



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	2
1.1	要旨	2
1.2	本稿の構成	2
第 2 章	研究背景	3
2.1	津波避難誘導における課題	3
2.1.1	訪日観光客数の増加と観光地における避難誘導の課題	3
2.1.2	津波避難タワー・津波避難ビル	4
2.1.3	二次被害の発生	6
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	6
2.2.1	ドローンによる避難誘導の先行研究と自治体の実証実験の事例	7
2.3	強化学習	8
2.3.1	マルチエージェント強化学習の基本概念	9
2.3.2	MA-POCA (MultiAgent POnthumous Credit Assignment)	11
2.4	ナビゲーションメッシュ	12
2.4.1	$a^*$ アルゴリズム	13
2.5	強化学習エージェントのデジタルツインへの応用	14
2.5.1	sim2real (Simulation to Reality)	15
第 3 章	提案手法と実験概要	17
3.1	提案手法の概要	17
3.2	既存研究との新規性	18
3.3	実験概要	19
3.4	シミュレーション前提条件	19
3.4.1	都市モデルの選定と避難所の配置条件	19
3.4.2	避難者の前提条件	20
3.4.3	エージェントの前提条件	20
3.5	避難者探索タスク実験方法	21
3.6	避難所誘導タスク実験方法	21

3.6.1	エージェントの観測	21
3.6.2	エージェントの行動	22
3.6.3	エージェントの報酬	22
3.6.4	評価と比較	22
<b>第 4 章</b>	<b>実験結果と考察</b>	<b>23</b>
4.1	避難者探索タスク実験	23
4.2	避難所誘導タスク実験	23
4.2.1	学習結果モデルの評価	23
<b>第 5 章</b>	<b>結果と考察</b>	<b>24</b>
<b>第 6 章</b>	<b>結論</b>	<b>25</b>
<b>参考文献</b>		<b>26</b>

# 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な課題である。また近年、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 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。

**我が国の観光客数増加** 我が国では、2007年に観光立国推進基本法<sup>\*1</sup>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。

---

<sup>\*1</sup> 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。

## 訪日外国人旅行者数・出国日本人数の推移

国土交通省  
観光庁

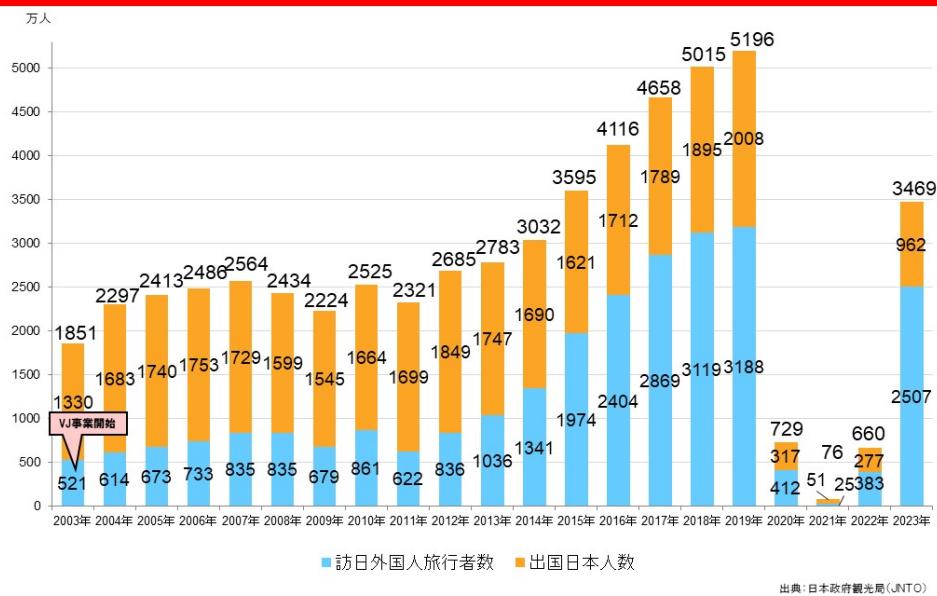


図2.1: 2003年～2023年の訪日外国人旅行者数の推移

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,以下のようないくつかの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.
- ・避難誘導に従わず周囲の人の動きに追従し,混乱を招く恐れがある.

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.

### 2.1.2 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<sup>\*2</sup>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.

\*2 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<sup>\*3</sup>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.



(a) 青森県八戸市の津波避難タワー



(b) 宮城県仙台市の津波避難ビル

図2.2

このような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。特に、観光地では景観等の問題から、十分な高さの堤防や防波堤を用意することが難しいといった問題もあり、津波避難対策を強化する施策としてこのような津波避難施設の設置は自治体を主導に行われている。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 [1]。

しかし、その母数が足りず想定される避難者の数をカバーしきれない等の指摘や、被害予測の改訂等で必要な高さ等の基準を満たせていない等の問題も存在する [?]。加えて、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結果がアンケート調査で判明している [2]。

このような状況下では、近隣の高台へ避難することが求められるが、土地勘のない観光客や外国人観光客に対してこれを求めるのはかなり難しく、既往研究の多くで指摘されている問題である。

また、観光地特有の問題として一部の避難所に避難者が集中し、避難完了時間が遅くなるというシミュレーション結果が示されており、適切に他の避難所（ないしは避難ビル）に避難者を誘導することの必要性が指摘されている [3]。

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 [4] 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ている。

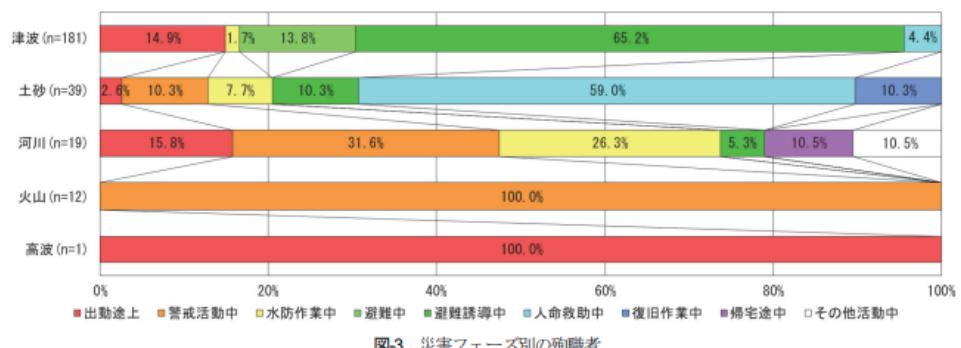


図2.3: 消防団員の災害フェーズ別殉職者の割合

以上より、津波災害時の消防団員におけるの 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 [5] 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

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<sup>\*4</sup>）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。

\*4 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。

	操縦方法	視界	飛行可能場所
000	レベル1	操縦飛行	目視内
	レベル2	自律飛行	目視内
	レベル3	自律飛行	目視外（補助者無し）
	レベル4	自律飛行	目視外（補助者無し）
			無人地帯
			有人地帯

表2.1: ドローンの飛行レベル別概要

近年我が国では、少子高齢化に伴う労働人口の減少の問題もあり、災害対応人材の不足が懸念されている背景がある。そのような人手不足に対応するため、ドローン等による災害対応の機械化・省人化が進められ始めている。総務省・消防庁が公開しているデータ [6]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れ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
	林野火災
	上記以外の火災
調査	火災調査
自然災害（地震・雨）	200
救助活動・捜索活動	861
その他※	392

図2.4: ドローンの運用種別ごとの累計活用件数

## 2.2.1 ドローンによる避難誘導の先行研究と自治体の実証実験の事例

ドローンを初めとする UAV の津波避難誘導に置ける活用方法を検討した既往研究が存在する。杉安らの研究では、津波避難時の迅速な避難行動を促進するために、UAV（無人航空機）の活用可能性を示し、避難誘導を視覚的に行うことの目的とし、福島県いわき市を対象に実証実験を行っている [7]。

また、本研究と関連する先行研究事例として、鈴木らが行った協調ドローンを用いた避難誘導支援システムの研究がある [8]。この研究では、ドローンを活用して、安全な避難経路を生成し、被災者を誘導するシステムを提案している。被災者は AR マーカーを身に着け、ドローンがこれを識別することで位置情報を取得し、後述する  $a^*$  アルゴリズムにより避難経路を探索し、計算した軌道情報に沿って対象者を誘導するものである。この研究では実機による誘導試験も行っており、ドローンによる避難誘導の実現可能性を示した。もう一つの先行

研究事例として、複数のドローンが連携して避難誘導を行うことを検証した高橋らの研究が存在する [9]。この研究では、自然災害時に UAV (無人航空機) を活用して避難誘導を行う支援システムの設計と試作について述べている。このシステムは複数の UAV が連携して避難者を誘導する仕組みを構築しており、沿岸部地域を対象にした実証実験まで行っている。UAV エージェントによる避難誘導プラン生成と経路選択、複数の UAV の協調による避難誘導機能の実現可能性が示された。

また、改正航空法の施行後、沿岸部の自治体を中心に、津波避難誘導を行うドローンの研究や実証実験が進められている。宮城県仙台市では、東日本大震災の際に津波避難誘導を行っていた自治体職員が津波に巻き込まれ犠牲になった事例を受け、津波避難を呼びかける手段として津波避難広報ドローンの研究が行われている [10]。この実証実験は、「自動運航のドローンにより津波避難広報を行うこと」及び「専用の LTE 通信網でドローンの制御等を行うこと」の 2 点において世界初の事例で、J アラート<sup>5</sup>による津波情報を受信した後、飛行経路上の気象条件を確認し自律的にドローンの飛行可否を判断した後、事前に定められた飛行ルート上を飛行しながら津波避難のアナウンスを行うというものである。また、ドローンの管制システムには専用の LTE 通信網を利用しておらず、令和 4 年 10 月に整備を完了し本格運用に入っている。

以上の様に、ドローンを用いた避難誘導システムの基礎研究や、自治体による避難誘導案内ドローンの整備など、災害時の避難誘導においてドローンの活用が検討、実装が進められている。

ただし、これらの研究事例においては、後述する提案手法で示す、本研究が行うような要避難者の位置分布に基づいた各ドローンの配置や移動、避難先の収容人数を考慮した避難誘導の最適化は行われていない。これらの点を後述する提案手法の章にて、本研究が示す新規性として具体的に述べる。

## 2.3 強化学習

強化学習とは、エージェント<sup>6</sup>と環境<sup>7</sup>との相互作用を通じ、得られる報酬<sup>8</sup>を最大化するエージェントの方策<sup>9</sup>を学習する機械学習アルゴリズムの種類である。

---

<sup>5</sup> 全国瞬時警報システム（J アラート）とは、弾道ミサイル情報、緊急地震速報、大津波警報など、対処に時間的余裕のない事態に関する情報を携帯電話等に配信される緊急速報システム

<sup>6</sup> 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。

<sup>7</sup> 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。

<sup>8</sup> 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

<sup>9</sup> 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

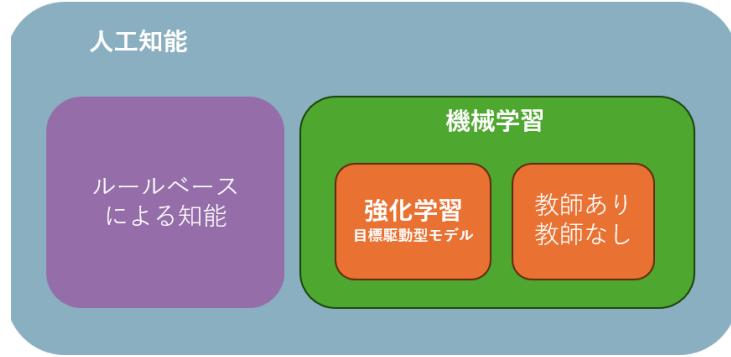


図2.5: 強化学習の枠組み概念図

教師あり学習・教師なし学習のデータ駆動型機械学習モデルと異なり, 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. このような特性から, 強化学習は目的駆動型モデルや行動駆動型モデルと呼ばれることもある. エージェントの一連の流れである「観測」, 「行動出力」, 「報酬獲得」のサイクルを決定と呼ぶ.

### 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
- $P(s'|s, \mathbf{a})$ : 状態  $s$  と行動の組み合わせ  $\mathbf{a} = (a_1, a_2, \dots, a_N)$  から次の状態  $s'$  への遷移確率
- $r(s, \mathbf{a})$ : 共有報酬関数
- $\gamma$ : 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：

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

## 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  $\pi$  に従ったときの期待累積報酬である：

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

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

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

## 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  $\mathbf{o}$  と行動  $\mathbf{a}$  を基に価値関数を近似する：

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

ここで  $f_\phi$  はパラメータ  $\phi$  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：

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

## 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：

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

## 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  $r(s, \mathbf{a})$  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) [11] 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<sup>\*10</sup> 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 [11]. 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<sup>\*11</sup>, そして, シングルエージェント強化学習手法 PPO<sup>\*12</sup> と MA-POCA の性能を比較した結果が示されている.

---

<sup>\*10</sup> 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.

<sup>\*11</sup>

<sup>\*12</sup>



図2.6: 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ぬ. 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.

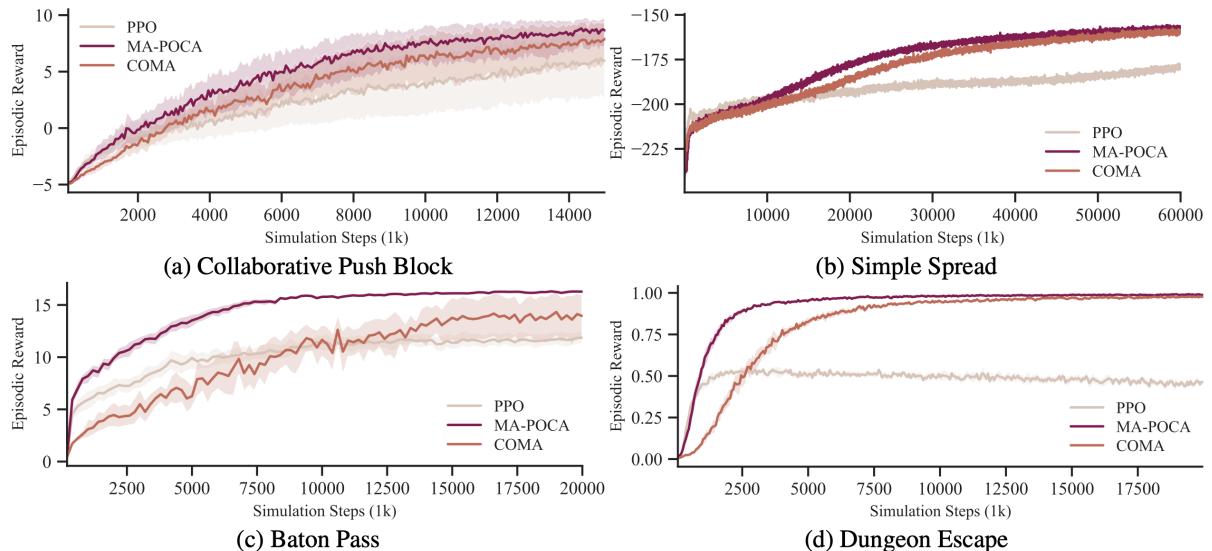


図2.7: 各環境での MA-POCA の性能評価結果

## 2.4 ナビゲーションメッシュ

デジタルゲームにおける人工知能 [12] には大きく 3 つの種類がある.

- キャラクター AI : ゲームやシミュレーション内で使用する NPC の頭脳
- メタ AI : ゲームやシミュレーション全体を監督し, 難易度等を調整する
- ナビゲーション AI : キャラクターの移動経路検索や障害物等の管理を行う

本研究では, エージェントの経路探索に上記のナビゲーション AI に区分される, ナビゲーションメッシュと呼ばれる機能を利用する. ナビゲーションメッシュは, ノード<sup>\*13</sup>の連結(グラフ)によって移動可能領域を覆うことで, キャラクターから移動可能範囲, 経路を認識させるためのゲーム AI 技術である.

都市モデル上でエージェントを動かす場合, あらゆる経路や移動手段が考えられ, エージェントが移動可能な道路や領域, 障害物として認識した上で行動するにはモデル訓練時の経験においてそれらを学習する必要がある. 本研究の目標は, エージェントが避難者を, 誘導人數や収容人數等を考慮し適切な避難所へ誘導することが目標なため, ナビゲーションメッシュを使用し事前にエージェントが移動可能な範囲を定めるものとする.

#### 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は 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グラフ探索アルゴリズムで, 経路をノードとして表現しグラフ探索を行う.

スタートノードからゴールノードまでの最短経路を探索する際に, 次の評価関数  $f(n)$  を用いる:

$$f(n) = g(n) + h(n)$$

ここで:

- $f(n)$ : ノード  $n$  の総評価値. スタートからゴールまでの推定コスト.
- $g(n)$ : スタートノードから現在のノード  $n$  までの実際のコスト.
- $h(n)$ : 現在のノード  $n$  からゴールノードまでの推定コスト (ヒューリスティック関数).

#### アルゴリズムの手順

A\* アルゴリズムは以下の手順で進行する:

1. スタートノードをオープンリストに追加し, 初期化する.
2. オープンリストから  $f(n)$  が最小のノードを選択する.
3. 選択したノードがゴールノードであれば, 経路探索を終了する.

---

<sup>\*13</sup> このノードをウェイポイントデータと呼び, ダイクストラ法等のグラフ探索アルゴリズムを用いて最短経路を検索することが可能になる

4. そのノードの隣接ノードを評価し, 以下を実行する :
  - 新しいノードであれば,  $f(n) = g(n) + h(n)$  を計算し, オープンリストに追加する.
  - 既に評価済みのノードであれば, より低いコストが見つかった場合に更新する.
5. 評価済みノードをクローズリストに移動し, 2 に戻る.

## ヒューリスティック関数

ヒューリスティック関数  $h(n)$  は, A\* アルゴリズムの効率と正確性を左右する重要な要素である. 一般的な選択肢として以下がある :

- マンハッタン距離: 格子状のグラフで利用される.
- ユークリッド距離: 2D または 3D 空間での最短直線距離を近似.

$h(n)$  が許容可能 (ゴールまでの実際のコストを過小評価しない) である場合, A\* アルゴリズムは最適解を保証する.

## 応用例

A\* アルゴリズムは, 以下のような応用分野で利用される :

- ゲーム AI: キャラクターの経路探索.
- ロボティクス: 障害物を回避する経路計画.
- 地図アプリケーション: 最短経路の検索.

## 2.5 強化学習エージェントのデジタルツインへの応用

デジタルツインとは, 現実空間に存在する建物や人流などの情報をリアルタイムで観測し, ネットワーク技術等を使用して仮想空間上に再現する技術のことである. 現実世界と対になる「双子」をデジタル空間上に構築し, モニタリングやシミュレーションを行うことで, 社会やビジネスプロセスを進化させることができ, 近年注目されている技術分野である.

CRDS<sup>14</sup>の調査によると, デジタルツイン関連の研究は, 工学分野や計算科学分野を中心として, 2016 年から 2021 年の過去 5 年間の研究論文数で約 30 倍に急増しており, 米国, ドイツ, 英国, 中国などでの研究開発が活発であり, 各国で大学, 公的研究機関, 民間企業が連携した研究プロジェクトが推進されていることが報告されている [13].

デジタルツインと強化学習を組み合わせることで, 現実空間で動作するロボットを仮想空間上で強化学習エージェントとして訓練をすることが可能になる. 具体的な事例としては,

---

<sup>14</sup> 国立研究開発法人科学技術振興機構研究開発戦略センター

Unity<sup>\*15</sup> を活用してロボットのデジタルツインを作成し、強化学習によるトレーニングを行うことで、仮想環境内での動作学習と実世界での性能向上を計るという研究がある [14]。また、我が国ではトヨタ自動車株式会社<sup>\*16</sup>と SCSK 株式会社<sup>\*17</sup>工場の製造ラインをデジタルツインで再現し、強化学習を活用してロボットの動作や生産プロセスの最適化を目指す取り組みがある [15]。

### 2.5.1 sim2real (Simulation to Reality)

sim2real とは、デジタル空間内のシミュレーションで学習したモデルを実世界でモデル適用、並びにタスクに適用する技術のことである。現実空間でエージェントの訓練環境を構築するよりも、デジタル空間でのシミュレーションを用いて訓練を行う方が、手軽に様々な環境条件を設定・試すことができ、低コストでモデルの実装を行うことができる。

一般に、ロボット分野において、シミュレーション上の訓練環境では、状態やその他要因により現実環境を完全に再現することは難しく、モデルの訓練環境と実環境（運用環境）とでギャップが生じる。このギャップが大きいと、エージェントが意図しない動きを行う可能性が高まり、安全上のリスクを伴う。また、モデル訓練時にエージェントが行う試行錯誤による事故や故障のリスクを軽減できる点も、sim2real 技術の利点である。

sim2real 技術を用いてドローンの飛行制御訓練モデルを作成し、現実空間での実機ドローン制御を行った研究が Rana Azzam らにより行われた。この研究では、深層強化学習 (Deep Reinforcement Learning, DRL) を用いて、動的環境における自律的かつゴール指向型のナビゲーションシステムを開発した。このシステムは、シミュレーション環境でエージェントをトレーニングし、高精度な UAV コントローラーモデルを使用することで、追加の sim2real 転送技術なしで現実世界へ適用されている。また、静的および動的障害物を回避しながら、安全かつ効率的に UAV をゴール地点まで誘導することを可能にしている。さらに、現実世界でのテストにおいて 90% の成功率を達成したことが報告されている [16]。

この研究は、消防や救助、監視、物流などのシナリオにおいて、高度な障害物回避機能を備えたリアルタイムのナビゲーションシステムとしての応用が期待されている。また、単一の UAV 制御にとどまらず、将来的には複数の UAV の協調制御への拡張も視野に入れている [16]。

デジタルツインと強化学習の融合は、システム設計、運用、制御のすべてにおいて革新的な可能性を提供する。この組み合わせは、安全性の向上、効率性の最大化、コスト削減を実現し、実世界の複雑な課題に対するソリューションを提供する手段として、今後ますます重要性を増していくだろう。製造業、建築設備、ロボット工学、モビリティ制御など、広範な分野での応

---

\*15 ユニティ・テクノロジーズ社が開発・提供するゲームエンジン。ゲーム開発の分野で世界シェアナンバー 1 を誇り、多くの RPG や位置情報ゲーム、VR コンテンツなどが制作可能。

\*16 愛知県豊田市に本社を置く日本最大手の自動車メーカー。

\*17 住友商事、住友グループのシステムインテグレータ企業

用が期待される中、これらの技術が社会全体にもたらす恩恵は計り知れない。

## 第3章

# 提案手法と実験概要

本章では、本研究が最終的に目指す津波避難誘導問題への解決策として、マルチエージェント強化学習と自律飛行型ドローンを組み合わせた提案手法について述べる。また、提案手法が既存研究と異なる点や新規性についても論じる。

### 3.1 提案手法の概要

本研究では、観光地や都市部といった地元住民以外にも多数の人々が屋外に存在する状況を想定している。これには、日常的に避難訓練を受けていない観光客や土地勘のない訪問者も含まれる。このような状況や、前章で述べた避難誘導における課題を背景に、地震発生後の津波避難という非常に緊急性の高い場面を想定し、避難誘導を行うための手法を検討する。従来、自治体職員や警察・消防隊員といった人間が担ってきた避難誘導を、自律飛行型ドローンが代替するシステムを構築することを目標とする。

具体的には、マルチエージェント強化学習を活用し、複数のドローンエージェントが協調して行動する能力を学習させることで、刻々と変化する被災地域の状況を動的に認識し、群衆の避難完了率を最大化することを目指す。また、避難者の位置、避難経路上の障害物、各ドローンの位置などをリアルタイムに反映するデジタルツイン環境を構築し、その環境内で学習済みのエージェントがシミュレーションを通じて最適な誘導方法を実行できるようにする。

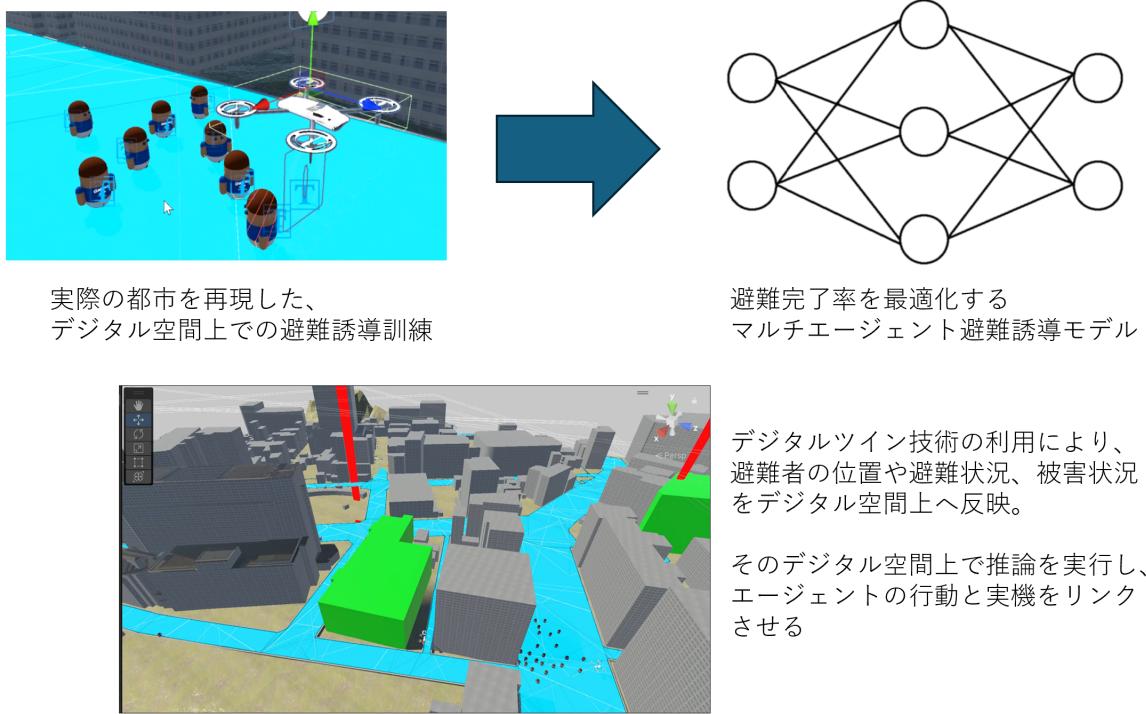


図3.1: 提案手法概略図

### 3.2 既存研究との新規性

研究背景で述べたとおり、本研究は既存の研究といいくつかの重要な相違点や新規性を有している。

まず、ドローンの防災活用はまだ研究が始まったばかりの新しい分野であり、本研究はその発展に寄与するものである。特に、複数のドローンを連携して運用するシステムに、AIや強化学習モデルを導入する点は、本研究の独自性を示す重要な要素である。

さらに、本研究では都市モデルを活用した訓練環境を構築し、デジタルツイン技術を通じて現実環境における運用を想定している。このように、現実世界での実用性を考慮した研究は、防災分野においても新しい試みである。

また、本研究は避難ビルの収容定員など、経路条件以外の要素を考慮した避難誘導モデルの作成にも取り組んでいる。従来の研究の多くが、避難者自身の行動最適化を通じて避難完了率の向上を目指しているのに対し、本研究では避難完了率を最適化できる避難誘導方策そのものを追求している点で特徴的である。

以上のような取り組みにより、本研究は現実世界での応用可能性を持つ動的な津波避難誘導システムの実現を目指している。

### 3.3 実験概要

本研究では、津波避難誘導問題を解決するために、マルチエージェント強化学習を用いたドローン避難誘導システムを提案する。本研究では、この問題を **1. 避難者探索タスク** と **2. 避難所誘導タスク** の 2 つの課題に分け、それぞれのタスクを解決するためのエージェントモデルを Unity 上のシミュレーションにより構築する。エージェントの訓練環境は実際の道路状況や避難所配置に限りなく近づけるため、都市モデルを活用したデジタルツイン環境を利用する。その後訓練済みエージェントモデルを利用した場合とルールベースで行動するエージェントを利用した場合それぞれのタスク遂行能力を評価するため、同一環境で比較シミュレーション実験を行う。なお、2. 避難所誘導タスクにおいては、ドローンエージェントの誘導がない場合、つまり避難者のみで避難行動を行う場合との比較も行う。その実験結果から、最終的な避難完了率の推移や経過時間などの指標を用いて、提案手法の効果検証と実現可能性を評価する。

### 3.4 シミュレーション前提条件

#### 3.4.1 都市モデルの選定と避難所の配置条件

環境としては、下記 3 つの都市の都市モデルを PLATEAU SDK for Unity を使い実際の都市環境に近いシミュレーション環境を再現する。なお、都市の選定基準については、3.1 章にて述べた想定場面を考慮するため下記の選定基準をもって決定した。

- 南海トラフ等で津波被害が想定されている沿岸地域であること
- 自治体の津波避難のハザードマップが参照可能であること
- 地元住民以外にも多数の観光客が見込まれる、比較的規模の大きな地域であること
- 津波避難ビルあるいは津波避難タワーが整備されている地域であること

以上の条件を元に、下記 3 つの都市をモデル都市として選択した。

1. 神奈川県横須賀市市役所本庁舎周辺沿岸地域
2. 静岡県沼津市沼津港周辺の一部地域

都市モデルと実際の避難ビル（避難タワー）との位置付けは、自治体公表のハザードマップより目算により算出し、都市モデル上で指定した。各モデル都市におけるシミュレーション対象範囲は以下地図の赤枠で示した範囲とする。

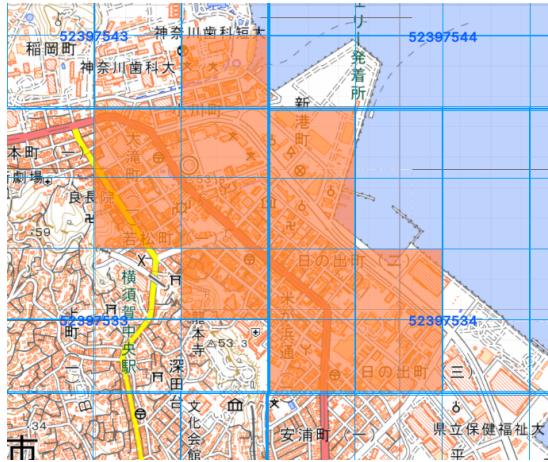


図3.2: 横須賀市でのシミュレーション範囲

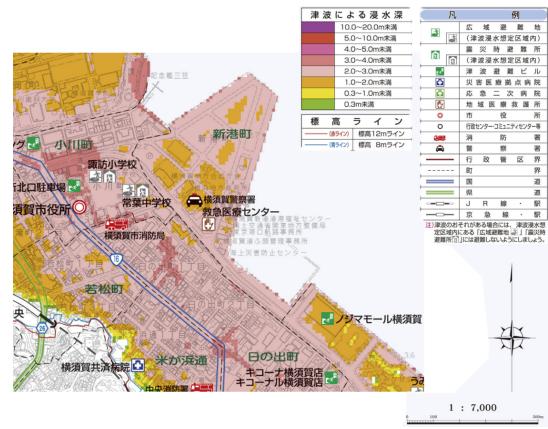


図3.3: 対象範囲の横須賀市のハザードマップ



図3.4: 構築したシミュレーション環境の例

### 3.4.2 避難者の前提条件

避難者の初期出現位置は環境内の道路上にランダムに配置するものとする。避難者の移動方法は内閣府が公表している津波避難ガイドライン基づき徒歩移動ないしは自転車による移動を想定する。そのため避難者の移動速度は、1m/s から 3.4m/s の範囲内でランダムに設定する。移動経路については、ナビゲーションメッシュにより指定位置までの道路上における最短経路を計算し、各避難者はその経路に従って移動するものとする。

### 3.4.3 エージェントの前提条件

エージェントの初期位置は、避難者の初期位置と同様に環境内の道路上にランダムに配置するものとする。なお、出現するエージェントの数は、環境内の避難所 1 つあたりにつき 1 機とする。移動方法の詳細は後述する各実験の章にて述べる。なお、強化学習アルゴリズムに

は MA-POCA を利用する。

### 3.5 避難者探索タスク実験方法

この実験ではエージェントは環境内の避難者を探索し、制限時間以内にできるだけ多くの避難者を見つけるタスクを行う。エピソード開始時、エージェントは環境内のランダムな道路上に配置される。その後エージェントは観測として、自身の環境内に位置情報と移動速度、現在発見した避難者の人数、制限時間と経過時間、他のエージェントの位置情報を取得する。またエージェントはレイキャスト<sup>\*1</sup>観測により、自身の付近に避難者がいる場合、その避難者の位置情報を観測することができる。その後、エージェントは行動として、自身の移動速度を  $1.0 \text{ m s}^{-1}$  から  $2.0 \text{ m s}^{-1}$  の間で移動速度を出力する。また、現在位置からの移動量を半径  $50 \text{ m}$  以内の範囲で決定し、移動を開始する。

### 3.6 避難所誘導タスク実験方法

この実験ではエージェントは事前に割り当てられた避難者グループを指定された避難所まで誘導するタスクを行う。エピドート開始時、エージェントは環境内のランダムな道路上に配置される。そのエージェントの周辺に 10 人から 40 人の避難者がランダムな人数配置される。次にエージェントは観測として自身の環境内での位置情報や割り当てられている避難者の人数、各避難所までの移動距離や収容人数などの情報を取得する。その観測に基づいてエージェントは行動として、自身の移動速度を  $1.0 \text{ m/s}$  から  $3.0 \text{ m/s}$  の間で、誘導先である避難所を 1 つ決定し移動することができる。避難者はエージェントに追従し、避難所に到達すると避難者は避難所に収容される。なお、避難所に収容される避難者の数は避難所の収容人数を超えることはないものとする。もし、避難時点でその避難所の収容定員を超える場合、避難者はエージェントに追従し続け、次のエージェントの行動決定まで待機する。制限時間を超えるか全ての避難者が避難所に収容された時にシミュレーションは終了する。

#### 3.6.1 エージェントの観測

エージェントは環境内の以下の情報を観測することができる。

- 自身の位置情報 (X, Y, Z 座標)
- 自身の移動速度 (浮動小数値)
- 割り当てられている避難者の人数
- 避難者グループの平均移動速度

---

<sup>\*1</sup> レイキャスト (Raycast) は、コンピュータグラフィックスや物理エンジンで使用される技術であり、特定の方向に仮想的な「線 (レイ)」を発射して、それがどのオブジェクトに当たるかを検出する手法。検知対象オブジェクトの位置情報を取得できる。

- 各避難所の位置情報 (X,Y,Z 座標)
- 各避難所までの移動距離
- 各避難所の現在の収容可能人数
- 他のエージェントの情報
  - 他のエージェントの位置情報 (X,Y,Z 座標)
  - 他のエージェントが移動する避難所の位置情報 (X,Y,Z 座標)
  - 他のエージェントが誘導している避難者的人数

### 3.6.2 エージェントの行動

エージェントは観測情報に基づいて以下の行動を取ることができる。

- 移動先の避難所
- 自身の移動速度

### 3.6.3 エージェントの報酬

各エージェントは誘導した避難所に自身に割り当てられている避難者が到達するとその人数に応じて +1 の正の個別報酬を得る。また、グループ報酬として、シミュレーション終了時点での最終的な避難完了率を計算し、その値に応じて 0~1 の正のグループ報酬を得る。

### 3.6.4 評価と比較

この実験では、マルチエージェント強化学習を用いたエージェントモデルとルールベースで行動するエージェントモデル、避難者単独行動のパターンの 3 つの行動パターンを比較する。ルールベースで行動するエージェントモデルは、自身から最も近い受け入れ可能な避難所までの最短経路を計算し、避難所に到達するまでの移動速度を一定に設定するものとする。避難者単独のパターンでは、各避難者は自身から最寄りの避難所までの最短経路を計算し、その経路に従って移動するものとする。なおこの時避難者は、移動先の避難所の収容人数を知ることはできず、避難所に到達した段階で収容可能人数を超える場合、次の最寄りの避難所に向かうものとする。

以上の 3 つの行動パターンにおいて、避難完了率、避難所到達までの経過時間、などの指標を用いて、マルチエージェントモデルの有効性を評価する。

## 第4章

# 実験結果と考察

本章では、沿岸地域における津波避難誘導をドローンエージェントに行わせる事を考え

### 4.1 避難者探索タスク実験

### 4.2 避難所誘導タスク実験

#### 4.2.1 学習結果モデルの評価

## 第5章

# 結果と考察

## 第 6 章

### 結論

# 参考文献

- [1]
- [2]
- [3] 北原武嗣, 岸祐介, and 久保幸獎. 高低差を考慮した津波災害時の群衆避難における経路選択に関する一検討. 土木学会論文集 A1 (構造・地震工学), 69(4):I\_1067–I\_1075, 2013. 鎌倉市材木座沿岸部を対象とした津波避難シミュレーションを通じて、パーソナルスペースや高低差の考慮が避難行動に及ぼす影響を検討。.
- [4] 山田忠, 後藤雄太, and 松枝心路. 風水害における消防団員の人的被害の特徴 – 1969 年から 2018 年までの災害を事例に –. 土木学会論文集 F6 (安全問題), 76(1):20–27, 2020.
- [5]
- [6] 総務省・消防庁. 無人航空機の災害時における活用状況等調査について. 消防の動き, 2, 2022.
- [7] 杉安和也, 高橋秀幸, 横田信英, and 片山健太. 津波避難時の誘導を目的とした uav 活用方法の検討. In 東日本大震災特別論文集 No.7, pages 7–10. 東北大学災害科学国際研究所, July 2018. 福島県いわき市での実証実験を含む.
- [8] 鈴木学, 浜克己, and 中村尚彦. 協調ドローンを用いた避難誘導支援システム. 計測自動制御学会論文集, 56(1):24–30, 2020. Vol.56, No.1, 24/30 (2020).
- [9] 高橋秀幸, 片山健太, 横田信英, 杉安和也, 北形元, and 木下哲男. Uav を活用した避難誘導支援システムの設計と試作. In FIT2018 (第 17 回情報科学技術フォーラム), pages 363–364. 情報処理学会, September 2018.
- [10] 仙台市. 津波避難広報ドローン事業, 2023. 最終アクセス日: 2024 年 12 月 20 日.
- [11] Andrew Cohen, Ervin Teng, Vincent-Pierre Berges, Ruo-Ping Dong, Hunter Henry, Marwan Mattar, Alexander Zook, and Sujoy Ganguly. On the use and misuse of absorbing states in multi-agent reinforcement learning. *arXiv*, 2111.05992v2(cs.LG), 2022.
- [12] 三宅陽一郎. 人工知能の作り方——「おもしろい」ゲーム AI はいかにして動くのか. 技術評論社, 東京, 2016. ISBN: 978-4-7741-8627-6.
- [13] 国立研究開発法人科学技術振興機構研究開発戦略センター. デジタルツインに関する国

内外の研究開発動向. Technical report, 国立研究開発法人科学技術振興機構, 3 2022.  
調査報告書 CRDS-FY2021-RR-09.

- [14] Unity Technologies. Made with unity : ロボットのデジタルツインの制作とトレーニング, 3 2021. 最終アクセス: 2024 年 12 月 19 日.
- [15] SCSK 株式会社. 工場の「デジタルツイン」で働き方を変える。トヨタ自動車が目指す魅力的な職場づくり, 2 2024. 最終アクセス: 2024 年 12 月 19 日.
- [16] Rana Azzam, Mohammad Chehadeh, Oussama Abdul Hay, Igor Boiko, and Yahya Zweiri. Learning to navigate through reinforcement across the sim2real gap. *TechRxiv Preprint*, October 2023. Preprint available on TechRxiv.