード付近に他のエージェントがいる場合に，その情報をエージェントに与える．例えば，図 1 において，赤いエージェントの視野の範囲はノード 0，1，2，4 であり，赤いエージェントはノード 0 の近くに他のエージェントが存在することを感知する．

このように，他の近いエージェントの位置が分かることで，衝突を避けるように学習させることができる．

3.2.　危険な行動の制限

[2]では，マルチエージェント強化学習で安全を確保するために，危険な行動を制限する手法

Safe Multi-Agent Reinforcement Learning for Drone Routing Problems

Masahiro KAJI†, Donghui LIN‡

†Department of Information Technology, Faculty of Engineering, Okayama University

‡Faculty of Environmental, Life, Natural Science and Technology, Okayama University

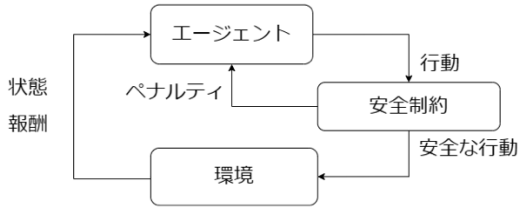

図 2：危険な行動の制限の流れ

が提案されている．本研究ではこの手法に着目し，DRP へ反映させる．ここでは危険な行動を次のように定義する．

・他のエージェントと同じエッジを通る行動
・他のエージェントと，次に目指すノードが同じである行動

図 2 に危険な行動を制限する流れを示している．方策によって得られた行動を安全制約によって確認し，危険な行動であれば安全な行動に変更する．その際，危険な行動であることを学習させるために，エージェントに負の報酬をペナルティとして与える．

安全制約では，例えば，図 1 において赤いエージェントがノード 1 にとどまっているときに，緑のエージェントがノード 1 に向かう行動が危険な行動と判断される．そして，緑のエージェントをノード 0 にとどまらせることで安全な行動にする．

4. 評価

提案手法について，実験は図 1 のような形状のマップで行う．図 1 はノード数が 3×3 のマップであるが，実験ではエージェント数を 4 として，ノード数が 5×4，8×5，10×8 のマップを使用する．また，各エージェントがゴールするまでの時間の和をコストとして比較する．衝突や設定している制限時間を超えることによって全てのエージェントがゴールできなかった場合，コストは（エージェント数×制限時間）とする．

視野を導入した手法と安全制約を導入した手法についての 4 エージェント，8×5 のマップでの実験結果が図 3，4 である．視野と安全制約の導入は共に，衝突率の減少に貢献することがわかる．また，コストを比較した結果，視野，安全制約の導入により学習効果は向上するといえる．

また，表 1 は様々なマップでの実験結果をまとめている．ここでの従来の手法は，視野は導入しているが，安全制約を導入していない手法を表し，提案手法は視野と安全制約を導入していることを表す．表 1 から，様々な広さのマップでも安全制約の導入によって，衝突率が 0 になり，コストも改善されているとわかる．

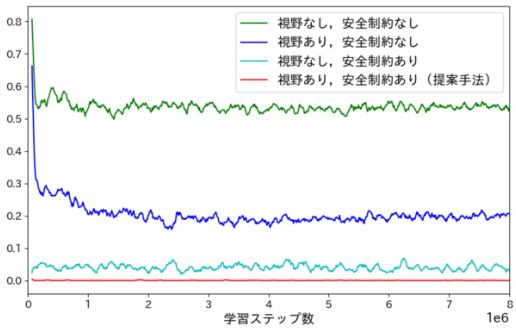

図 3：衝突率

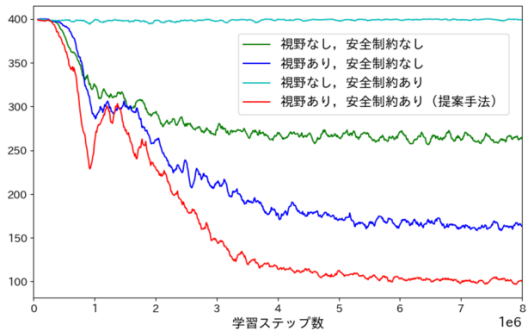

図 4：コスト

表 1：様々なマップでの安全制約の比較

	衝突率		コスト	
	従来	提案	従来	提案
5×4	0.28	0.00	188	93
8×5	0.20	0.00	170	102
10×8	0.15	0.00	193	166

5. まとめ

本稿では，DRP をマルチエージェント強化学習で解く際の安全を確保するための手法について提案した．また，その手法が安全性と学習効果の面から有効であることを示した．

6. 謝辞

本研究は，日本学術振興会科学研究費基盤研究（B）（21H03556，2021 年度～2023 年度）の補助を受けた．

参考文献

[1]青山 秀紀，丁 世堯，林 冬惠，“ドローン配送計画最適化問題のための最短経路情報を利用したマルチエージェント強化学習”，人工知能学会全国大会論文集，2022，JSAI2022 巻，第 36 回 (2022)

[2] ElSayed-Aly, Ingy, et al. “Safe Multi-Agent Reinforcement Learning via Shielding.” International Conference on Autonomous Agents and MultiAgent Systems (AAMAS). 2021.