CF-003

海上警備：深層強化学習を用いた巡視戦略の最適化

Maritime Security : Optimizing Patrol Strategy Through Deep Reinforcement Learning

沖本 天太 *
Tenda Okimoto

山陰 将典 †
Masanori Yamakage

1. 序論

海洋国家である日本は，四方を海に囲まれており，広大な領海と排他的経済水域 (EEZ) を管理する責務を負っている．日本における海洋安全保障の強化は，海洋の平和と安全，持続可能な利用と開発，海洋環境の保全に加え，漁業資源や海底資源の確保，海上交通や航空交通の安全など，国家の平和と繁栄に不可欠な国家レベルの重要課題の一つである．海洋安全保障の強化に欠かせないのが海上警備であり，海上での人命・財産の保護を含む治安維持，領海内へ侵入しようとする不審船の発見・追跡・対処等が主な活動として挙げられる．

海上警備では，刻一刻と変化する動的環境下において，アジリティの高い巡視計画が求められる．領海や EEZ への侵入を試みる不審船は，巡視船の監視を回避しようとするため，その行動は予測困難である．また，不審船は相手が人間であるという性質上，領海警備網を回避するような戦略的行動を取る可能性も考慮する必要があり，実務・学術の両面から海上警備に関する研究が求められている．海上警備に関する既存研究には，警戒区域や警戒レベルの予測に基づき巡視船の経路を決定するような巡視経路問題 [1, 2, 4]，巡視船，巡視領海，巡視時間帯等の巡視業務の集合に対して，与えられた制約を満たすように，巡視船を巡視領海内に配置する巡視船配置問題 [3, 5, 7]，ドローンと巡視船を併用した空と海からの海上警備割当問題 [6] 等がある．

本研究では，刻一刻と変化する動的環境下において，アジリティの高い海上警備計画の構築を目的とする．具体的には，深層強化学習 (Deep Reinforcement Learning, DRL) を用いて，不審船の動的な行動を学習・予測するモデルを提案する．提案手法では，巡視船の初期配置を学習する手法として深層 Q-Learning (DQL) 及び，巡視船の巡視戦略を学習する手法として Deep Recurrent Q-Network (DRQN) を採用し，巡視船が海上警備環境との相互作用を通じて逐次的に最適な行動方策を学習する．実験では，不審船の行動予測・発見及び，巡視戦略の学習に対する提案手法の有効性を示す．最後に，海上警備に関する既存研究はいくつか存在するが，日本国内を対象とした研究は少なく，DRL を用いた海上警備に関する研究は寡聞にして見当たらない．

2. 準備

2.1. 強化学習

強化学習は代表的な機械学習の一つであり，意思決定者である「エージェント」と「環境」の相互作用に基づく逐次的意思決定過程問題である．強化学習ではマルコフ決定過程が用いられる．マルコフ決定過程とは，S を状態空間，A をエージェントが取る行動からなる集合，P を環境の状態遷移確率，R を報酬関数，γ を割引率 $(0 \leq \gamma \leq 1)$ とし，$\langle S, A, P, R, \gamma \rangle$ の組により定義される．強化学習は，エージェントの利得（割引報酬の総和）$G = \sum_{t=0}^{T-1} \gamma^t r^{(t+1)}$ の最大化を目的とする．ここで，T は時系列，$t \in T$ は時刻，$r^{(t+1)}$ は時刻 $t+1$ における報酬を表し，本論文では $\gamma = 0.95$ としている．さらに，密漁船は強化学習の一手法である Q 学習を用いて行動戦略を学習する．具体的には，密漁船は，海洋資源に関する分布を状態空間とし，分布上で得られる報酬が最大となるような行動戦略を学習する．

2.2. 深層強化学習

深層強化学習 (Deep Reinforcement Learning, DRL) は，複雑な環境における意思決定問題を解くことを目的とした，強化学習と深層学習を統合した手法である．DRL の目的は，エージェントが環境との相互作用を通じて最適な行動戦略を学習することであり，特に，高次元の状態空間を扱う問題に適している．深層学習はニューラルネットワークを用いて，複雑なデータから特徴量を抽出し，学習していく技術である．DRL では，特に，高次元の観測データを効率的に扱うために，ニューラルネットワークを Q 関数の近似に活用する．具体的には，Deep Q-Network (DQN) では，エージェントが自身の経験をリプレイバッファ D に，タプル形式 (s_t, a_t, s_{t+1}, r_t) で保存する．このバッファは時間的相関を切り離し，経験を効率的に再利用するための重要な役割を果たす．DQN では損失関数 $L(\omega)$ を最小化するように学習する．

$$L(\omega) = \mathbb{E}_{(s_t, a_t, r_{t+1}, s_{t+1}) \sim D}[(r_{t+1} + \gamma \cdot \max_a Q(s_{t+1}, a; \omega^-) - Q(s_t, a_t; \omega))^2] \quad (1)$$

ここで，ω^- はターゲットネットワークの重みを表す．本研究では，巡視船の行動戦略を学習するために DRL を用いる．具体的には，巡視船をエージェント，密漁船の行動範囲や海洋資源の分布を含む空間を状態空間とし，巡視船が最適な巡視ルートを学習する．その際，巡視船は密漁船を発見することで報酬を得るが，資源の保全や巡視範囲の制約条件に基づく戦略的な巡視も求められる．ここでは，DQN を用いることにより，巡視船が環境との相互作用を通じて，密漁船の動きを予測し，効率的な巡視行動を学習するプロセスを実現する．

2.3. 長・短期記憶 (LSTM)

再帰型ニューラルネットワーク (Reccurent Neural Network, RNN) は，シーケンスや時系列データを扱うために設計されたニューラルネットワークであり，過去の情報を基に現在の出力を計算し，データの時間的

*神戸大学大学院海事科学研究科
†神戸大学海事科学部

構造を考慮することができる．しかし，標準的な RNN には勾配消失問題や勾配爆発問題といった課題が存在する．これらは時間的な依存が長くなるにつれて誤差勾配が適切に伝播しなくなる現象を指す．長・短期記憶 (Long Short Term Memory, LSTM) とは，標準的な RNN の一種であり，RNN が抱える「勾配消失問題」や「勾配爆発問題」を克服するために設計されたモデルである．RNN では，時間的に離れた情報の依存関係を学習する際に誤差が伝播しづらくなるため，長期依存関係の処理が困難となる．一方，LSTM では内部にセル状態とゲート機構を導入し，重要な情報を選択的に保持し，不要な情報を忘却する仕組みを備えている．LSTM の動作は以下の主要なゲートによって制御される．

1. 忘却ゲート：セル状態から不要な情報を削除する．ここで，f_t は忘却ゲートの出力，W_f は重み行列，b_f はバイアス項，σ はシグモイド関数を表す．

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{2}$$

2. 入力ゲート：セル状態に新しい情報を追加する．ここで，i_t は入力ゲート，$\tilde{C}_t$ は候補セル状態を表す．

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i),$$
$$\tilde{C}_t = tanh(W_c \cdot [h_{t-1,x_t]+b_C}]) \tag{3}$$

3. 出力ゲート：次の隠れ状態 h_t を決定する．

$$h_t = o_t \cdot tanh(C_t),$$
$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{4}$$

4. セル状態：時刻 t のセル状態 C_t は以下で更新する．

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \tag{5}$$

2.4. 最適化アルゴリズム (Adam)

Adam (Adaptive Moment Estimation) は，確率的勾配降下法の拡張版として広く利用されている最適化アルゴリズムである．Adam は，学習率の適応的な調整を行いながら，一階および二階のモーメントを利用して効率的にパラメータの更新を行う手法である．その特徴は，計算効率が高く，メモリ要件が少ないこと，さらに学習率の調整が容易である点にある．このため，大規模データセットや高次元パラメータ空間を扱うディープラーニングにおいて標準的な選択肢であると言える．Adam は，以下の更新則に基づいて動作する．まず，時刻 t におけるパラメータ θ_t の勾配を $g_t \nabla_\theta J(\theta_t)$ とし，1 階モーメント m_t と 2 階モーメント v_t を更新する．

$$m_t = \beta_1 m_{t-1} + (1-\beta_1) g_t \tag{6}$$

$$v_t = \beta_2 v_{t-1} + (1-\beta_2) g_t^2 \tag{7}$$

ここで，β_1 と β_2 は，一階および二階モーメントの減衰率を表すハイパーパラメータであり，典型的には $\beta_1 = 0.9$, $\beta_2 = 0.999$ が用いられる．次に，これらのバイアスを補正するために以下に示す補正項を計算する．

$$\hat{m}_t = \frac{m_t}{1-\beta_1^t}, \hat{v}_t = \frac{v_t}{1-\beta_2^t} \tag{8}$$

最後に，パラメータは以下に従って更新される．ここで，η は学習率，ϵ は微小値（分母ゼロの回避）を表す．

$$\theta_{t+1} = \theta_t - \eta \frac{m_t}{\sqrt{\hat{v}_t} + \epsilon} \tag{9}$$

3. 深層強化学習を用いた巡視戦略

深層強化学習を用いて，不審船の動的な行動を学習・予測するモデルを提案する．本手法では，巡視船の最適な初期配置を学習する深層 Q-Learning (DQL) と，巡視船の最適な巡視戦略を学習する Deep Recurrent Q-Network (DRQN) を採用し，巡視船は環境と相互作用を通じて逐次的に最適な行動方策を学習する．まず，学習環境について概説する．次に、巡視船の学習アルゴリズムと，本手法のシミュレーションフローを示す．

3.1. 学習環境の設計

巡視船が巡視行動を学習する環境は，実際の海洋環境を簡略化して設計する．本研究では，巡視領海を一辺 150km，500km の長方形のエリアとする．さらに，1 グリッドを 50km の正方形（巡視船の監視可能な範囲の半径）とし，巡視船と密漁船は 3×10 の二次元グリッド上で各々の行動戦略を実行する（図 1 を参照）．

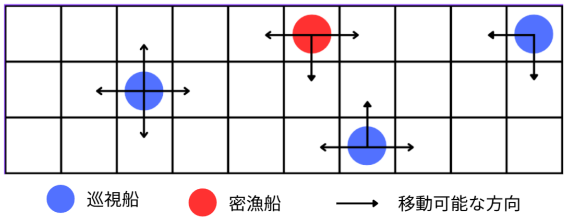

図 1: 二次元グリッド環境のイメージ図．

気象および潮流のモデル化と影響

本研究では，気象および潮流が環境に与える影響を考慮し，動的な環境モデルを構築した．具体的には，潮流が高い場合には資源の流動性が増加することを反映し，重み付きテーブルの値を上昇させた．また，潮流が活発であるほど，密漁船は，その領域に集中する可能性があるため，スポーン確率も上昇するように設計した．さらに，潮流の影響を資源の減少速度にも反映し，フェード速度を早めることで資源分布の動的変化をシミュレーションした．一方，気象条件が良好な場合，密漁船の活動が活発になることを想定し，スポーン確率が上昇するように設定した．以上の設定を用いて，動的環境下における海洋領域を構築し，不審船の行動や重み付きテーブル上の資源分布の変化を再現した．

巡視船と密漁船

巡視船は監視する領海全体（二次元グリッド環境）を効率的にカバーすることを目的とする．まず，初期配

置の最適化として，巡視領海を巡視する他の巡視船と巡視エリアが被らないような巡視船の位置を選択する．また，高資源の領海エリアを監視する位置を選択する．次に，巡視ルートの選択として，時間と共に動的に変化する海洋環境の要因（潮流・気象条件など）や，密漁船の行動に基づき巡視戦略を逐次的に学習する．これに対し，密漁船は巡視船による監視を回避しながら資源を獲得することを目的とする．具体的には，まず，高資源の領海エリアにおいて資源を獲得するような行動（移動）を選択する．また，巡視船が近づいてきた場合は，監視を回避するための行動（移動）を選択する．

状態空間 (S)

海洋環境の状態は，資源の分布（重み付きテーブル），動的な変化を反映した潮流および気象条件，巡視船および密漁船の位置，時系列情報と DRQN を用いて処理した過去の状態遷移に関する情報を統合的に勘案した高次元ベクトルとして定義される．状態ベクトルに関しては，(i) 初期配置を学習する場合は，単一の状態として処理される，(ii) 動的な巡視ルートの選択では，シーケンス長 L を持つ時系列データとして処理される．

行動空間 (A)

グリッド環境内の任意のセルを巡視船の行動として定義し，巡視船の初期配置を学習する．また，巡視ルートの選択に関しては，グリッド環境のセルの上下左右の 4 方向から 1 つを選択して移動する（図 1 を参照）．

報酬関数 (R)

- 巡視船の初期配置の学習に関しては，資源分布，密漁船の出現率，他の巡視船との監視範囲の重複に基づいて報酬・ペナルティをそれぞれ与える．具体的には，(i) 高資源エリアに配置された場合は正の報酬を付与する．(ii) 密漁船が出現しやすいエリアに近い場合は報酬を加算する．(iii) 他の巡視船と監視範囲が重複する場合はペナルティを課す．
- 巡視ルートの選択は，巡視船が密漁船を補足した場合は正の報酬を付与する．また，高資源の領海エリアを巡視した場合は中程度の報酬を与える．
- 密漁船に関しては，以下の条件に基づいて報酬・ペナルティをそれぞれ与える．(i) 高資源の領海エリアへ到達した場合，正の報酬を付与する．(ii) 監視を回避した場合，微小な報酬を付与する．(iii) 巡視船に発見された場合，負の報酬を付与する．

3.2. 巡視船の学習アルゴリズム (DRQN)

巡視船の巡視戦略の学習に関しては，密漁船の過去の行動パターンに基づく最適な巡視行動の選択を可能とする Deep Recurrent Q-Network (DRQN) を活用する．DRQN は，入力層，リカレント層，全結合層の三層構造を持つ．入力層では，シーケンス長 L と状態次元 d_s の入力データを受け取る．リカレント層では，LSTM セルを使用し，時系列依存性を学習する．全結合層では，隠れ状態を行動空間の次元数 d_s にマッピングし，各行動に対する Q 値を出力する．モデルの学習は式 (1) で示した損失関数 $L(\omega)$ を最小化することで行われる．学習アルゴリズム DRQN をアルゴリズム 1 に示す．

Algorithm 1 巡視船の学習アルゴリズム (DRQN)

Require: 環境 $\mathcal{E}$，巡視船エージェント π_ω，経験リプレイバッファ D
Require: エピソード数 N，最大タイムステップ T，探索率 ϵ
for episode = 1 to N **do**
  初期状態 s_1 を取得し，LSTM 層の隠れ状態 h_1 を初期化
  for t = 1 to T **do**
    ϵ-greedy 方策で行動 a_t を選択
    行動 a_t を実行し，報酬 r_t と次の状態 s_{t+1} を観測
    経験 $(s_t, h_t, a_t, r_t, s_{t+1})$ をバッファ D に保存
    LSTM 層の隠れ状態 h_{t+1} を更新
    if 経験リプレイバッファ D のサイズが十分 **then**
      ミニバッチを D からサンプリング
      損失関数を最小化するように ω を更新（LSTM 層を含む）
    end if
  end for
end for

3.3. シミュレーションフロー

巡視船と密漁船の学習は，以下の流れで行われる．

1. 環境の初期化：重み付きテーブル，潮流・気象条件，密漁船の初期配置（ランダム）が設定される．
2. 巡視船と密漁船の行動選択：巡視船は過去の状態シーケンスを DRQN に入力し，逐次的に行動を選択する．密漁船は Q 学習を用いて行動を選択し，重みの大きいエリアを目指し，巡視船を回避する．
3. 行動と報酬の計算：巡視船と密漁船はタイムステップ毎に行動する．また，巡視船は密漁船を補足した時と重みの大きいエリアを巡視した時に報酬が与えられる．密漁船は重みの大きいエリアに到達した時と，巡視船を回避した時に報酬が与えられる．巡視船に補足された時はペナルティが課される．
4. 学習：初期配置は DQL を用いて行動価値関数を更新し，次エピソードの初期配置を改善する．巡視船の巡視ルートは，経験リプレイとターゲットネットワークを活用して，巡視船の行動価値関数を学習する．密漁船は Q 学習を用いて行動価値関数を更新し，巡視船を回避する戦略を最適化する．
5. 結果の記録と可視化：エピソード終了後，巡視船の初期配置，移動軌跡，成功率を記録し，学習プロセスの進行を評価する．具体的には，巡視船の移動ヒートマップや重み付きテーブルとの相関分析を通じて，学習結果の有効性を視覚的に確認する．

提案手法は，巡視船が効率的，かつ，柔軟な巡視戦略を学習することを目的としている．まず，初期配置の学習では，DQL を活用することにより，高資源の領海エリアを優先的に巡視し，巡視船間の監視範囲の重複を最小化する配置を学習する．この手法により，従来の固定的な初期配置と異なり，動的な資源分布や密漁船の出現に適応した最適な初期配置が選択可能となる．

さらに，巡視ルートの選択では，DRQN の時系列に対する処理能力を活かし，動的環境の変化を考慮した巡視行動を学習する．この手法により，潮流や気象条件などの外的な要因が巡視戦略に与える影響を適切に反映することができ，巡視船が最適な巡視ルートをリアルタイムに選択できるようになる．これらの戦略は，高資源の領海エリアを効率的に巡視しつつ，密漁船を迅速に補足する能力を強化することが可能となる．

最後に，提案手法では，巡視船の初期配置と巡視ルートの選択の双方を統合的に最適化することにより，巡視活動の効率性を向上させている．具体的には，巡視船の初期配置を適切に決定することで巡視基盤を整えることができ，さらに，密漁船の動向に柔軟に対応することでアジリティの高い巡視戦略の学習が期待される．

4. 実験

本章では，巡視船の巡視戦略の最適化を目的としたシミュレーション評価を行う．まず，実験設定および評価指標について概説する．次に，実験結果として，巡視船の初期配置，移動軌跡，成功率をそれぞれ与える．

4.1. 実験設定

以下，仮想環境，密漁船の生成・動作モデル，エピソードと進行，報酬値についてそれぞれ説明する．まず，仮想環境の設定に関しては，密漁船と巡視船の相互作用を模倣するため，実際の密漁船監視シナリオを基に仮想環境を構築した．また，環境設計では，資源の動的な変化などを考慮し，現実的な動作を再現した．

密漁船の行動に関しては，資源分布を表す重み付きテーブルに基づいてモデル化した．グリッド環境上の各セルには 0 から 2 の値（重み）が割り当てられており，密漁船は重みが大きい領海エリアを目指して移動する．また，周辺の重みが現在の位置より低い場合には，その場に停滞するとした．密漁船の出現頻度に関しては時間帯に応じて動的に変化するよう設計した．具体的には，1 エピソード 24 タイムステップのうち，3 から 10 ステップの間で最も活動が活発になり，この時間帯では，30 %の確率で密漁船が出現するように設定した．それ以外の時間帯では出現確率は 10 %に設定した．

エピソードとタイムステップに関しては，1 エピソードを実世界の 1 日と想定している．また，1 エピソードは 1 タイムステップ，すなわち，エージェントが 1 回行動する間隔を 1 時間と仮定した 24 タイムステップから構成されている．タイムステップの間隔は，密漁船と巡視船が 1 タイムステップで 1 グリッドを移動できる時間，すなわち，以下の式を用いて算出している．

$$\frac{1\,\text{グリッドの距離}}{\text{巡視船の時速}} = 1\,\text{タイムステップの間隔} \quad (10)$$

巡視船と密漁船の報酬に関しては，各々の目的を達成するのための行動を学習するように設計した．巡視船は効率的に密漁船を監視または補足する行動を取り，密漁船は巡視船の監視を逃れ，長期間生存する行動を取るように学習する．以下，詳細な報酬設計を示す．

- 密漁船の報酬
  - 監視回避による報酬：密漁船が巡視船に監視されずに行動できた場合，1 タイムステップ毎に微小な報酬を付与する．この報酬設計により，密漁船の監視回避は継続される．
  - 長時間滞在による報酬：密漁船が巡視船に発見されずに消滅した場合，報酬を付与する．
  - 監視によるペナルティ：密漁船が巡視船によって監視された場合，ペナルティを付与する．

  以下，本実験で用いた具体的な数値を与える．

$$\text{密漁船の報酬} = \begin{cases} 2 & \text{巡視回避の微小報酬} \\ 15 & \text{寿命全うによる報酬} \\ \text{-10} & \text{監視時のペナルティ} \end{cases}$$

- 初期配置に対する報酬
  - (a) 資源量に基づく評価：選択された初期配置における重み付きテーブルの値を報酬として加算する．これにより，高資源の領海エリアを優先的に巡視するように学習される．
  - (b) 密漁船の出現率の予測：巡視船の初期配置が，密漁船の出現率の高い領海エリアに近い場合，報酬を加算する．これにより，密漁船を効率的に発見する初期配置が学習される．
  - (c) 監視エリアの重複回避：巡視船同士の初期配置が重複する場合，ペナルティを課す（重複数に比例して減点）．これにより，各巡視船の初期配置が分散されるように学習される．

$$\text{巡視船の報酬} = \begin{cases} +\omega(x,y) & \text{(a)} \\ +p_{spawn} & \text{(b)} \\ -5 \times \text{重複数} & \text{(c)} \end{cases}$$

- 巡視ルートに対する報酬
  - 密漁船の補足報酬：巡視船が密漁船を補足した場合，報酬を付与する．これにより，密漁船を効率的に追跡・補足する行動が学習される．ここで，巡視船が密漁船を監視･補足するとは，グリッド環境内の同じ領海エリア（セル：1 グリッド= 50km の正方形）内に，巡視船と密漁船が同時に存在することを指す．
  - 巡視･探索行動の奨励：巡視船が密漁船を発見できなかった場合，巡視船は報酬を得ることができない．これにより，巡視・探索行動を通じて密漁船を発見する戦略が学習される．

$$\text{巡視行動の報酬} = \quad 10 \quad \text{密漁船補足の報酬}.$$

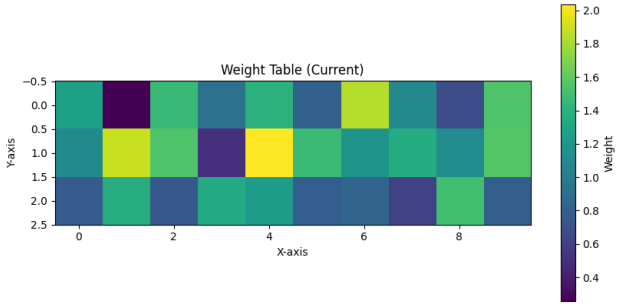

図 2: (i) 重み付きテーブルヒートマップ

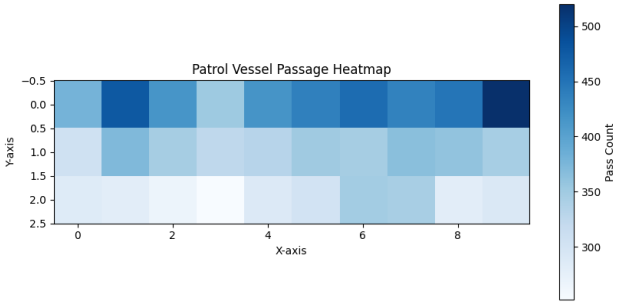

図 4: (iii) 移動軌跡ヒートマップ

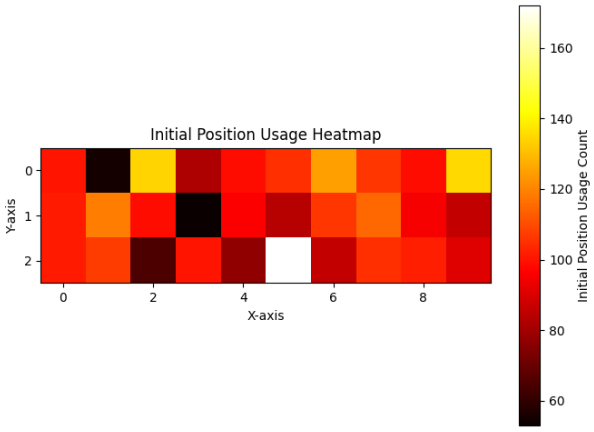

図 3: (ii) 初期配置ヒートマップ

4.2. 評価指標

本研究では，巡視船の巡視戦略（行動）の正当性の評価指標として，各グリッドに対する重み付きテーブル，初期配置・軌跡のカウント数をまとめたヒートマップ，学習の成功率をそれぞれ示す．初期配置に関するヒートマップは初期配置に選ばれたグリッドをカウントする．軌跡のカウント数は巡視船が通ったグリッドをそれぞれカウントする．これらを，重み付きテーブルと比較することにより，巡視船が，どの領海エリアを優先的に巡視・監視したかを視覚的に与える．また，学習の成功率に関しては，各エピソードにおける成功率を $S_t = C_t/P_t$ で与える．ここで，C_t はエピソード t における巡視船が密漁船を補足した数，P_t はエピソード t におけるスポーンした密漁船の数を表す．ただし，密漁船のスポーン数は乱数により確率的に決定されているため，各エピソード単体の成功率には，ばらつきが生じる可能性がある．そこで，学習の妥当性を適切に評価するために，以下に示す過去 W エピソードの平均を用いたスムージング成功率 $\bar{S}_t$ を用いて計算する．

$$\bar{S}_t = \frac{\sum_{i=t-W+1}^{t} S_i}{W} \tag{11}$$

ここで，W は移動平均のウインドウサイズを表し，本論文では $W = 10$ と設定した．また，式 (11) の分子は，過去 W エピソードにおける成功率の合計を表す．

4.3. 実験結果

実験では，まず巡視船の巡視行動を 3 つのヒートマップ，(i) 重み付きテーブルヒートマップ，(ii) 初期配置ヒートマップ，(iii) 移動軌跡ヒートマップを用いて評価する．次に，提案手法の有効性を評価するために，巡視船が密漁船を監視できた割合を成功率として解析する．

(i) 重み付きテーブルヒートマップ：領海内の資源分布を可視化するため，本研究で用いた重み付きテーブルを二次元グリッド上にプロットした．図 2 に重み付きテーブルヒートマップを示す．各セルは濃淡色で色付けされており，重みが大きい，すなわち，資源が集中しいるセルは濃色で表している．

(ii) 初期配置ヒートマップ：巡視船の初期配置を可視化するため，巡視船がエピソード毎に初期化された位置をカウントし，その頻度を二次元グリッド上にプロットする．巡視船の初期配置は，他の巡視船と領海エリア（セル）が重複しないように設計されている．図 3 に初期配置ヒートマップを示す．図より，巡視船の初期配置は広範囲に分散されていることが視覚的に確認できる．さらに，この初期配置ヒートマップと，図 2 の重み付きテーブルヒートマップとの相関関係を分析した結果，巡視船は重みが大きい領海エリアに初期配置される傾向があった．具体的には，初期配置のカウント数が最も大きいエリア $(x = 5, y = 2)$ は，図 2 の重み付きテーブルヒートマップの $(x = 4, y = 1)$ および $(x = 6, y = 0)$ の近傍であることが分かる．このことは，巡視船の初期配置が資源（重み）に基づいて設計されており，重みの大きい領海エリアを優先的に巡視していることを示している．

(iii) 移動軌跡ヒートマップ：巡視船の移動軌跡を可視化するため，巡視船が監視行動を行った軌跡をカウントし，その頻度を二次元グリッド上にプロットする．図 4 に移動軌跡ヒートマップを示す．図より，巡視船の移動軌跡は重みが大きいセル，すなわち，図 2 の重み付きテーブルヒートマップにおける重みが大きい高資源のセル（領海エリア）と，密漁船がスポーンするセル（領海エリア）に集中しており，巡視船が密漁船および資源分布を考慮した巡視戦略を学習していることが確認された．

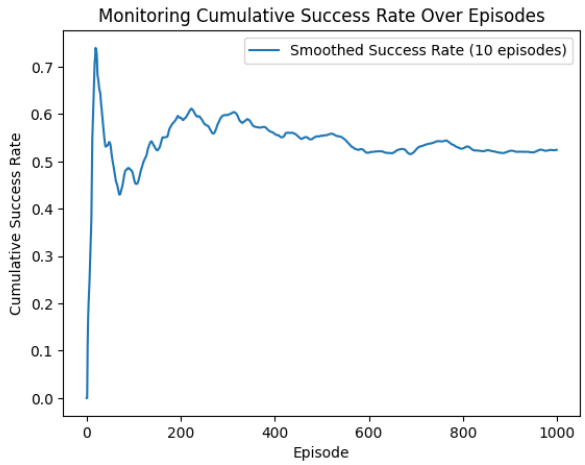

図 5: エピソードと成功率

次に，提案手法の有効性を評価するために，巡視船が密漁船を監視できた割合を成功率として解析した．図 5 にエピソードと成功率の結果を示す．ここで，x 軸はエピソード数，y 軸は成功率を表す．図より，シミュレーションの初期段階では，成功率が低いが，エピソードが進むにつれ，成功率が上昇していることが分かる．この結果から，巡視船が深層強化学習を通じて巡視戦略を効率的に学習していることが確認される．例えば，エピソード 100 時点での成功率は約 45%であったが，エピソード 250 時点での成功率は約 60%に達していた．このことは，巡視船がエピソードを重ねるごとに密漁船の行動を効果的に予測し，監視能力を向上させていることを意味する．加えて，成功率の増加傾向に関しては，巡視船が密漁船の戦略的な回避行動にも適応していることを表している．図 4 の巡視船の移動軌跡ヒートマップから，エピソード後半においても，巡視船が高資源の領海エリアを重点的に巡視し，密漁船の監視回避行動に対応できていることが確認された．この成功率の上昇傾向は，提案手法が動的な環境からのフィードバックを効率的に活用し，動的かつ柔軟な巡視戦略を学習していることを示している．以上より，提案手法では，エピソード中の報酬設計により，巡視船の短期的な巡視戦略ではなく，監視成功率が高くなるような長期的な巡視戦略を学習していることが分かった．

5. 結言

海洋安全保障の強化は海洋国家である日本をはじめ，国際社会全体の平和と繁栄に不可欠である．海洋安全保障の強化に欠かせないのが海上警備であり，海上での人命・財産の保護を含む治安維持，領海内へ侵入しようとする不審船の発見・追跡・対処等が主な活動として挙げられる．海上警備では，刻一刻と変化する動的環境下において，アジリティの高い巡視計画が求められる．本研究では，刻一刻と変化する動的環境において，不審船に対する巡視船の監視能力の向上を目的とし，深層強化学習を用いた巡視船の巡視戦略学習モデルを提案した．シミュレーションの結果，巡視船は密漁船の行動を予測して発見することができた．また，巡視戦略の学習は，エピソードを重ねるごとに安定的に向上し，動的環境下においても一貫した成果を示した．これらの結果は，深層強化学習の適用の有効性を示しており，提案手法の妥当性を裏付けるものであった．

本研究は，海上警備において，動的環境下における巡視戦略の新たな可能性を提供する一方で，いくつかの課題も挙げられる．今後の課題として，まず，提案モデルでは，各巡視船が独立して行動しているが，巡視船同士が情報を共有する仕組みを導入することで巡視効率の向上が期待される．情報共有による協調戦略は，巡視船間の役割分担や巡視範囲の最適化を可能にし，密漁船に対する対応力を強化するものであると考える．次に，実際の運用では，巡視船毎に性能が異なることを考慮する必要がある．例えば，巡視範囲や航行速度が異なる船を最適に組み合わせることで，巡視効果を最大化しつつ，出航コストを最小化する戦略を構築することが求められる．最後に，仮想環境に基づく現在のシミュレーションを拡張し，実際の密漁船の行動データや海洋条件を取り入れることで，提案手法の実用性を向上させることが重要であると考える．

参考文献

[1] I. Çapar, B. Keskin, and P. Rubin. An improved formulation for the maximum coverage patrol routing problem. *Computers & Operations Research*, 59:1–10, 2015.

[2] X. Chen, S. Wu, Y. Liu, W. Wu, and S. Wang. A patrol routing problem for maritime crime-fighting. *Transportation Research Part E: Logistics and Transportation Review*, 168:102940, 2022.

[3] P. Chircop, T. Surendonk, M. van den Briel, and T. Walsh. On routing and scheduling a fleet of resource-constrained vessels to provide ongoing continuous patrol coverage. *Annals of Operations Research*, 312:723–760, 2022.

[4] B. Keskin, S.-R. Li, D. Steil, and S. Spiller. Analysis of an integrated maximum covering and patrol routing problem. *Transportation Research Part E: Logistics and Transportation Review*, 48(1):215–232, 2012.

[5] T. Surendonk and P. Chircop. On the computational complexity of the patrol boat scheduling problem with complete coverage. *Naval Research Logistics*, 67:289–299, 2022.

[6] T. Zhu, Y. Xiao, and H. Zhang. Maritime patrol tasks assignment optimization of multiple USVs under endurance constraint. *Ocean Engineering*, 285:115445, 2023.

[7] 沖本天太. 海上警備：ロバストな巡視船再配置問題. **日本オペレーションズ・リサーチ学会**, pages 172–173, 2024.