別紙 1

論文の内容の要旨

論文題目	静的・動的環境における通信なしマルチエージェント強化学習の協調行動を導く合理的目的設定
学位申請者	上野　史

本論文は，現実問題に起こる通信遅延や情報の不確かさに対処するために，複数のエージェント間の協調行動を通信なしに導く強化学習手法を提案するとともに，変化のない静的環境に加えて不測の事態などで変化する動的環境にも対応できるように拡張し，それらの有効性を検証することを目的とする．その実現に向け，各エージェントの目的を合理的に設定し，逐次的にエージェントの行動を決定させることで協調行動を導くアプローチをとる．また，提案手法を迷路問題に適用し，静的環境のみならず動的環境においても協調行動を早く獲得可能であることを明らかにする．本論文は全８章で構成され，その内容の要旨は以下の通りである．

第１章では，マルチエージェント強化学習の説明を通して，情報通信がなくてもエージェント間で協調行動を実現することの重要性を示す．また，従来研究の問題点を説明し，本研究の位置づけと研究目的を述べる．

第２章では，マルチエージェント強化学習における設計に着目する．具体的には，学習エージェントの構成を説明した後，エージェント間の協調について焦点をあてる．特に，エージェント間の通信頻度と環境の変動度合により達成可能な協調の違いを説明し，エージェントの設計の難しさを述べる．続いて，従来研究をまとめるとともに，現在のマルチエージェント強化学習はエージェント間の情報共有に基づいているため，情報共有ができない場合は，学習環境の不安定性から合理性を示すことができないことを論じる．

第３章では，取り上げる迷路問題とその特徴について説明する．また，動的変化は問題の性質の違いから，環境の大きさと形状，エージェントのスタート・ゴール位置が変化する「空間的環境変化」と，エージェント・ゴール数が変化する「エージェント・ゴール数変化」に分類する．

第４章では，静的環境において，全てのエージェントが他のエージェントに対して譲歩行動を導く目的を学習することで，情報通信なしに協調行動を獲得可能な手法(Profit Minimizing Reinforcement Learning: PMRL)を提案する．また，譲歩行動を達成するための報酬値(内部報酬値)の設定による学習結果の収束性を示し，その条件を明らかにする．シミュレーション実験では，従来手法のQ学習とProfit Sharing(PS)，提案手法のPMRLを比較し，Q学習エージェントやPSエージェントは協調行動を獲得できないのに対し，PMRLエージェントは適切な目的を決定することで競合を解消でき，ほぼ最短ステップでゴールに到達することを示す．

第５章では，「空間的環境変化」が起こる動的環境において，適切な目的に切り替える手法(Profit Minimizing Reinforcement Learning with Oblivion of Memory: PMRL-OM)を提案する．具体的には，環境変化に伴って達成すべき目的が変化する可能性があるため，達成した目的と達成までにかかった行動数をセットで記憶し，一定数以上記憶したときに最初に記憶したセットから削除して，常に最新の情報を維持することで適切な目的に変更する．シミュレーション実験では，PMRL，PS，PMRL-OMを比較し，PMRLエージェントやPSエージェントは環境変化後の環境に追従できないのに対して，PMRL-OMエージェントは環境変化後の最新の情報(ゴールまでのステップ数)を活用することで目的を変化し，ほぼ最短ステップでゴールに到達することを示す．

第６章では，「エージェント・ゴール数変化」が起こる環境において，環境変化に応じて決めた範囲内のエージェントのみが目的を変更する手法(Profit Minimizing Reinforcement Learning for Dynamic Rewards and Environmental States: PMRL-DRES)を提案する．具体的には，全てのエージェントは達成すべき目的を変更する必要はなく，必要なエージェントのみ変更すればよいことから，その対象エージェントを決める．シミュレーション実験では，PMRL-OM，PS，PMRL-DRESを比較し，PMRL-OMエージェントやPSエージェントは環境変化への適応に限界があるのに対し，PMRL-DRESエージェントは環境変化後に学習領域を分割することで，適切な目的を選択でき，環境変化の種類やタイミングに関わらず，ほぼ最短ステップでゴールに到達することを示す．

第７章では，現実問題へ適用する上での提案手法の設計論について述べる．具体的には，提案手法の迷路問題に対する適用限界について説明する．また，現実問題に展開する上で必要な情報に加え，エージェントの状態，行動，報酬の設定方法を明らかにし，提案手法の適用範囲について論じる．

最後に，第８章では上記の各章にて得られた知見をまとめ，本研究の成果について述べるとともに，今後想定外の問題へ適用する上での課題を述べる．

別紙２

論文審査の結果の要旨

学位申請者氏名　　上野　史

審査委員主査　　髙玉　圭樹

委員　　西野　哲朗

委員　　大須賀　昭彦

委員　　庄野　逸

委員　　佐藤　寛之

本論文は，複数のエージェント間の協調行動を通信なしに導く強化学習手法を探究し，変化のない静的環境に加えて不測の事態などで変化する動的環境にも対応可能な手法を提案するとともに，それらの有効性を迷路問題への適用を通して検証するものである．

第１章では，マルチエージェント強化学習の背景を述べるとともに，従来研究が着目してこなかったエージェントと環境のモデル化の重要性を説明している．さらに，静的環境のみのらず動的環境で協調行動を学習することの価値を指摘し，研究目的を述べている．

第２章では，本論文で扱うマルチエージェント強化学習について概観している．具体的には，エージェントの定義と強化学習のメカニズムを紹介した後，従来手法の説明を通してマルチエージェントシステムにおける協調行動についてまとめている．その中で，通信せずに協調行動を学習する方法は実現困難であり，複雑な協調行動の学習のためにより多くの情報を利用する研究が主流となっていることから，実応用性への問題点を論じている．

第３章では，提案手法を評価するためのテストベッドとして迷路問題を選定し，迷路問題に動的変化の要素を組み込むための設計方法を述べている．特に，動的変化は二種類に分けることができ，環境の大きさと形状，エージェントのスタート位置，ゴール位置が変化する「空間的環境変化」と，エージェント数とゴール数が変化する「エージェント・ゴール数変化」に分類している．

第４章では，「静的環境」において通信なしに協調行動を獲得するために，譲歩行動のための目的を学習する手法(Profit Minimizing Reinforcement Learning: PMRL)を提案し，その合理性を明らかにしている．具体的には，協調行動のために

必要な目的を達成するための報酬値(内部報酬値)を設定している．静的環境のシミュレーション実験では，PMRLエージェントは適切な目的を決定することで他のエージェントとの競合を解消でき，ほぼ最短ステップでゴールに到達することに成功している．

第５章では，「空間的環境変化」が起こる動的環境において通信なしに協調行動を獲得するために，環境変化に伴って常に最新の情報を維持しながら，適切な目的を学習する手法(Profit Minimizing Reinforcement Learning with Oblivion of Memory: PMRL-OM)を提案している．また，PMRL-OMはPMRLが持つ合理性を維持できることを明らかにしている．動的環境のシミュレーション実験では，PMRLエージェントは環境変化前の情報に引きずられるのに対し，PMRL-OMエージェントは環境変化後の最新の情報(ゴールまでのステップ数)を活用し，ほぼ最短ステップでゴールに到達することに成功している．

第６章では，「エージェント・ゴール数変化」が起こる動的環境において通信なしに協調行動を獲得するために，環境変化に応じて学習する範囲を適切に制限し，その範囲内で目的を変更する手法(Profit Minimizing Reinforcement Learning for Dynamic Rewards and Environmental States: PMRL-DRES)を提案している．動的環境のシミュレーション実験では，PMRL-OMエージェントは環境変化への適応に限界があるのに対し，PMRL-DRESエージェントは設計した全環境変化に対して適切な目的を選択でき，ほぼ最短ステップでゴールに到達することに成功している．特に，PMRL-DRESエージェントは環境変化前の学習結果を利用することで，環境変化後に学習をやり直すよりも，環境変化の影響を受けにくいことも明らかにしている．

第７章では，現実問題へ適用する上での提案手法の設計論について論じている．具体的には，第４章から第６章の実験で得られた知見をまとめ，その限界について述べている．特に，エージェント数とゴール数が異なる場合の協調行動獲得の難しさを指摘するとともに，内部報酬値の設計法について論じ，実応用として複数ロボットによる物流システムへの展開について述べている．

最後に，第８章では上記の各章で得られた知見をまとめ，本研究の成果について述べるとともに，今後様々な問題へ適用する上での課題を述べている．

以上のように，本論文は，エージェント間の通信なしに協調行動の学習を可能し，マルチエージェント強化学習における適用範囲を大きく広げるものである．したがって，本論文は博士(工学)の学位論文として十分な価値を有するものであると認める．