マルチエージェント継続協調巡回における分業の創発と変化に対する柔軟性の評価

杉山 歩未　Sea Vourchteang　早野 真史　菅原 俊治

近年の情報通信やセンサ技術の発達により，動的かつ大量に発生する情報やタスクを短時間で処理し，効率的な作業を実現することが望まれている．このような問題では単一のロボットや計算機では性能に限界があり，それらを複数組み合わせたシステムによって効率を向上させることが期待される．一般に，複数者での作業では分業を適切に行うことが効率化のために有効であるが，複雑かつ動的に変化する環境では適切な分担を事前にエージェントに割り当てることは困難である．そこで，エージェントが自律的にタスクの分担を決めることが期待されるが，複数の自律エージェントがシステム全体の効率と通信コストのトレードオフを考慮しながら効率的な分業体制を創発するためのメカニズムは明らかになっていない点が多い．本研究では，継続協調巡回問題を例に我々が提案した単純な交渉と自律的学習を組み合わせることで分業の創発を促進する手法に対し，分業創発の要因についての分析と本手法が異なる種類の環境の変化にどのように対応するかを調査し，自律エージェント群が変化に柔軟で持続的なシステムを構築するためのメカニズムについて考察を行う．結果から，本手法が複数の役割分担を自然に発生させ，それらの役割分担や異なる粒度の学習が混在することが効率的かつ柔軟な分業体制を創発させることを示唆した．

1　はじめに

動的に発生するタスクや環境の変化に柔軟に対応可能な自律エージェントによるシステムが多くの分野で期待されている．近年のロボットやセンサなどのデバイスやネットワークの発達により，動的に発生する多数のタスクやデータを，要求される時間内で処理するシステムが求められている．このような要求を解決するためには，ロボットやプログラムを制御する多数のエージェントが自律的に対応することが重要となる．しかし，多数の自律エージェントによる作業において，エージェントが自己の効率の最大化を目指しながらもエージェント間でも競合を回避しながらシステム全体の効率を高い状態で持続させるメカニズムは，エージェント間の相互作用の複雑さから未だ明らかとなっていない．

本研究では，継続協調巡回問題 (Continuous Cooperative Patrolling Problem, CCPP) を定義し，複数の自律エージェントが効率的な分業体制を創発するメカニズムと，そのようなボトムアップに作られた体制の，変化に対する柔軟性を評価する．[3] は近年のマルチロボットによる巡回アルゴリズムの比較をおこなっており，複数者による巡回問題に求められる拡張性や変化への柔軟性，通信の制約などの特徴から，複数エージェント巡回問題で得られた知見が巡回問題だけでなく，マルチロボットシステム全体で有効活用できると主張している．既存の巡回問題に関連する研究の多くは，予め決められた各ノードへの訪問要求を満たしつつ，各ノードへの訪問間隔を最小化することを目的としたものが多かった．しかし，現実の多くの問題では，処理・検出すべきイベントは確率的に発生し，その発生確率も動的に変化する場合も多い．十分な時間があれば，既存のアルゴリズムによって最適なルートを再計算することで対応可能だが，持続的なシステムのためにはそのような動的な変化にもシステムを停止させずに，自律的に対応することが望まれる．そこで，CCPP では既存の巡回問題の設定を拡張し，継続的な活動のために充電やサーバー

Ayumi Sugiyama　Vourchteang Sea　Masashi Hayano　Toshiharu Sugawara, 早稲田大学基幹理工学研究科情報理工・情報通信専攻, Dept. of Information and Computer Science, Waseda University.

のメンテナンス等を考慮してエージェントが定期的に休止する必要のあるモデルを提案している．また，現実のように通信範囲や通信間隔に制限を設け，そのような制約の元で自律的な複数のエージェントが効率的な巡回を継続するためのメカニズムを考察する．

本稿では，分業の創発を促進する我々の提案手法 [14] [11] [10] [12] に対し，環境の急激な変化に対する追従性を評価する追加的な実験を行い複数の自律エージェントがそのような変化に効率よく対応するための要素を分析する．[12] では，途中でエージェントの一部が停止するという変化に対する追従性の評価を行った．結果から，提案手法によってエージェント間で主に担当する場所や巡回の方法などにおいて役割分担を創発し，あらかじめ分業の割り当てを与えられていないにも関わらず効率的な分業体制を構築し，その役割を柔軟に変化させることで変化に柔軟なシステムを維持できることが分かった．しかし，エージェントの停止はシステムの内部的な変化であったため，環境のイベント発生確率の急激な変化などの外部的な変化に対しての考察は行われていなかった．既存の多くの巡回問題に関する研究においても，このような二面的な変化に対してシステムの継続性・柔軟性の点で深く考察されたものは見受けられていない．本研究では，イベント発生確率が急激に変化するという追加的な実験の結果から，これまでに得られていたエージェントの停止という変化に対応するメカニズムとは別のメカニズムによって，環境の急激な変化に適応していることが明らかとなった．そして，本実験により得られた結果とこれまでに得られていた知見との統合的な分析を行い，異なる種類の変化にも柔軟に対応するためのメカニズムについての考察を行った．

2 関連研究

複雑な問題を複数人で効率的に行うために分業を行うことは，古くからその有用性が認識されている．産業革命時に Adam Smith は分業について言及し，分業の利点として (1) 特定の作業への特化による熟練度の向上 (2) 作業の変更に伴うコストの削減 (3) 各作業の効率化のための道具開発の促進の 3 点を述べている [9]．また，採餌問題もエサをイベントととらえると巡回問題と共通する要素が多くあり，複数の自律的な行動主体による分業についての研究が行われている [6] [13] [5] [7]．しかし，これらの研究では各エサに割り当てるエージェントの最適な比率が与えられており，どのようにそれぞれのエージェントが割り当てられるかということに主眼を置いていた．

巡回問題においては分業を実現するために大きく分けて 2 つのアプローチがある．1 つは環境をいくつかのサブエリアに分割し，各サブエリアを一体もしくは数体のエージェントに割り当てる領域分割手法 [1] [4] であり，もう一方は環境を分割せずに，効率的な巡回のための目標や経路を学習する手法 [8] [2] である．領域分割によるアプローチは比較的簡単に競合の発生を防ぐことが可能であり，有用なアプローチである．しかしながら，このような領域分割のアプローチではサブエリアの分け方や各領域に割り当てる最適なエージェントの比率があらかじめ与えられており，領域が連続しているゆえに環境の変化時に収束が遅く柔軟性に欠けるといった課題がある．そこで我々は後者のアプローチに注目し，[14] では清掃問題を題材にいくつかの目標決定戦略から最適な戦略を学習する Adaptive meta target decision strategy(AMTDS) を提案し，[11] ではエージェント間の交渉がなくとも環境のイベント発生確率を学習することで間接的に他者のふるまいを考慮する AMTDS with learning of dirt accumulation probability(AMTDS/LD) を提案した．また，[10] では清掃問題から巡回問題にモデルを一般化して継続協調巡回問題として再定義し，責任をもつ範囲という概念をエージェントに導入し，非常に簡単な交渉手法によって互いのイベント発生確率に関する情報を交換することで，分業創発をボトムアップに促進させる手法 (AMTDS with learning of event probabilities and enhancing divisional cooperation, AMTDS/EDC) を提案した．

しかし，これらの研究では効率性と通信時のコストにのみ着目しており，[3] が主張した巡回問題における課題である変化への柔軟性については考慮されていなかった．そこで我々は [12] でエージェントの一部が停止するという内部的な変化に対して我々の提案手法がどのように対応できるかを調査し，提案手法によっ

て発生した役割をエージェントが自律的に変化させることで，変化に対して効率の低下を防ぐ柔軟なシステムとなることを示した．しかし，[12] は創発された役割分担が変化にどのように対応するかというメカニズムの初期的な実験であり，もう一つの大きな変化である外部的な環境の変化に対しての考察はされていなかった．そこで本研究では，追加的な実験として環境のイベント発生確率が急激に変化するという外部的な変化への対応を調査する．手法自体は，異なる種類の変化に対する影響の違いに主眼をおくため [12] と同一のものを用いるが，結果から，異なる種類の変化に対しては異なるメカニズムがはたらいて変化への柔軟性を維持していることが分かった．本稿ではその違いを分析し，役割分担だけでなく異なる粒度の学習が共存していることで多様な変化への柔軟性をシステムが有することを示す．

3 モデル

3.1 環境

エージェントが巡回を行う環境を $G = (V, E) \subset \mathbb{R}^2$ とし，G はあらかじめエージェントに与えられていると仮定する．マップ生成も重要な課題の一つであるが，これに関しては多くの有用な研究がなされており，それらを活用することを想定している．$V = \{v_1, \ldots v_m\}$ をノードの集合とし，ノード v は座標 x_v，y_v をもつ．E はエッジの集合であり，エージェントはエッジで接続された隣接ノードへ移動が可能である．ただし，障害物 $R_o \subset \mathbb{R}^2$ がノード上に存在する場合，そのノードへは移動できない．また，すべてのノードはイベントの発生確率 $p(v)$ $(0 \leq p(v) \leq 1)$ をもち，高い値の $p(v)$ ほど頻繁な訪問を要求する．時刻 t にノード v において，検出されずに残存しているイベントの数を $L_t(v)$ と表す．$L_t(v)$ は $p(v)$ に基づいて毎 tick ごとに以下の式で確率的に更新する

$$L_t(v) \leftarrow \begin{cases} L_{t-1}(v) + 1 & (\text{イベント発生時}) \\ L_{t-1}(v) & (\text{その他}). \end{cases} \tag{1}$$

エージェントがノード v を訪問したとき，残存していたイベントは処理され $L_t(v) = 0$ となる．我々は離散時間 $tick$ を導入し，イベントの発生やエージェントの移動は各 tick 内で行われる．

CCPP における巡回の目標は，環境内に残存するイベント量を継続的になるべく小さく維持することである．そのため，本研究における巡回効率の評価指標を時刻 t_s から t_e 間の D_{t_s,t_e} として以下のように定義する

$$D_{t_s,t_e}(s) = \sum_{v \in V} \sum_{t=t_s+1}^{t_e} L_t(v), \tag{2}$$

ここで，$t_s < t_e$ であり，s はエージェントが選択した目標決定戦略である．$D_{t_s,t_e}(s)$ の値が小さいほど作業効率が良いとする．また，もう一つの評価指標として $U_{t_s,t_e}(s)$ を以下のように定義する

$$U_{t_s,t_e}(s) = \max_{v \in V, t_s < t \leq t_e} L_t(v). \tag{3}$$

$U_{t_s,t_e}(s)$ は環境内に蓄積した最大のイベント数を表す．実問題ではある一定以上の成果を保っているのであれば稼働台数を減らしたいなどのエネルギー消費を考慮したい場合も多くある．そこで，本モデルではその成果の指標となる最大蓄積イベント数による評価も提案している．

3.2 エージェント

エージェントの集合を $A = \{1, \ldots, n\}$ と定義する．時刻 t におけるエージェント i の位置を $v_t^i \in V$ と表現する．本研究では継続的な巡回活動を考慮するため，エージェントは有限容量のバッテリをもち，定期的に充電基地 v_{base}^i に戻り充電する必要がある．詳しい制御方法は [14] を参照されたい．また，エージェントはノード v が最後に訪問され，イベントが処理された時刻 t_{visit}^v を用いて，現在の時刻 t までの経過時刻 $I_t^i(v)$ を以下のように計算する

$$I_t^i(v) = t - t_{visit}^v. \tag{4}$$

ここで，本稿ではエージェントは赤外線素子や GPS 機能を利用することで，自分と他者の位置を把握できると仮定した．実問題においては位置情報を完全に取得できない場合や位置情報にノイズが発生する場合は考えられるが，今回は分業の創発と変化への追従性に焦点を絞るため，このような仮定をおいた．この仮定により，本稿では経過時刻 $I_t^i(v)$ は各エージェントにおいて他者が訪問した時刻を考慮した値となり，エージェント間で統一の値をもつ．しかし，$I_t^i(v)$ は各エージェントが独立にもつ値のため，他者の位置が

把握できない場合にも自然に拡張可能である．

エージェントは各ノード v の**重要度** $p^i(v)$ を独立に学習・記録し，その集合 $P_v^i = \{(v, p(v)) | v \in V\}$ をもつ．学習の詳細については 4 節にて後述する．重要度は数値で表現でき，環境において設定されているイベントの発生確率 $p(v)$ と異なり，各エージェントが独立にもつ，あるノードがどれだけ重要かという経験に基づく知識を表す．また，エージェント i は学習した各ノードの重要度 $p^i(v)$ を用いて，時刻 t におけるノード v の推定優先度 $EL_t^i(v)$ を次の式で定義する

$$EL_t^i(v) = p^i(v) \times I_t^i(v). \tag{5}$$

エージェントは戦略によっては推定優先度を基に訪問すべきノードを決定する．

本モデルでは一般的な通信の制約を考慮した通信制御を想定した．エージェントは通信範囲 d_{com} (> 0) を有し，エージェント i がエージェント j と通信ができるのはエージェント間のユークリッド距離 $m(v^i, v^j)$ を用いて，時刻 t において $m(v_t^i, v_t^j) < d_{com}$ の時とする．また，過度な通信によるコストの増加を防ぐため，i は j と最後に通信をおこなった時刻 $T_{last}^{i,j}$ を保持し，最低通信間隔 $T_{limit}^{i,j}$ (> 0) を定義し，$T_{limit}^{i,j} \geq t - T_{last}^{i,j}$ のときは通信を行わないと仮定した．

4 提案手法

本研究とも密接に関わるため，ここで我々の提案した自律的な学習と 1 対 1 間の交渉というミクロな行動戦略からシステム全体のマクロな分業創発を促進させる AMTDS/EDC 手法[12] について説明する．本手法では，エージェントは複数の行動戦略から適切だと思われる戦略を学習しながらも，その戦略の決定に影響する環境に対する知識を表す $p^i(v)$ も学習する．さらに，経験から学習した $p^i(v)$ に関する知識を簡単な交渉手法によって他者と交換し，責任をもつ範囲の改善を試みる．

4.1 目標決定戦略

エージェントの巡回は (1) 戦略 s に従い目標ノード v_{tar} を決定．(2) v_{tar} までの経路を生成．(3) 生成した経路を通り v_{tar} に到着．という流れを繰り返すことで行う．以下でエージェントがとりうる戦略のコンセプトを簡単に述べる．

ランダム法 i は環境全体からランダムに目標を選択する．

貪欲法 i は推定優先度 $EL_t^i(v)$ の上位 N_g ノードから目標を選択する．

訪問間隔優先法 i は環境全体から最後に訪問されてからの間隔 $I_t^i(v)$ の上位 N_i ノードから目標を決定する．

戦略型目標決定法 i は近隣の巡回を優先するかの閾値 E_B を学習し，近隣の推定イベント蓄積数が E_B 以下になるまで近隣を優先して巡回する．

[11] では，エージェントが学習した各ノードの知識も利用して，これらの戦略から適切な戦略を学習する手法として AMTDS/LD を提案している．AMTDS/LD においてエージェント i は任意の戦略の集合 $S = \{s_1, \ldots, s_n\}$ から最適と思われる戦略 s^* を学習しながら巡回を行う．各エージェントは $p(v)$ を未知とし $p^i(v)$ の学習を行う．後述する本提案手法の AMTDS/EDC においても，同様の戦略学習を行うが，AMTDS/LD では計算コストを考慮してエージェント間の交渉は行っておらず，単一エージェントによる学習と間接的に環境から他者のふるまいを学習するのみとなっている．経路生成手法については今回の主目的と外れるため [14] を参照されたい．

4.2 重要度の学習とエージェント間交渉

交渉においては公平性のための交渉と改善のための交渉を行う．公平性のための交渉では，責任をもつタスク量の多いエージェントが少ないエージェントにいくつかのノードの責任を委託するよう打診し，打診されたエージェントは自分の方が少ないコストでそのノードを巡回できると考えられるのであれば責任の一部を引き受ける．この仕事量の分配によって，両方のエージェントが一部のノードへの集中と広範囲への探索を行えるようになる．改善のための交渉では，責任量の同等なエージェント同士が，互いにより少ないコストで巡回できるノードがあればその責任を交換する．2 つの交渉手法において，エージェントは責任の委譲を打診されたノードが，自分の方が少ないコストで巡回できると思われる場合のみその打診を受

諾する．たとえば，エージェント i の責任量が他者に比べて多い場合でも，他にその作業をより小さいコストで行えるエージェントがいない場合，交渉は成立しない．経済学における比較優位の法則のように，あるエージェントの責任量が大きくなるほど他のエージェントがそのエージェントの仕事を代わることによるインセンティブは増加する．これにより，自分の現在の状況を考慮しつつ，単に 1 対 1 間で責任を分けるのではなくシステムの全体にとって利益となるように情報を交換することで，互いに多くのイベントの回収を目指せるというインセンティブをもつ．また，提案する交渉手法は他のオークションベースなどの交渉プロトコルと異なり，交渉はあくまで責任の委託であり，あるノードをどちらのエージェントがどれくらい担当するかは交渉によって完全には決められず，各エージェントが自分の知識に基づき決定することに注意されたい．

4.2.1 重要度の学習と責任をもつ範囲

エージェント i はノード v に訪れたとき，$I_t^i(v)$ を用いて $p^i(v)$ を次式で更新する．

$$p^i(v) \leftarrow \begin{cases} (1-\beta)p^i(v) + \beta\dfrac{1}{I_t^i(v)} \ (L_t(v) > 0 \text{ のとき}), \\ (1-\beta)p^i(v) \ (L_t(v) = 0 \text{ のとき}), \end{cases} \tag{6}$$

ここで，α $(0 < \beta \leq 1)$ は学習率である．

また，エージェント i が責任をもつ範囲 V_{self}^i $(\subset V)$ を導入する．責任をもつとは，戦略 s による目標ノード v_{tar}^i の候補を環境全体の V でなく，V_{self}^i に限定することで巡回対象の選択と集中を行うことをあらわす．ただし，訪問間隔優先法とランダム法は環境内で見逃されているノードの探索が目的であるため，これらの戦略では v_{tar}^i を V から選択する．エージェント i は充電基地に戻った際，V_{self}^i を更新する．V_{self}^i の初期値は環境全体とし，$V_{self}^i = V$ とする．V_{self}^i のサイズを N_{self}^i とし，この初期値も $N_{self}^i = |V|$ となる．充電基地に戻った際，i は P^i の要素を $p^i(v)$ の降順に並び替え，上位 N_{self}^i のノードを V_{self}^i として再定義する．$p^i(v)$ が同値のノードが複数あった場合はランダムに決定される．

本研究では交渉にかかるコスト抑制のため，交渉には V_{self}^i から計算される 2 つの値のみを用いる．1 つは責任をもつノードの重要度の総和 p_{sum}^i (≥ 0) であり，以下のように計算する

$$p_{sum}^i = \sum_{v \in V_{self}^i} p_t^i(v).$$

重要度 $p_t^i(v)$ が高いノードは頻繁な訪問を求められるので，p_{sum}^i は i が責任をもつ作業量を表現している．2 つめは作業の重心 $C^i = (x_c^i, y_c^i)$ であり，$p^i(v)$ の重みを考慮して以下のように設定する

$$x_c^i = \sum_{v \in V_{self}^i} \frac{p^i(v)}{p_{sum}^i} x_v, \text{ and } y_c^i = \sum_{v \in V_{self}^i} \frac{p^i(v)}{p_{sum}^i} y_v. \tag{7}$$

ここで，$v = (x_v, y_v)$ である．ノード v_p からノード v_q への最短経路長を $d(v_p, v_q)$ とすると，$d(v_c^i, v) < d(v_c^j, v)$ のとき，ノード v はエージェント i の作業の中心にエージェント j よりも近く，訪問するために払うコストが小さいことをあらわす．

4.2.2 公平性のための交渉

エージェントは他者と交渉を行い，各々の V_{self}^i の要素の改善を試みる．この際，通信コストを考慮し，p_{sum}^i と C^i のみの比較によって交換する情報を決定する．公平性のための交渉では，自分より少ない責任をもつ他者に自分にとって重要でないタスクを任せることで，自分は重要な作業に集中し相手は広く探索を行えるようにすることが狙いである．式 (8) の条件を満たす場合，担っている責任に差があると判断し公平性のための交渉を行う

$$1 + T_c < p_{sum}^i / p_{sum}^j. \tag{8}$$

T_c $(0 < T_c \ll 1)$ は責任の差を判定する閾値である．この時，i は $p^i(v)$ の降順に集合 $V_{self}^{i,j} = \{v \in V_{self}^i \mid d(C^i, v) > d(C^j, v)\}$ を計算する．i は $V_{self}^{i,j}$ の $p^i(v)$ の小さい値から，$e_g (0 \leq e_g)$ 個のノードを選択し，その $p^i(v)$ の値を j に次式によって渡す．

$$p^j(v) \leftarrow p^j(v) + p^i(v) \times \delta \tag{9}$$

$$p^i(v) \leftarrow p^i(v) \times \delta. \tag{10}$$

$\delta (0 < \delta < 1)$ は情報を受け渡す割合を表す．式 (9), (10) に示す $p^i(v)$ 値の受け渡しは，i が部分的に自分の責任を j に委託することを表す．e_g は交渉時に情報を受け渡されるノードの数であり，p_{sum}^i と p_{sum}^j の比率から以下のように決定する．

$$e_g = \min\left(N_{self}^i - 1, N_{gmax}^i, \left\lfloor \frac{p_{sum}^i}{p_{sum}^j} \times \gamma \right\rfloor\right), \tag{11}$$

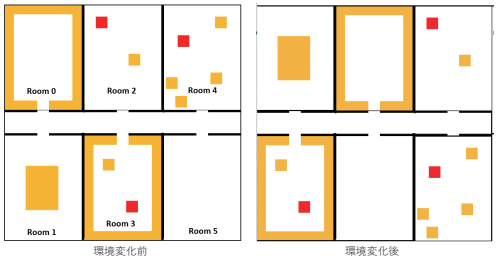

図 1　実験環境

ここで，N^i_{gmax} $(0 < N^i_{gmax} \leq N^i_{self})$ は急激な変動を防ぐための上限であり，γ は受け渡すノード数を調整するパラメータである．情報の受け渡しが終わった後，エージェント i と j は V^i_{self} のサイズ N^i_{self} を次式で更新する

$$N^i_{self} \leftarrow N^i_{self} - e_g \tag{12}$$

$$N^j_{self} \leftarrow \min(|V|, N^j_{self} + e_g). \tag{13}$$

4.3　改善のための交渉

もう 1 つの交渉は，責任をもつ作業量が均衡したときに，責任をもつノードを交換することで互いの責任を持つノードの改善を試みる交渉である．この交渉では，お互いに相手の方がより少ないコストでできる作業を交換することで，大きな変動を与えずに互いの巡回コストを低下させ，効率を向上させることが狙いである．式 (14) の条件を満たす場合に以下の改善のための交渉を行う．

$$1 - T_c < p^i_{sum}/p^j_{sum} < 1 + T_c. \tag{14}$$

i は $V^{i,j}_{self}$ の大きい $p^i(v)$ の大きい値から e_g 個のノードの $p^i(v)$ を式 (9), (10) に従い j に渡す．e_g は次式で決定する

$$e_g = \min\left(N^i_{self} - 1, N^i_{cmax}\right), \tag{15}$$

ここで，$N^i_{cmax}(> 0)$ は上限である．情報の受け渡しが終わった後，エージェントは N^i_{self} を式 (12) に従い更新する．交換する $p^i(v)$ の値を考慮し，N^i_{cmax} は N^i_{gmax} と比較して小さな値とする．式 (14) の条件を満たすとき，上記の過程は i と j で双方向に発生する．

表 1　実験パラメータ

目標選択戦略	パラメータ	値
貪欲法	N_g	5
訪問間隔優先法	N_i	5
AMTDS/LD	α	0.1
	β	0.05
	ε	0.05
AMTDS/EDC	N_{gmax}	100
	N_{cmax}	10
	T_c	0.05
	γ	10
	δ	0.5
通信の制約	d_{com}	5
	T_{limit}	10800

5　評価実験

5.1　実験設定

提案手法により創発された分業と環境の変化への追従性の関係性を調査するためにシミュレーションによる評価実験を行った．ここでは，[12] で示されたエージェントの停止という変化への対応と比較して，異なる種類の変化，具体的には環境のイベント発生確率の急激な変化に対応するための要素を分析する．本研究でエージェントが巡回を行う環境は，各ノードのイベント発生確率 $p^i(v)$ に偏りがあり，障害物によっていくつかの部屋 $Room\ N$ $(N = 0, \ldots, 5)$ に分かれている 101×101 の大きさのグリッド構造とした．その環境を図 1 に示す．各ノードの $p(v)$ の値は図 1 に示す色によって赤色部：10^{-3}，オレンジ色部：10^{-4}，その他：10^{-6} とし，障害物 R_o は黒色で表している．また，本研究の目的のため，環境のイベント発生確率を途中で大きく変化させる．100 万ステップ時に図 1 に示す環境変化後のように，各部屋のイベント発生確率の特徴を時計回りに 1 部屋ずつずらし，200 万ステップ時に元の状態に戻した．[12] では，イベント発生確率の変化の代わりに 100 万ステップ時にランダムに 10 のエージェントを停止させ，200 万ステップ時に停止したエージェントを復帰させている．

エージェント数は 20 で，全エージェントの充電基

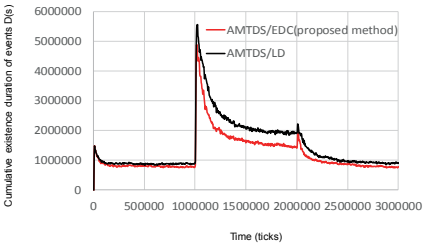

図 2　環境変化時の巡回効率 $D(s)$ の推移

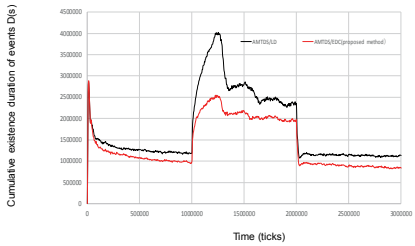

図 3　エージェント停止時の巡回効率 $D(s)$ の推移

地は $v_{base} = (0, 0)$ とした．全てのエージェントは充電基地からスタートし，充電のために定期的に戻る必要がある．エージェントのバッテリの設定は容量を 900，容量 0 からの満充電に 2700tick を要し，隣接ノードへの移動は 1 消費すると設定した．よって，エージェントの巡回サイクルは最大 3600tick となり，評価指標の $D_{t_s,t_e}(s)$ も 3600tick ごとに集計した．全てのエージェントは戦略として AMTDS/LD か提案手法の AMTDS/EDC のいずれかを用いる．ここで，AMTDS/LD はエージェント間交渉は行わないが環境の学習のみによって間接的に他者のふるまいを学習するものであり，単一エージェントの学習フレームワークに近い手法である．AMTDS/LD でもわずかながら主に訪問する場所の偏り，いわゆる分業が発生することが確認されているがその程度は小さいものであった．また，モデルおよび提案手法で述べたパラメータは表 1 のように設定した．本研究では [12] で得られたエージェントの停止に対する結果と，今回行う環境の急激な変化に対する結果を比較するため，表 1 のパラメータは [12] で用いたものと同様の値を用いた．

5.2　結果と考察

環境変化時およびエージェント停止時の評価指標 $D(s)$ の推移を図 2, 3 に示す．両方の変化に対して，提案手法である AMTDS/EDC は効率の悪化を AMTDS/LD に比べて防いでおり，$D(s)$ も早く低下させていることが見て取れる．詳細は [12] にて述べているが，エージェント停止時に提案手法である AMTDS/EDC が大きく効率低下を防いでいるのは，

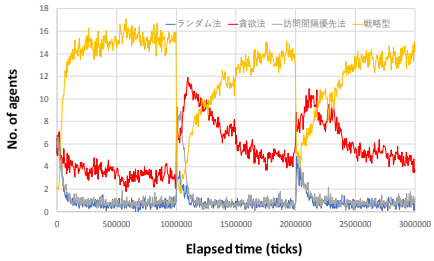

図 4　イベント発生確率変化時の選択戦略の推移 (AMTDS/LD)

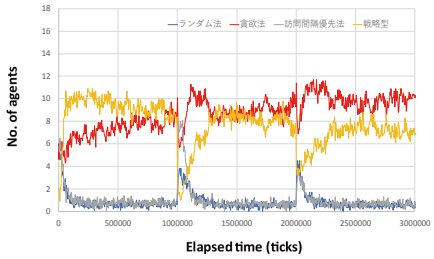

図 5　イベント発生確率変化時の選択戦略の推移 (AMTDS/EDC)

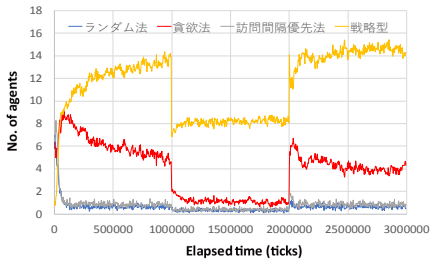

図 6　エージェント停止時の選択戦略の推移 (AMTDS/LD)

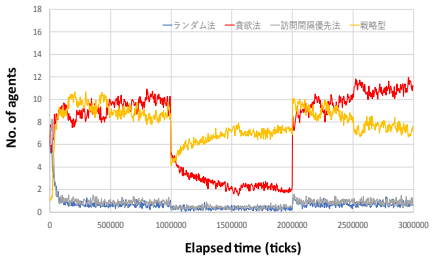

図 7　エージェント停止時の選択戦略の推移 (AMTDS/EDC)

提案手法によって主に担当する範囲に大きな偏りが発生するのと同時に何台かのエージェントは似たような場所を巡回するいわゆるチームのような働きが創発されたこと，責任の量はほぼ同等ながら責任を持つ範囲にも偏りが生じ，一部のエージェントは特定の範囲に集中してその場所を精度良く巡回するスペシャリストのような役割を，別のエージェントは広範囲を巡回するゼネラリストのような役割を担うようになったことが要因である．これらの役割は提案手法の交渉によって決定的に定まるのではなく，エージェントが自律的に調整を行うことで変化時には役割を柔軟に変更させ，素早く変化に対応することが可能となった．

しかし，本実験では [12] で得られた役割分担だけでなく，素早い戦略の転換によって変化に対応したことがみてとれた．環境変化時の AMTDS/LD および AMTDS/EDC によって各エージェントが選択していた目標決定戦略を図 5,4 に，エージェント停止時の AMTDS/LD および AMTDS/EDC によって各エージェントが選択していた目標決定戦略を図 7,6 に示す．[12] では AMTDS/LD では近領域を優先する戦略型が優勢であったが AMTDS/EDC では各エージェントが連続しないノードに対しても責任範囲を決定でき，その知識を利用することで近領域以外にも柔軟に巡回する貪欲法で高い効率をだせていた．しかし，図 6 と図 7 をみて分かるように，両手法においてエージェント停止時にも貪欲法か戦略型をほぼ全てのエージェントが選んでいた．ところが，図 4 と図 5 をみると，AMTDS/LD では戦略型が優勢なことは変わらないが，環境変化時には両手法において変化の直後に訪問間隔優先法が急激に増え，その後貪欲法と戦略型が増加している．今回比較した 2 つの変化において最も異なる点は，エージェント停止時は残されたエージェントにとって追加分のイベント発生領域が増えるが，エージェントの台数は減っており考慮する他者の数は減ったが，環境変化時にはこれまでに学習した各ノードのイベント発生確率の増減の修正と，新たに発生したイベントに対する他者のふるまいの変化の両方に追従する必要がある点である．つまり，環境変化時の方が新たに学習し直す要素が多く，学習に時間がかかると考えられる．本実験では，[12] では見られなかった戦略の素早い変更が AMTDS/LD および AMTDS/EDC で起きている．これは，環境変化時には各ノードの重要度の再学習に時間がかかるため，戦略という大きな方針が先に学習されることで，知識の活用を行う貪欲法や戦略型から，環境の探索を行う訪問間隔優先法への変更が促され，これにより変化に対応したと考えられる．それに加え，AMTDS/EDC では訪問間隔優先法にて探索優位になった状態から，[12] で述べたようにそれぞれが探索した新しい環境の知識をより低コストで巡回できるエージェントに委託しながら役割を柔軟に変化させることで，スペシャリストによる各ノードに対するイベント発生確率の学習が促進され，近領域を優先する—ある種の連続領域の領域分割のような巡回をする— AMTDS/LD よりも早く変化に対応できたと考えられる．つまり，環境変化に対してはエージェントの停止時と異なり，大きな方針である戦略学習と各ノードに対するイベント発生確率の学習という小さな粒度の学習という 2 種類の粒度の異なる学習が存在することで，変化に柔軟に対応しているということが分かった．

[12] において多様な役割が存在し，それを交渉によって決定するのではなく自律的に決めることで柔軟な役割の変更を可能とし，変化への柔軟性を向上しているという知見を得ていたが，本研究において異なる環境の変化に対する実験を追加的に行いその差異を比較することで，戦略学習という大きな行動決定のための方針の学習と，イベント発生確率の学習とその再交渉という各戦略を遂行する際に参考とする情報の学習という小さな粒度の学習をエージェントが行っ

ていることで，変化やエージェントによる変化の検知に応じて役割や方針を自律的かつ柔軟に変更することでシステムの効率性を維持しながら柔軟性をも向上させることを示唆している．

6 結論

本研究では，高度な自律性と持続可能性・複数エージェントによる協調が求められる問題の 1 つとして CCPP を定義した．既存研究における内部的な変化と異なる，外部的な変化を評価する追加的な実験を行い，異なる種類の環境変化への追従性を調査してその要因を分析することで，創発された分業がシステムの柔軟性向上に与える要因を考察した．具体的には，提案手法はエージェントの停止という変化にはボトムアップに発生した多様な役割分担を柔軟に変化させることで効率向上と変化への対応性の向上を実現するのに加え，環境のイベント発生確率の急激な変化という外部的な変化には，戦略の学習と戦略の内部で参照する情報の学習という異なる粒度の学習が混在することで，変化に対し迅速な対応をしつつ高い効率を発揮することを示した．提案手法ではこれらの異なる役割分担や異なる粒度の学習をエージェントが自律的に行うことで，通常時の効率を損なうことなく，それぞれの変化に対応した役割や方針の変化を自律的に行い変化に柔軟なシステムを持続することを示唆した．

今後は，[3] が主張したように，この研究において得られた知見や手法を一般的なマルチロボット・マルチエージェントシステムにおいて適用させ，一般性の高い手法の提案を試みる．また，結果から多様性を担保する自律的な決定が低コストの交渉ながらも効率と変化への柔軟性の向上をさせることが示唆されたが，これらの多様性とシステムの性能についてのより一般的な理論の発見が求められる．

謝辞 本研究は JSPS 特別研究員奨励費 JP16J11980 の助成を受けたものである。

参 考 文 献

[1] Ahmadi, M. and Stone, P.: A Multi-Robot System for Continuous Area Sweeping Tasks, *Proceedings of the 2006 IEEE International Conference on Robotics and Automation*, 2006, pp. 1724 – 1729.

[2] Cheng, K. and Dasgupta, P.: Dynamic Area Coverage using Faulty Multi-Agent Swarms, *Intelligent Agent Technology, 2007. IAT '07. IEEE/WIC/ACM International Conference on*, nov. 2007, pp. 17 –23.

[3] David, P. and Rui, R.: A Survey on Multi-robot Patrolling Algorithms, *Technological Innovation for Sustainability: Second IFIP WG 5.5/SOCOLNET Doctoral Conference on Computing, Electrical and Industrial Systems, DoCEIS 2011, Costa de Caparica, Portugal, February 21-23, 2011. Proceedings*, Camarinha-Matos, L. M.(ed.), Springer Berlin Heidelberg, 2011, pp. 139–146.

[4] Elor, Y. and Bruckstein, A. M.: Multi-a(ge)nt Graph Patrolling and Partitioning, *Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technologies*, 2009, pp. 52–57.

[5] Ikemoto, Y., Miura, T., and Asama, H.: Adaptive Division of Labor Control for Robot Group, *Proceedings of the 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Oct 2009, pp. 2409–2414.

[6] Jones, C. and Mataric, M. J.: Adaptive division of labor in large-scale minimalist multi-robot systems, *Proceedings of the 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003)*, Vol. 2, Oct 2003, pp. 1969–1974 vol.2.

[7] Lee, W. and Kim, D.: Adaptive Division of Labor in Multi-Robot System with Minimum Task Switching, *Proceedings of the Fourteenth International Conference on the Synthesis and Simulation of Living Systems (ALIFE 14)*, Jul 2014, pp. 750–756.

[8] Sampaio, P., Ramalho, G., and Tedesco, P.: The Gravitational Strategy for the Timed Patrolling, *Tools with Artificial Intelligence (ICTAI), 2010 22nd IEEE International Conference on*, 2010, pp. 113–120.

[9] Smith, A.: *An Inquiry into the Nature and Causes of the Wealth of Nations*, History of Economic Thought Books, No. smith1776, McMaster University Archive for the History of Economic Thought, 1776.

[10] Sugiyama, A., Sea, V., and Sugawara, T.: Effective Task Allocation by Enhancing Divisional Cooperation in Multi-Agent Continuous Patrolling Tasks, *2016 IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI)*, Nov 2016, pp. 33–40.

[11] Sugiyama, A. and Sugawara, T.: Meta-strategy for Cooperative Tasks with Learning of Environments in Multi-agent Continuous Tasks, *Proceedings of the 30th Annual ACM Symposium on Applied Computing*, 2015, pp. 494–500.

[12] Sugiyama, A. and Sugawara, T.: *Improvement of Robustness to Environmental Changes by Au-*

tonomous Divisional Cooperation in Multi-agent Cooperative Patrol Problem, Springer International Publishing, 2017, pp. 259–271.

[13] Suseki, K., Mizuguchi, T., Sugawara, K., and Kosuge, K.: Proportion Regulation for Division of Labor in Multi-Robot System, *Proeedings of the 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Aug 2005, pp. 2339–2344.

[14] Yoneda, K., Sugiyama, A., Kato, C., and Sugawara, T.: Learning and relearning of target decision strategies in continuous coordinated cleaning tasks with shallow coordination, *Web Intelligence*, Vol. 13, No. 4(2015), pp. 279–294.