2024 年度　修士論文

津波避難誘導のマルチエージェント強化学習とドローンによるアプローチの検討

23VR008N　　高林秀

指導教員　三宅陽一郎

立教大学大学院

人工知能科学研究科　人工知能科学専攻

概要

ここに概要を書く [**?**]

目次

第 1 章

はじめに

本章では, 本論文の要旨および構成について述べる.

1.1　要旨

首都直下型地震や南海トラフ地震をはじめとする, 大地震の 30 年以内の発生確率が 70%～80% と非常に高くなっていることに加え, 近年の豪雨など, 将来の大規模災害のリスクが著しく高まっている現状がある.

1.2　本稿の構成

まず, 第 2 章において本稿の内容を理解するのに必要な事前知識, 研究背景について述べる. 具体的には, 以下の項目について説明する.

- 強化学習についての基本説明
- マルチエージェントアルゴリズム MA-POCA（MultiAgent POsthumous Credit Assignment）について
- 本研究の社会的背景・課題について

次の第 3 章においては, 本研究の研究手法についての説明を行う. 第 4 章では, マルチエージェント強化学習エージェントによる, 津波避難誘導のシミュレーション実験の結果と考察を行う. 第 5 章では, 実験結果をまとめ, 本研究の応用, 今後の研究の展望を述べる.

第 2 章

研究背景

本研究を理解する上で必要な概念である, 強化学習とそのアルゴリズムである MA-POCA の理論や, 関連する研究について述べる. また, 本研究を行うことになった社会的背景についても述べる.

2.1　津波避難誘導における課題

本章では, 我が国での津波避難誘導における課題について取り上げ, 後述する提案手法の研究背景の理解を補助するものとする.

災害大国である我が国において, 地震発生後の津波避難誘導オペレーションは非常に重要である. 特に近年, 津波以外にも異常気象等による気象災害の激甚化もあり, 避難誘導の遂行にあたって, 益々その危険性も増していると推察される.

2.1.1　津波避難タワー・津波避難ビル

我が国には, 津波避難タワーや津波避難ビル[*1]が建設されており, 津波からの公的な避難先の 1 つとして提供されている. 当該施設の建設にあたっては, 避難経路や避難時間などの基準が国から示されており, 自治体により適切な位置に建設が進められている. このような施設は, 津波から命を守る手段として非常に重要であるが, 避難者の行動, 配分によっては収容定員を超過し, 適切な避難が行えない可能性があることが示されている [**?**]

2.1.2　訪日観光客数の増加と観光地における避難誘導の課題

近年の大幅な観光客増加と, 観光地における避難誘導の課題, その関連性について述べる.

[*1] 津波浸水が想定される地域において, 地震発生時に住民が一時的, または緊急に避難・退避するための人工施設を言う. 内閣府が 2005 年に策定した「津波避難ビル等に係るガイドライン」に沿って進められ,2011 年の東日本大震災の発生を受け,「津波防災地域づくりに関する法律」によって津波防災対策が制度化された.

我が国の観光客数増加　我が国では,2007 年にに観光立国推進基本法[*2]が施行され, 国として観光客数の増加が進められてきた. 観光庁の調査によれば, 我が国における訪日外国人観光客数は増加の一途を辿っている. 下図は観光庁が公開している,2003 年から 2023 年までの訪日外国人観光客数の推移を示したグラフである.

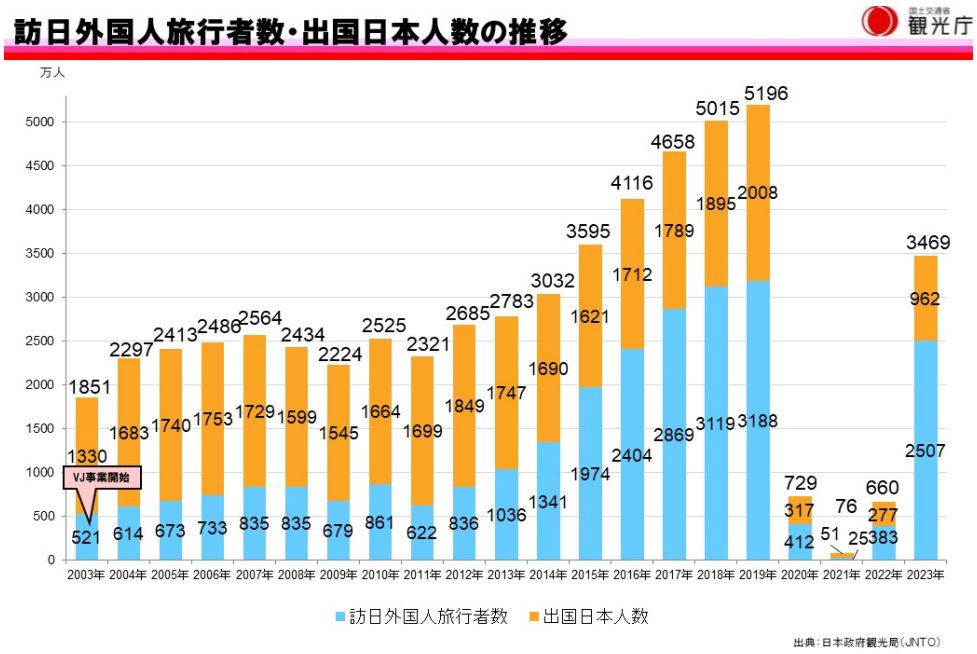


図2.1　2003 年～2023 年の訪日外国人旅行者数の推移

上図を読み解くと,2003 年から 2019 年にかけて, 訪日外国人観光客数は倍以上に増加していることがうかがえる.2020 年から 2022 年にかけては, 著しく観光客数が減少しているが, これは新型コロナウイルスの世界的流行による影響であると考えられる.

また,2023 年は新型コロナウイルスによる行動自粛が解除されたことを受け, 観光客数は 2015 年と同等水準まで回復しており, 今後も増加するものと推察される.

このような観光客数の急激な増加は, 観光地の災害時の避難誘導タスクにおいて, 以下のような問題を生じさせ適切な避難誘導を行えない可能性がある.

- 観光客の土地勘がないため, 的確な避難誘導が必要
- 観光客数は時間や季節によって変動するため, 特定の避難所に多数の避難者が向かい, 収容不足となる可能性がある.
- 避難誘導に従わずに周囲の人の動きに追従し, 混乱を招く恐れがある.

[*2] 議員立法により平成 18 年 12 月 13 日に成立し, 平成 19 年 1 月 1 日から施行されている. 本法律において, 観光は 21 世紀における日本の重要な政策の柱として初めて明確に位置づけられた.

観光地における避難誘導の課題　ほとんどの観光客は土地勘がないとともに, 防災意識もあまり高いとは言えない [?]. ゆえに, 単独では適切な避難行動がとれない可能性が高く, このような, 観光客の避難に関する問題は, 多くの関連研究でも指摘されている.

以上の背景から, 今後発生しうる, 南海トラフ地震などの巨大地震とそれにより発生する津波からの避難に関して, その対策は進められてきてはいるものの, 地元住民だけでなく観光客も含めた避難に関しては多くの課題を残している現状がある. また, 避難する人だけでなく, 避難者を適切な場所へ誘導する人員の安全確保にも課題が残されている.

2.1.3　二次被害の発生

津波避難誘導（あるいは, 他の災害における避難誘導）においては, 発災直後から二次被害にあう危険性が高い地域で活動しなければならないため, 現場で誘導を行う警察や消防員等の安全確保が問題になっている.

風水害時における人的被害の特徴　以下は, 我が国で発生した 1969 年から 2018 年までの災害を対象に, 消防団員が殉職した事例を消防白書や新聞記事, 既往研究などから把握し，殉職時の状況を分析した結果が, 山田らの研究 [?] によって報告されている.

> 図-3 より，津波は，出動途上，水防作業中，避難中，避難誘導中，人命救助中に殉職者を出したことがわかった．なかでも避難誘導中と避難中を合わせると全体で約 80% を占めており，避難に関係する時に殉職者が出ている.

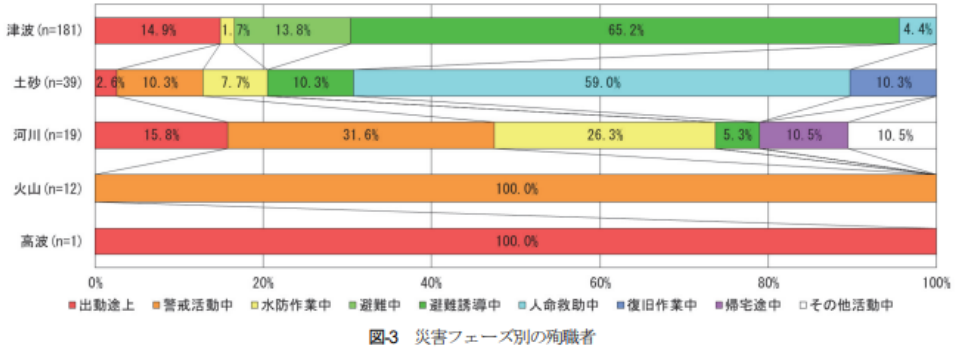


図-3　災害フェーズ別の殉職者

図2.2　消防団員の災害フェーズ別殉職者の割合

以上より, 津波災害時の消防団員おけるの 2 次被害に関しては, 避難誘導中が最も多い結果であることが示されている. 上記は消防団員に限定した統計であるが, 同じく避難誘導を行うすべての人員においても同様の傾向があると推察される.

また, 東日本大震災のケースにおいても, 避難誘導にあたった警察職員や自治体職員の多数が地域住民の避難誘導中に津波に巻き込まれ殉職された事例 [?] が報告されており, このような二次被害の防止は避難誘導において重要な意味を持つ.

2.2　既存のドローンの災害対応における活用事例と航空法改正

総務省・消防庁が公開しているデータ [?] によると、全国の消防本部におけるドローンの活用率は年々上昇しており、2017 年には 9.6% だったものが,2021 年には 52.9% と全国半数以上の消防本部でドローンの利活用が進めらたことが報告されている。

運用種別		累計件数
火災	建物火災	４０２
	林野火災	１５７
	上記以外の火災	１４３
調査	火災調査	１，８９６
自然災害（地震・雨）		２００
救助活動・捜索活動		８６１
その他※		３９２

図2.3　ドローンの運用種別ごとの累計活用件数

加えて、我が国では,2022 年に航空法が改正され, これまで規制されていたドローンの有人地帯目視外飛行（レベル 4 飛行[*3]）が解禁された. これにより, これまでドローンの活用が規制されていた防災分野での利活用や研究が大きく進んだ背景がある.

2.2.1　ドローンによる避難誘導の先行研究

2.3　強化学習

強化学習とは、エージェント[*4]と環境[*5]との相互作用を通じ、得られる報酬[*6]を最大化するエージェントの方策[*7]を学習する機械学習アルゴリズムの種類である.

[*3] 無人機の運用・操縦方法をレベル別に定めたもの. レベル 4 では操縦者が直接目視で機体を見ていなくても有人地帯でドローンを飛ばすことが可能になった.

[*4] モデルを訓練するための主体。環境に対して行動を出力する。

[*5] エージェントがいる世界、モデルの訓練を行うための様々な機能や状態を提供する。

[*6] エージェントの行動の良し悪しを判断する評価値. 行動に対する環境からの評価

[*7] ポリシーとも呼ばれる。環境の状態に基づいて、次の行動を決定するためのルール

図2.4　出典：布留川英一著 ML-Agents 実践ゲームプログラミングより

教師あり学習・教師なし学習の機械学習アルゴリズムと異なり、事前に訓練データを作成する必要はなく、訓練に必要なデータはエージェントが環境から得るものである. 強化学習は与えられた環境の中で、最適な戦略行動 (方策) を分析することが目的となる。

2.3.1　マルチエージェント強化学習の基本概念

マルチエージェント強化学習 (Multi-Agent Reinforcement Learning, MARL) では、複数のエージェントが環境と相互作用し、それぞれが自身の行動方策を学習しながら、協調または競争を行う。以下にその基本的な数式を示す。

環境の定義

環境は、部分観測可能マルコフ決定過程 (Decentralized-POMDP) として定義される：

$$\mathcal{M} = \langle N, S, \{O_i\}_{i=1}^{N}, \{A_i\}_{i=1}^{N}, P, r, \gamma \rangle$$

ここで：

- N：エージェントの数
- S：環境の状態空間
- O_i：エージェント i の観測空間
- A_i：エージェント i の行動空間
- $P(s'|s, \boldsymbol{a})$：状態 s と行動の組み合わせ $\boldsymbol{a} = (a_1, a_2, \ldots, a_N)$ から次の状態 s' への遷移確率
- $r(s, \boldsymbol{a})$：共有報酬関数
- γ：割引率

エージェントの行動方策

各エージェント i は、観測 O_i に基づき行動を選択する方策 $\pi_i(a_i|o_i)$ を学習する。エージェント全体の方策は次のように表される：

$$\pi(\boldsymbol{a}|\boldsymbol{o}) = \prod_{i=1}^{N} \pi_i(a_i|o_i)$$

状態価値関数と行動価値関数

- 状態価値関数 $V^\pi(s)$ は、状態 s から始まり方策 π に従ったときの期待累積報酬である：

$$V^\pi(s) = \mathbb{E}_\pi \left[\sum_{t=0}^{\infty} \gamma^t r(s_t, \boldsymbol{a}_t) \mid s_0 = s \right]$$

- 行動価値関数 $Q^\pi(s, \boldsymbol{a})$ は、状態 s で行動 $\boldsymbol{a}$ を取った場合の期待累積報酬である：

$$Q^\pi(s, \boldsymbol{a}) = r(s, \boldsymbol{a}) + \mathbb{E}_\pi \left[\sum_{t=1}^{\infty} \gamma^t r(s_t, \boldsymbol{a}_t) \right]$$

集中化された Critic

MARL では、集中化された Critic を用いて全エージェントの観測 $\boldsymbol{o}$ と行動 $\boldsymbol{a}$ を基に価値関数を近似する：

$$Q^\pi(s, \boldsymbol{a}) = f_\phi(s, \boldsymbol{a})$$

ここで f_ϕ はパラメータ ϕ を持つ関数近似器（通常はニューラルネットワーク）である。

Advantage 関数

アクター・クリティックアルゴリズムでは、Advantage 関数を用いて方策の更新を行う：

$$A^\pi(s, \boldsymbol{a}) = Q^\pi(s, \boldsymbol{a}) - V^\pi(s)$$

方策の更新

エージェントの方策は、Advantage 関数を最大化するように勾配上昇法で更新される：

$$\nabla_\theta J(\pi_\theta) = \mathbb{E}_{\pi_\theta} \left[\nabla_\theta \log \pi_\theta(a|s) A^\pi(s, a) \right]$$

協調と競争

協調タスクでは、全エージェントがグループ報酬 $r(s, \boldsymbol{a})$ を最大化する。一方、競争タスクでは、各エージェントが自分の報酬を最大化する。

2.3.2 MA-POCA（MultiAgent POsthumous Credit Assignment）

環境内のエージェントの個体数の増減に対応し、エージェント間の協調行動を重んじるようなタスクを学習するのに適しているアルゴリズムが MA-POCA（MultiAgent POsthumous Credit Assignment）[?] である。

MA-POCA は既存のマルチエージェントアルゴリズムと比較して、以下の特徴を持つ.

- 環境内のエージェント数の増減に対応した学習が可能
- エピソード内でエージェントが生成・消滅するタスクや、標準的な協調タスクにおいて、既存手法を大幅に上回る性能を示した

例えば、実世界で動くようなドローンをエージェントとして、その群衆飛行を考えた時、あるバッテリーが切れたり、故障したりすることが考えられ、エージェントが他のエージェントよりも先に行動不能（早期終了）になる場合が考えられる。既存のマルチエージェントアルゴリズムは、エージェントがエピソード[*8] 終了前に消滅した場合, そのエージェントの行動出力に関係なく状態を固定することでこれを再現する。これを吸収状態と言い、このようにすることで Critic への入力数を固定したまま学習を行うことが出来るが、同時に無駄な情報を入力しているとも捉えることができ、環境内のエージェント数が多いほどこの問題は顕著に出現することが指摘されている。

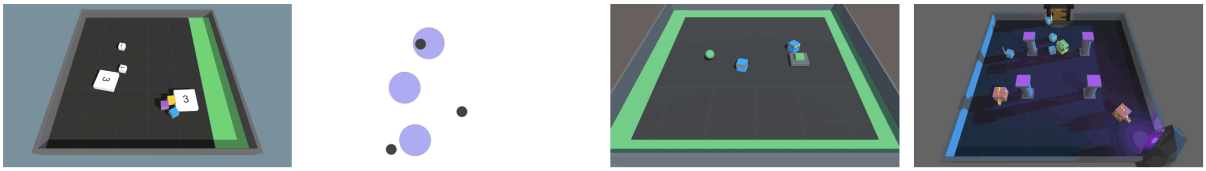
早期終了になったエージェントは、与えられたグループ報酬を経験することができない為, 自身の行動のグループにおける価値を計算することができない。MA-POCA は、この問題を解消するために提案されたアルゴリズムで、エージェントが早期終了しても価値を伝搬させるアルゴリズムとなっている。

MA-POCA の性能評価　MA-POCA は既存の MARL 手法よりも多くの場合で性能が向上することが報告されている [?]。下記のような 4 つの実験環境において, マルチエージェント強化学習手法 COMA[*9]、そして、シングルエージェント強化学習手法 PPO[*10] と MA-POCA の性能を比較した結果が示されている。

[*8] エージェントが環境と相互作用してタスクを完了するまでの一連のステップのこと。例えば、迷路のスタート地点からゴール地点までの移動がこれに該当する。一方、「ステップ」とは、そのエピソード内でエージェントが 1 回行動を選択し、環境から報酬と次の状態を受け取る単位時間のことである。

[*9]

[*10]



(a) (b) (c) (d)

図2.5　MA-POCA の性能評価を実施した環境

(a) **Collaborative Push Block** エージェント（青，黄，紫）は白いブロックを緑の領域まで押す。大きなブロックはより多くのエージェントが押す必要がある。

(b) **Simple Spread** エージェント（紫）は互いにぶつかることなく、ターゲット（黒）をカバーするように移動しなければならない。

(c) **Baton Pass** 青いエージェントが緑色の food をつかみ、緑色のボタンを押すと別のエージェントが生まれ、次の food をつかむことができるようになるので、それを繰り返す。

(d) **Dungeon Escape** 青いエージェントは緑のドラゴンを倒し、そのうちの 1 人を犠牲にしてカギを出さなければならない。チームメイトは鍵を拾って、ピンクのドラゴンを避けながら、ドアまでたどり着くタスク。

下図は、上記 4 環境における、累積報酬の推移を示している。このように、MA-POCA は既存の MARL 手法よりも多くの場合で性能が向上することが報告されている。

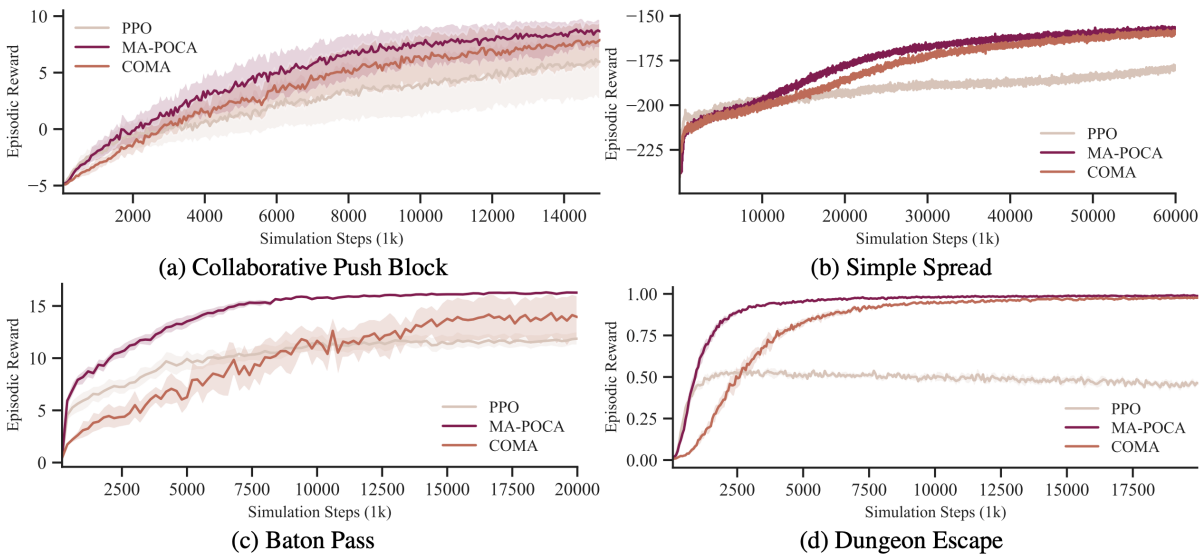


図2.6　各環境での MA-POCA の性能評価結果

2.4　ナビゲーションメッシュ

2.4.1　a* アルゴリズム

キャラクターの移動経路を探索するナビゲーション機能のアルゴリズムとして、a* アルゴリズムが広く利用されている。a* アルゴリズムは、最短経路を探索するためのアルゴリズムで、

第 3 章

提案手法

第 4 章

実験結果と考察

ddd

第 5 章

結果と考察

第 6 章

結論