研究会論文

Technical Report /

不均質な環境における拡張協調期待戦略の効率と特性

大塚 知亮
Tomoaki Otsuka
早稲田大学基幹理工学研究科情報理工・情報通信専攻
Department of Computer Science and Communications Engineering, Waseda University
oh-tsuka.21@suou.waseda.jp

菅原 俊治
Toshiharu Sugawara
（同　　上）
sugawara@waseda.jp

keywords: 囚人のジレンマ, 複雑ネットワーク, 強化学習, マルチエージェントシステム

Summary

本論文では, 複雑ネットワーク上での繰り返し囚人のジレンマゲームにおいて, 協調をネットワーク上に拡散させる協調期待戦略に裏切りエージェントの判定を追加した, 拡張協調期待戦略を提案する. 協調期待戦略は裏切り行動がナッシュ均衡である囚人のジレンマゲームにおいて, インタラクションによって相互協調が成立した際に, 周囲に協調が拡散したことを期待して, 一定回数連続して協調行動を行う手法である. この手法は相互評価や罰則などの本来のゲームに不要なインタラクションを発生させずに, 比較的簡単なアルゴリズムでネットワーク上に協調行動を拡散させることが可能な手法である. しかし, これまでの研究ではネットワーク上の全てのエージェントがこの協調期待戦略をとるエージェントと仮定しており, 単純な強化学習により裏切り行動に収束するエージェントが混在することで協調を維持できずに, 徐々に崩壊することがわかっている. そこで協調期待戦略に比較的簡単なアルゴリズムでインタラクションのみから裏切りエージェントと一時的に判定し, 裏切りエージェントに対して協調期間の協調を行わない拡張協調期待戦略を導入した. 実験から本戦略が頑健であり, 広く協調を拡散できることを示す.

1. は じ め に

近年のテクノロジーの発展により, コンピュータシステムは様々な用途で用いられるようになった. 特に, スマート端末や小型で高性能なコンピュータの普及と各種のサービスサーバとの連携により人間の代理として働くエージェントを様々な電子機器に搭載する技術が注目されている. こうしたエージェントはユーザの代理として情報の収集や共有, ユーザ間での交渉や調整を自動化することを目的としている. このとき, エージェントはエージェント間のインタラクションにおいて, 各ユーザにとって適切な動作と行動戦略を取ることが求められる. 従って, 協調・調整・競争などを通じて, 個々の条件を考慮しつつシステムが構成する社会全体のバランスをとった行動選択ができるように, エージェントが学習する機能が必要となる.

このような状況に対し, マルチエージェントシステム (MAS) の分野において, エージェント間の社会的協調や調整・取引を効率的に実現するために, 社会全体で整合性のある戦略や行動規範（ノルム）を周囲のエージェントとのインタラクションから学習する手法が提案されている [Shibusawa 14][Sen 10][Yu 13a][Yu 15][Jianye 15][Walker 95]. これらの研究では, 利用者間の関係をエージェントのネットワークで, インタラクションにおける利得構造をゲームでそれぞれをモデル化し, その環境下でエージェントが過去のインタラクションを強化学習によって最適な戦略を決定するアルゴリズムを提案している. しかし, これらの研究の多くは協調ゲームなどナッシュ均衡が社会全体においても最適解である問題を仮定しているものが多い.

エージェント間においても現実の人間社会と同様に, それぞれの利益を優先して行動することでエージェント間での競合が強くなり, 社会全体で利益を下げ合うジレンマ的状況が存在する. これは, 例えば, エージェントが個人や組織の代理として動作することや, エージェント自体の活動が共有するサーバや通信帯域を計算資源とすることに起因する. こうした社会的ジレンマをモデル化したゲームの 1 つが囚人のジレンマゲームである. 囚人のジレンマゲームでは,「裏切り (D)」行動がナッシュ均衡となるが, この状況はパレート効率的ではなく, 全員が「協調 (C)」行動をとることがパレート効率的な戦略である. こうした社会的ジレンマの状況は低価格競争や公共財に対する行動戦略など様々な現実世界の事象があてはまり, このゲーム構造において, エージェントが社会全体とし

て協調を促進できるメカニズムの解明や実装は非常に重要である.

そこで我々は大規模なエージェントネットワークにおいて, 無期限繰り返しの囚人のジレンマゲームを行い, エージェントがアルゴリズムとして単純な戦略の学習をすることで, ネットワーク全体に協調を促進させる協調期待戦略を提案してきた [渋澤 15]. ここでは, エージェントが自身のおかれた環境は社会的ジレンマ構造であると推定し, 協調行動がパレート効率的であると認識している状況を想定している. また, エージェントはネットワーク全体の状態（他のエージェントの戦略）を観測できず, 局所的なエージェントの行動のみを観測可能としている. そこで, エージェントは最近のインタラクションで相互に協調行動が成立した場合に, 周囲で協調行動が拡散していることを期待し, 一定期間 (協調期間と呼ぶ), 協調行動を試行する. この戦略の下, 我々は協調期間を 3 とすることで, 完全グラフ, Barabasi-Albert モデルによって構築したネットワーク (以降, BA ネットワーク), Connectng Nearest Neighbor モデルによって構築したネットワーク (以降, CNN ネットワーク) などで協調行動が全体に広がることを示した.

一方 [渋澤 15] では, 全てのエージェントが協調期待戦略を取ることを想定していたため, 協調期待戦略のエージェントとインタラクションの結果を Q 学習する (通常は均衡点 (D, D) を学習) ことで利得を最大化する Q 学習エージェントを一定数混在させ, 協調期待戦略の頑健性を調査した [大塚 16]. その結果, 協調期間 3 において, 完全グラフでは Q 学習エージェントが約 5%以上混在するとネットワーク全体での協調が維持できず, さらに正規ネットワークでは Q 学習エージェントが存在しない状態でも長期間にわたる協調は維持できないことがわかった.

そこで本論文では, この問題の解決方法として, 自分の協調期間において連続して裏切り行動を行うエージェントを裏切りエージェントと判定し, 協調期待戦略による強制的な協調を行わず, 裏切りエージェントの戦略変化に対応して協調に切り替える, 拡張協調期待戦略を提案する. 本戦略は単純ながらも裏切り行動中心のエージェントとは協調せず, 他方, 提案戦略同士の間では協調を続け, これがネットワーク全体に広がることを実験により示す. またその拡張期待戦略の特性と限界についても議論する.

本書の構成は以下のとおりである. まず, 次節で準備として囚人のジレンマゲームとネットワークモデル, エージェントの行動決定手法について説明する. 第 3 節では拡張協調期待戦略を提案する. 第 4 節では完全グラフと正規ネットワークを用いて評価実験を行い, 拡張協調期待戦略の影響を調査する. 第 5 節では実験結果の考察を述べ, 第 6 節で結論と今後の課題を述べる.

2. 準備と協調期待戦略

2·1 囚人のジレンマゲーム

囚人のジレンマゲームは 2 人ゲームで, プレイヤーは 2 種類の行動 $S = \{C(\text{協調}), D(\text{裏切り})\}$ を取る. 本研究では現実社会での継続的な利用を想定し, エージェントがゲームの試行回数を知らない, 無期限繰り返しの囚人のジレンマゲームを行う. また, 囚人のジレンマゲームにおける 2 種類の行動 C, D に対し, 以下の利得行列を定義する. 例えば, 2 プレイヤーがそれぞれ C,D の行動を選択

	C	D
C	R, R	S, T
D	T, S	P, P

図 1 囚人のジレンマゲームの利得行列

した場合, C を選択したプレイヤーは利得 S を, D を選択したプレイヤーは利得 T を得る. 囚人のジレンマゲームの利得行列では以下の関係が成り立つ. $T > R > P > S$ かつ $2R > T + S$. この条件により相手の行動にかかわらず, ナッシュ均衡は両者の行動が D となる. このとき両プレイヤーは利得 P を得るが, パレート効率的な状況は 2 プレイヤーが互いに C を選択した場合である. また, 本研究ではこれまでの研究 [渋澤 15] と同様に, 囚人のジレンマゲームの利得行列は $R = 3$, $S = 0$, $T = 5$, $P = 1$ とする.

2·2 ネットワークモデル

本節では本稿の実験で用いるネットワークと, それらが持つ特徴を説明する.

§1 完全グラフ

完全グラフはネットワーク内の全ノード間にリンクが貼られたネットワークである. n 個の頂点からなる完全グラフのリンク数は ${}_nC_2 = n(n-1)/2$ であり, 平均次数は $n-1$ である.

§2 正規ネットワーク

正規ネットワークは全てのノードが同じ数の隣接エージェントを持つネットワークである. ここでは正規ネットワークはノードを円周上に並べ, 各ノードからその両隣 m 個先のノードまでリンクを貼ることで正規ネットワークを構築した. そこで, n 個の頂点からなる正規ネットワークのリンク数は nm であり, 平均次数は $2m$ である.

2·3 エージェントの行動決定

本研究では, 以下の状況を想定する. まず, 協調期待戦略をとるエージェントは自己の利得行列のみを知る. さらに, 他のエージェントの利得行列は知らないが, 周囲も同じ行列を持つと仮定し, 自身のおかれた環境が社会的ジレンマ構造であり, 協調行動がパレート効率的である

と推定している. そのため周囲とともに協調し合うことが最善だが, 自分だけが協調すると損をする環境であることを知っている. また, どのエージェントもネットワーク全体の状態を観測できず, 隣接するエージェントとのインタラクションの結果のみを観測可能とする. このようなゲームで利得を最大化しようと行動するエージェントは実験の初期に拡張協調期待戦略によって行動する協調期待戦略エージェントか, 本節で述べる統合法によって行動する Q 学習エージェントかを決定し, 実験中に戦略の変更は行わない. どちらの戦略で行動を決定する場合も各エージェントはその隣接エージェントごとに, 強化学習によりこれまでのインタラクションの結果を学習し, 行動を決定する. また, 行動にランダム性を取り入れるためにインタラクションごとに ε-greedy による戦略の変更も行う.

§1 統合法による行動決定

統合法は, エージェント i が各隣接エージェント j に対する行動 s_j^i から自らの行動 s^i を決定するための手法である. また, エージェント i の隣接エージェントを N_i とする. エージェント i はラウンド t において, 以下に示す優先度 $p^i(s)$ の値が最大となる行動 $s^i \in S$ を選択する. 各行動で優先度が同値の場合はランダムに行動 s^i を選択する.

$$s^i(t) = \arg\max_{s \in S} p^i(s) \tag{1}$$

ここでは優先度 $p^i(s)$ を

$$p^i(s) = \sum_{j \in N_i} \delta(s, s_j^i(t-1)) \tag{2}$$

$$s_j^i(t-1) = \arg\max_{s \in S} Q_{t-1}^i(s, j) \tag{3}$$

と定義する. なお, 関数 δ は任意の戦略 s_1, s_2 について

$$\delta(s_1, s_2) = \begin{cases} 1 & \text{if } s_1 = s_2 \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

となるデルタ関数である. 従って, この場合, 統合法は多数決による決定となる.

§2 Q 学習エージェント

Q 学習エージェントは隣接するエージェントへの行動決定に Q 学習のみを用い, 協調期待戦略をとるエージェントと同様に統合法による行動決定を行い基本的な行動を決定する. ラウンド t において, エージェント i は隣接エージェント j に対して行動 s をとる時の Q 値 $Q_t^i(s, j)$ を以下の式に従って更新する. このとき, 前ラウンド $(t-1)$ におけるその戦略での利得を $r_{j,t-1}^i$ とする.

$$Q_t^i(s, j) = (1 - \alpha_i) \cdot Q_{t-1}^i(s, j) + \alpha_i \cdot r_{j,t-1}^i \tag{5}$$

エージェント数 300 の完全グラフとエージェント数 1000, 平均次数 10 の正規ネットワークにおいて, Q 学習エージェントのみで, 繰り返し囚人のジレンマゲームを行った結果の協調成立率を図 2 に示す. 協調成立率は 1 ラウンドに行われる全インタラクションのうち, エージェント間のインタラクションで両エージェントが協調行動をとった割合である. 図 2 を見ると完全グラフと正規ネットワークの両方でネットワーク全体で協調成立率が 0 になっている. 初期の段階では最初の行動がランダムに決定されるため, 相互協調が成り立つ 25%程度のインタラクションが協調になり, それを学習したエージェントが協調を維持しようとするが, 初期に裏切りを学習したエージェントによって, ネットワーク全体に裏切りが広がることがわかる.

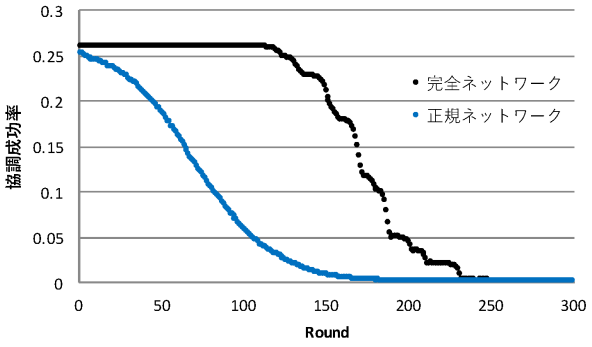

図 2 Q 学習エージェントによる繰り返し囚人のジレンマゲーム

§3 協調期待戦略

協調期待戦略 [大塚 16] をとるエージェント i に協調試行期間を表現するパラメータ L_i を導入し, $L_i > 0$ の時に協調期間にあるとする. また, 1 実験を通して変動しない協調期間定数 L を導入する. エージェントは最近のインタラクションから互いに協調行動が成立した場合に, 周囲で協調行動が促進されていることを期待して $L_i = L$ とし, $L_i > 0$ の期間, 協調行動を試行する. ここで L_i は正の整数とする. エージェント i は $L_i > 0$ のとき次の隣接エージェントのインタラクションで協調行動をとり, L_i の値を 1 減少させる. 協調期間であっても, 隣接エージェントと相互に協調行動をとった場合には $L_i = L$ と再設定し, 協調期間を延長する. この期間中は ε-greedy によるランダムな行動決定をせず, 協調行動を選択する. 協調期待戦略をとるエージェントも協調期間にない場合は Q 学習エージェントと同様に, 統合法に基づいて隣接エージェントへの戦略を決定する.

3. 提 案 手 法

協調期待戦略では協調期間中に一度でも他のエージェントと相互協調が成立すると協調期間を協調期間定数に再設定し, 本来の Q 学習の結果を学習した統合法による行動が裏切り行動であっても協調を持続する. そのため, 裏切り行動に収束した Q 学習エージェントが混在すると協調期待戦略エージェントの協調行動の Q 値が下がり, さ

らに協調期間が途切れると協調期待戦略エージェントも裏切り行動に収束する. そこで, 本節では裏切りエージェントを判定し, 裏切りエージェントと判定したエージェントに対する行動を変化させる手法を提案する. ここで述べる手法による戦略を拡張協調期待戦略とする.

3·1 裏切りエージェントの判定

拡張協調期待戦略エージェント i は裏切り判定閾値 $M \geq 0$ を導入し, その隣接エージェント j に対して裏切り行動のカウンター dc_{ij} を持つ. dc_{ij} は実験初期に全て 0 で初期化する. エージェント i が協調期間中 ($L_i > 0$) にのみ隣接エージェント j に裏切られた場合このカウンターに 1 を加算し, $dc_{ij} \geq M$ のときエージェント i はエージェント j を裏切りエージェントと判定する. しかし, エージェント i は協調期間中かどうかによらず, エージェント j が協調行動を選択した場合に dc_{ij} を 0 に初期化する.

3·2 裏切りエージェントへの行動決定

拡張期待戦略エージェント i は裏切りエージェントに対して, 統合法による基本行動の決定に考慮せず, 協調期間中でも協調を行わないという 2 つの異なる行動を行う.

§1 統合法による基本行動の決定に考慮しない

統合法の行動決定のセクションの式 (2) に以下の式に再定義する. エージェント i が裏切りエージェントと判定していない隣接エージェントの集合を j_c とし, この j_c に対し優先度 $p^i(s)$ を以下とする. に対する行動 s^i_j から自らの行動 s^i を決定するための手法である.

$$p^i(s) = \sum_{j \in N_i} \delta(s, s^i_{j_c}(t-1)) \tag{6}$$

§2 協調期間中であっても強制的な協調を行わない

エージェント i が裏切りエージェントと判定した ($dc_{ij} > 0$) エージェント j に対してはインタラクションの際にエージェント i が協調期間中であるかによらず, エージェント i は統合法による基本行動でエージェント j とインタラクションを行う. そのため, このインタラクションによってエージェント i の協調施行期間 L_i は減少させない. 一方インタラクションによって協調が成立した場合は協調試行期間の再設定は行う.

4. 実　　験

4·1 実 験 内 容

協調期待戦略及び拡張協調期待戦略を取るエージェントのネットワークに一定比率で Q 学習エージェントを混在させることで, ネットワーク上での囚人のジレンマゲームを通して協調をどの程度拡散させ, 維持できるかを調査する. 各エージェントはあらかじめ協調期待戦略により行動するか, 合理的に行動するかを決定し, 戦略は変更しないものとする. 各エージェントはそれぞれのゲームごとにそれぞれのエージェントごとに学習した Q 値を元に, 統合法によって基本行動を決定する. ここで決定した基本行動を元にそれぞれのエージェントは隣接エージェントとランダムな順番で 2 人ゲームの囚人のジレンマゲーム（インタラクション）を 1 回ずつ行う. またその結果に応じた報酬を受け取り, Q 値を逐次更新する. 全てのエージェントが隣接エージェントと 1 回ずつインタラクションを行うまでを 1 ラウンドとし, 協調成立率の変化を調査する. すべての実験を通して完全グラフにおけるエージェント数は 300, 正規ネットワークにおけるエージェント数は 1000, 正規ネットワークにおける平均次数は 10 とし, Q 学習の学習率 $\alpha = 0.1$, 裏切りエージェントの判定閾値 $M = 2$, 学習行動戦略として ε-greedy ($\varepsilon = 0.05$) を採用した. また, [Yu 13b] にならい, 90%以上のエージェントが協調を維持した場合に協調の拡散に成功したと判断する. しかし, Q 学習エージェントは Q 学習の結果のみを元に行動するため, 実験中裏切り行動に収束する. そのため, 本論文では（拡張）協調期待戦略のエージェント間でのインタラクションにおいて 90%以上で協調が成立した場合に協調の拡散に成功したとする. 以下の実験結果は全て 30 回の試行に基づく平均である.

4·2 評 価 実 験

拡張協調期待戦略との比較として, 協調期待戦略のエージェントネットワークに Q 学習エージェントを混在させ, 協調期待戦略の頑健性を調査する. この結果, 完全グラフにおいては協調期間が 3 のとき, 10%以上の Q 学習エージェントが混在すると協調を促進できないことがわかった. ここでは分かりやすさのため, 完全グラフ上で Q 学習エージェントを 20%混在させた場合の協調期待戦略エージェント間の協調成立率を図 3 に示す. $L = 3$ の場合, 1300 ラウンドあたりから協調成立率が一定になっているが, より長く実験を続けると段階的に協調成立率は下がることを確認している.

次に正規ネットワークにおいて協調期待戦略エージェントのみで実験を行った結果を図 4 に示す. 図 4 からもわかるように, $L = 3$ でも協調成立率が徐々に下がり, さらに長いラウンド数実験を続けると協調成立率は 0 になる. このことから, 協調期待戦略では正規ネットワークにおいては Q 学習エージェントが混在しない状況であっても $L = 3$ では協調を促進できないことがわかる. これらの結果は [大塚 16] の結果と一致する.

4·3 完 全 グ ラ フ

完全グラフにおいて拡張協調期待戦略の Q 学習エージェントに対する限界を調査する. エージェント数を 300 とし, 協調期間定数 $L = 3$ において, Q 学習エージェントの割合を 50%から 90%に変化させたときの拡張協調期待戦略エージェント間の協調成立率を図 5 に, このときの全エージェント間での協調成立率を図 6 に示す. また表

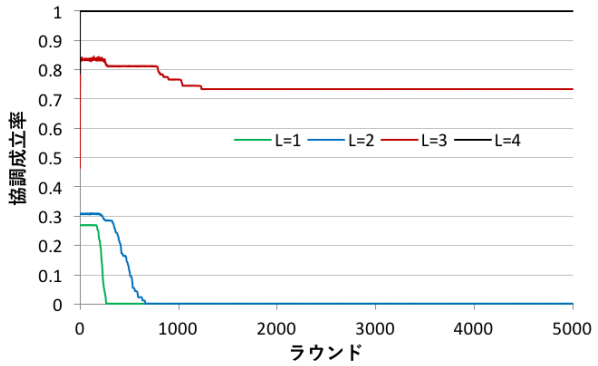

図 3 20%の Q 学習エージェントを含む完全グラフ上における協調期待戦略エージェント間での協調成立率

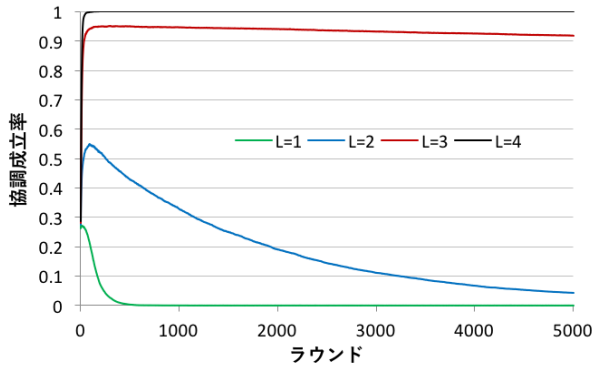

図 4 協調期待戦略エージェントのみの正規ネットワーク上における協調成立率

1 に協調期間定数 $L=3$ のとき, 5000 ラウンド時点で, Q 学習エージェントを裏切りエージェントと判定した割合及び拡張協調期待戦略エージェントを裏切りエージェントと誤判定していた割合を示す.

協調期間定数 $L=3$ の拡張協調期待戦略では Q 学習エージェントの割合が 10%のときはもとより, 60%存在しても協調成立率は 90%であり, 拡張協調期待戦略エージェント間で協調が維持されることがわかった. Q 学習エージェントの割合が 70%のときであっても 80%程度の協調成立率を維持しているが, ラウンド数を伸ばすと, 緩やかに協調成立率は落ちる. また, Q 学習エージェントが 90%混在するときでも初期 100 ラウンド程度では拡張協調期待戦略エージェント間では協調成立率がほぼ 100%に達し, その後 Q 学習エージェントの割合によって協調成立率が変動している. これは実験の初期の状態ではエージェントの Q 値が初期値 (0) で, 協調か裏切りがランダムの確率で選択されることに起因している. しかし, 図 6 からもわかるように Q 学習エージェントの大半が裏切り行動に収束しており, さらに 150 ラウンド程度でほぼ全ての Q 学習エージェントが裏切り行動に収束する. その結果, 拡張協調期待戦略エージェントの協調期間が途切れ, 協調の Q 値が低いと裏切りに収束したと考えられる.

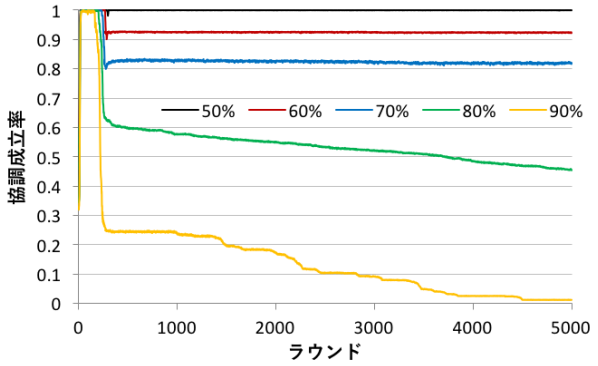

図 5 $L=3$ における完全グラフ上における拡張協調期待戦略エージェント間での協調成立率

一方, 図 6 を見ると拡張協調期待戦略エージェントが高い協調成立率を維持している場合であっても Q 学習エージェントが裏切りに収束しているため, ネットワーク全体では協調が促進されていないことがわかる. 例えば Q 学習エージェントの割合が 50%であっても, Q 学習エージェントが関わるインタラクションは 75%程度ある.

表 1 から協調の維持に成功した Q 学習エージェントが 60%以下では Q 学習エージェントの判定率も 90%以上の高い判定率となった. Q 学習エージェントは 5%の確率で ε-greedy によるランダムな行動選択をしており, 裏切りエージェントの判定閾値が 2 のため, 95%の判定率は全ての Q 学習エージェントを同定している.

次に協調期間定数 $L=2$ において, Q 学習エージェントの割合を 10%から 40%に変化させた場合の拡張協調期待戦略エージェント間の協調成立率を図 7 に示す. この結果 $L=2$ では Q 学習エージェントが 10%でも協調を促進できないことがわかる. 図 7 を見ると $L=2$ では図 5 の $L=3$ の初期に見られた協調成立の急激な上昇がないため, 裏切りエージェントの割合よりも協調期間定数 L 及び初期の Q 学習の初期値の設定に依存すると考えられる. 協調期間定数の値によらず, 拡張協調期待戦略エージェント間での協調成立率は 30%程度で安定しているが, 本稿の実験からはこの理由を推測できる結果は得られなかった.

4·4 正規ネットワーク

エージェント数 1000 の正規ネットワークにおける拡張協調期待戦略の Q 学習エージェントに対する限界を調査する. 正規ネットワークでは円周上に並べたエージェントそれぞれで m 個隣のエージェントまでリンクを貼ることでネットワークを作るが, ここでは $m=5$ とし, それぞれの隣接するエージェント数, つまり平均次数は $2m=10$ である. まず, 図 8 に拡張協調期待戦略エージェントのみで, 協調期間定数 L を 2 と 3 に設定した場合の協調成立率を示す. この結果, 拡張協調期待戦略によって協調期待戦略で協調の収束に失敗した $L=3$ さらに $L=2$ におい

表 1 完全グラフにおける裏切りエージェントの判定率 ($L = 3$)

ネットワーク中の Q 学習エージェントの割合	10%	20%	30%	40%	50%	60%	70%	80%	90%
Q 学習エージェントの判定率	95.0	95.1	95.1	95.1	95.0	91.0	88.2	65.6	2.38
拡張協調期待戦略エージェント間の誤判定率	0	0	0	0	0	3.43	4.64	16.5	1.29

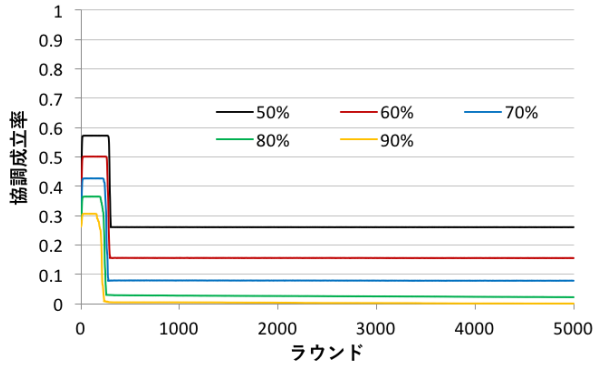

図 6 $L = 3$ における完全グラフ上における全エージェント間での協調成立率

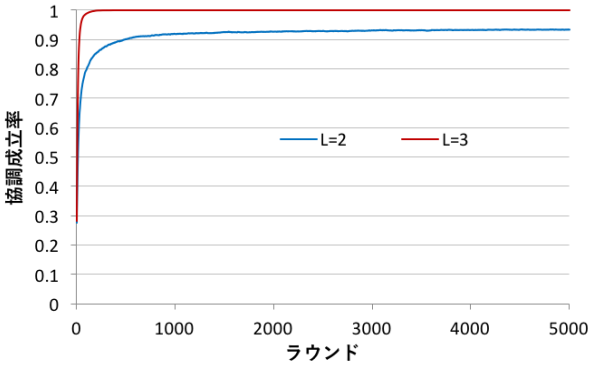

図 8 拡張協調期待戦略エージェントのみの正規ネットワーク上における協調成立率

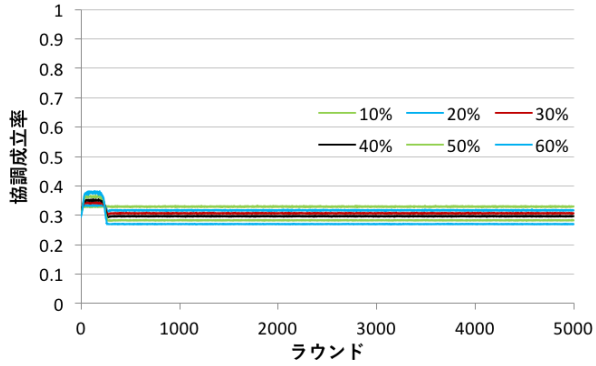

図 7 $L = 2$ における完全グラフ上における全エージェント間での協調成立率

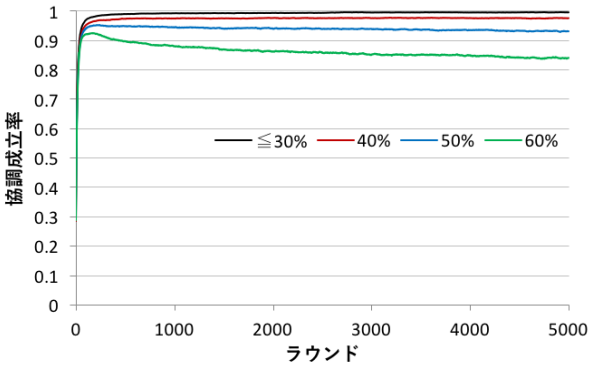

図 9 正規ネットワーク上における拡張協調期待戦略エージェント間の協調成立率

ても協調の収束に成功した. これは, これまでの協調期待戦略と比較して, 大きく異る結果となった. この要因として, Q 学習エージェントが混在しない状況であっても裏切りエージェントの判定を行うことで, 初期にランダムで裏切り行動をとる半数程度のエージェントを裏切りエージェントとすることで, 協調期待戦略による協調が広がるまでの間, 一時的に裏切り行動を学習しているエージェントに協調期間を消費しなくなったためと考えられる.

次に, 協調期間定数 $L = 3$ において, Q 学習エージェントの割合を 30%から 60%に変化させた場合の拡張協調期待戦略エージェント間の協調成立率を図 9 に, 表 2 に協調期間定数 $L = 3$ のとき, 5000 ラウンド時点で, Q 学習エージェントを裏切りエージェントと判定した割合及び拡張協調期待戦略エージェントを裏切りエージェントと誤判定していた割合を示す. 協調期間定数 $L = 3$ では Q 学習エージェントが 40%程度混在しても協調が維持された. 50%でも 5000 ラウンド目では協調成立率は 90%以上となったが, ラウンド数を増やして行くと徐々に協調成立率が下がり収束に失敗した. 表 2 から, 協調の収束に成功した Q 学習エージェントの割合が 40%以下では, Q 学習エージェントの判定率が 90%以上となった. また, Q 学習エージェントの割合が 10%程度であっても拡張協調期待戦略エージェント間の誤判定が発生した. これは平均次数が 10 である正規ネットワーク上において局所的な裏切りの輪が発生し, その輪に含まれる拡張協調期待戦略エージェントが裏切りに収束したためである.

次に, 協調期間定数 $L = 2$ において, Q 学習エージェントの割合を 10%から 50%に変化させた場合の拡張協調期待戦略エージェント間の協調成立率を図 10 に示す. 協調期間定数 $L = 2$ では Q 学習エージェントが 10%以下では協調の収束に成功し, 20%程度においても協調成立率が 90%には及ばないが僅かな増加傾向にある状態となった.

表 2　正規ネットワークにおける裏切りエージェントの判定率 ($L = 3$)

ネットワーク中の Q 学習エージェントの割合	10%	20%	30%	40%	50%	60%	70%	80%	90%
Q 学習エージェントの判定率	95.3	95.1	94.2	91.5	84.4	71.8	54.8	34.4	14.4
拡張協調期待戦略エージェント間の誤判定率	0.003	0.02	0.16	0.50	1.66	3.25	4.79	5.47	3.34

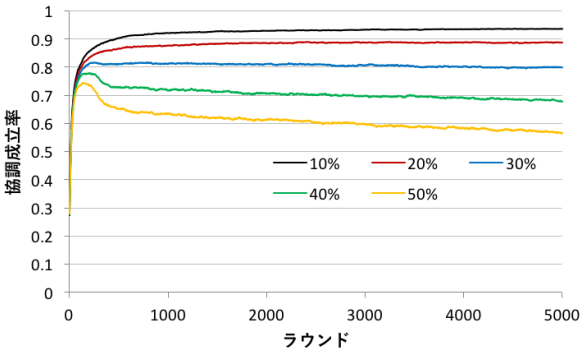

図 10　$L = 2$ における正規ネットワーク上における拡張協調期待戦略エージェント間での協調成立率

5. 考　　察

以上の実験結果から裏切りが適切と学習する Q 学習エージェントを導入しても, 拡張協調期待戦略はそれを同定し, 拡張協調期待戦略をもつエージェント同士での協調が広がることがわかった. つぎに実験の結果から, 完全グラフと正規ネットワークの違い及び今日長期間の違いによる拡張協調期待戦略の特徴および裏切りエージェントの判定閾値について考察する.

図 5, 図 9 から, 協調期間定数 $L = 3$ のとき, 完全グラフではの Q 学習エージェントの割合が 60%でも協調を維持できたが, 正規ネットワークでは 40%となった. 一方図 7, 図 10 から, $L = 2$ のときでは, 完全グラフでは Q 学習エージェントの割合によらず協調成立率を維持できなかったが, 正規ネットワークでは Q 学習エージェントが 10%程度混在しても協調が維持された. この違いは完全グラフと正規ネットの局所性の差, とインタラクションの順序をランダムに決定していることに起因すると考えられる. 本実験では実験の最初では, ランダムに行動を決定している. そのため, 任意の 2 エージェントで協調が成立する確率は 4 分の 1 であり協調期待戦略では, 協調の広がりが遅くなるが, 拡張協調期待戦略では裏切りエージェントの判定によりそれらを区別でき, 図 5 や図 6 で顕著なように初期段階での拡張協調期待戦略間での高い協調成立が実現されている. さらに, 裏切りエージェントと一時的に判定されても戦略の寛容性から相互協調に入ることが出来る. しかし, Q 学習エージェントが裏切りに収束するため, これにより拡張協調期待戦略間での協調期間中の協調行動が途切れるかどうかが, 協調を維持できるかの境界となる.

完全グラフでは, 全てのエージェントがランダムに相手を決めるため, 一定の間隔で協調が成立する. 一方で, 正規ネットではインタラクションの順序はランダムだが, ネットワーク上に Q 学習エージェントもしくは拡張協調期待戦略エージェントが偏って存在することで, 協調, 裏切りの広がり方が局所的に緩やかになる. その為, 正規ネットワークではランダムな順序でインタラクションが行われるときに $L = 2$ でも拡張協調期待戦略による協調がある程度維持でき, Q 学習エージェントが 10%程度混在してもそれが保てる. これは局所的な協調への収束があれば, そこに徐々に協調が広がり, 一時的に裏切りを選択していた拡張協調期待戦略エージェントも協調の輪に含まれるようになる. これは社会的ジレンマ構造を背後に持つマルチエージェントシステムを構築するときに重要な要素である.

また, 表 1, 表 2 から拡張協調期待戦略中での裏切りに収束するエージェントが増えるに従って誤判定も増加する. 裏切りエージェントを 90%以上判定できているときに収束に成功しているようにも読み取れるが, これは裏切りエージェントの判定方法が, 自身が協調期間中に裏切り行動を 2 回連続して取られたときとしているため, 周囲に Q 学習エージェントが多数いると拡張協調期待戦略エージェントも協調期間から出て裏切り行動を取ることがあり, これが誤判定に繋がったと考えられる.

本実験では裏切り判定閾値を $M = 2$ として実験をした. 本論文では示さなかったが, 裏切り判定閾値を変動させる実験も行い, $M = 1$ で一部の環境では $M = 2$ よりも協調性効率を高めることが出来たが, 同時に拡張協調期待戦略エージェント間での裏切り判定率を高めてしまい, $M \geq 3$ では殆どの場合で $M = 2$ より協調成立率が下がる結果であった.

6. 結論と今後の課題

本論文では, 協調期待戦略に裏切りに収束したエージェントを同定し, 協調期間による協調を行わない拡張協調期待戦略を提案し裏切り行動に収束する Q 学習エージェントが混在する環境においても拡張協調期待戦略に比べ高い確率で協調を拡散できることを示した. 一方で, 今回の実験では完全グラフ, 正規ネットしか考慮しておらず, より現実に近い, Watts-Strogatz(WS) モデルや Barabasi-Arbert(BA) モデルのネットワーク上での調査や拡張協調期待戦略によるネットワーク全体での利得への影響の調査が考えられる.

◇ 参 考 文 献 ◇

[Jianye 15] Jianye, H., Sun, J., Huang, D., Cai, Y., and Yu, C.: Heuristic Collective Learning for Efficient and Robust Emergence of Social Norms, in *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*, AAMAS '15, pp. 1647–1648, Richland, SC (2015), International Foundation for Autonomous Agents and Multiagent Systems

[Sen 10] Sen, O. and Sen, S.: Effects of Social Network Topology and Options on Norm Emergence, Vol. 67, No. 5, p. 056194 (2010)

[Shibusawa 14] Shibusawa, R. and Sugawara, T.: Norm Emergence via Influential Weight Propagation in Complex Networks, in *2014 European Network Intelligence Conference, ENIC 2014*, pp. 30–37 (IEEE Xplane, 2014)

[Walker 95] Walker, A. and Wooldridge, M.: Understanding the Emergence of Conventions in Multi-Agent Systems, in *Proceedings of the First International Conference on Multiagent Systems*, pp. 384–389, The MIT Press (1995)

[Yu 13a] Yu, C., Zhang, M., Ren, F., and Luo, X.: Emergence of Social Norms Through Collective Learning in Networked Agent Societies, AAMAS '13, pp. 475–482 (2013)

[Yu 13b] Yu, C., Zhang, M., Ren, F., and Luo, X.: Emergence of Social Norms Through Collective Learning in Networked Agent Societies, in *Proceedings of the 2013 International Conference on Autonomous Agents and Multi-agent Systems*, AAMAS '13, pp. 475–482, Richland, SC (2013), International Foundation for Autonomous Agents and Multiagent Systems

[Yu 15] Yu, C., Lv, H., Bao, H., and Li, Y.: Emergence of Social Norms through Collective Learning and Information Diffusion in Complex Relationship Networks, *International Joint Agents Workshop and Symposium(IJAWS2015)* (2015)

[渋澤 15] 渋澤 亮介, 菅原 俊治：複雑ネットワーク上での囚人のジレンマゲームにおける協調の促進について, Joint Agent Workshops & Symposium (2015)

[大塚 16] 大塚 知亮, 菅原 俊治：協調期待戦略による協調促進の頑健性について, Joint Agent Workshops & Symposium (2016)