

## 科学研究費助成事業（科学研究費補助金）研究成果報告書

平成24年 6月 8日現在

機関番号：82626

研究種目：基盤研究（C）

研究期間：2009～2011

課題番号：21500153

研究課題名（和文）マルチエージェントの学習過程に注目した系安定化・全体最適化に関する研究

研究課題名（英文）Researches on System Stabilization based on Learning Process of Multiple Agents.

研究代表者

野田 五十樹（NODA ITSUKI）

独立行政法人産業技術総合研究所・サービス工学研究センター・研究チーム長

研究者番号：40357744

研究成果の概要（和文）：

複数のエージェントが同時学習する際に生じるジレンマに起因する系の不安定を解決するため、エージェント学習の環境変化に対する不感度とエージェントの受ける情報配信のバランスを制御する方法を開発した。その結果、強化学習のステップサイズパラメータを自動的に制御するRASPや学習におけるExploration/Exploitationを調整する方法を始め、エージェント学習の制御パラメータを制御して系の安定さを向上させる手法を導出・提案した。

研究成果の概要（英文）：

In order to overcome difficulties of unstable situation caused by dilemma of simultaneous learning of multiple agents, we proposed several methods to control insensitiveness to environmental changes and information flow for the multiple agents. RASP is a method to control stepsize parameter of reinforcement learning automatically according to the degree of changes of environment. We also investigate a method to control Exploration/Exploitation ratio to stabilize learning processes.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2009年度	1,100,000	330,000	1,430,000
2010年度	1,300,000	390,000	1,690,000
2011年度	1,000,000	300,000	1,300,000
総計	3,400,000	1,020,000	4,420,000

研究分野：総合領域

科研費の分科・細目：情報学 知能情報学

キーワード：知的エージェント、マルチエージェント社会シミュレーション、強化学習、学習パラメータ

## 1. 研究開始当初の背景

複数の意思決定主体（エージェント）が独立して判断・行動を行うマルチエージェントシステムでは、各エージェントがより良い行動をしようとする学習・適応に関して、ジレンマを抱えたシステムになっている。すなわち、各エージェントが自らが実際に行った行動の経験から学習する場合、そのエージェン

トが置かれた環境は各行動に対して安定した反応を返す必要がある。一方、あるエージェントが学習によって振舞いを変化させた場合、その変化が他のエージェントにとっての環境の攪乱要因になってしまう。この攪乱が新たなエージェントの振舞いの変化につながり、その連鎖反応により系全体が不安定になる場合がある。例えば、じゃんけんの戦

略を一般的な強化学習手法である GIGA で複数エージェントが同時に学習した場合、Nash 均衡である最適解(すべての手を 1/3 の確率で出す)に収束せず、学習過程が発振してしまうことが知られている。同様の現象は、交通渋滞やネットワーク負荷分散など一般的に見ることができる。

エージェントの学習を加速する手段として、他のエージェントの経験を共有して利用する方法が良く用いられるが、この学習の高速化が上記のジレンマを増長させてしまうという問題も存在する。すなわち、経験の共有によりエージェントの学習が均質化されて全エージェントが同じ選択をするようになるため、特定の資源への集中が起き、全体として発振現象が生じて系が不安定になる。例えば、社会シミュレーションにおいても、VICS (Vehicle Information and Communication System) の普及がかえって交通渋滞を引き起こす例や、テーマパークなどで各アトラクションの待ち時間情報を広く公開するほど、逆に混雑を悪化させてしまう例が報告されている。

## 2. 研究の目的

本研究では、マルチエージェントの学習過程と環境の変化におけるジレンマに起因する系の不安定さを解決するため、エージェントの学習の不感度と不均質性に着目し、系の状態に応じてこれらの尺度が適切になるようにエージェントの学習パラメータや行動選択機構、エージェント間での環境情報の共有方法を調整し、系を安定化・全体最適化する手法を構築することを目指した。

マルチエージェント学習のジレンマに対し、系全体の状態と学習の感度と均質さに注目し、系全体の状態によって学習の不感化・不均質化を制御する方法の研究を進めた。具体的には、感度や均質さに関係する学習パラメータやエージェント間情報共有機構を取り上げ、系の状態によって学習の不感化・不均質化を行う方法を開発した。ここで重要なのは、系の状態が安定化し理想状態に近づいたとき、この不感化・不均質化を自動的に低減できるようにする手法である。上でも述べているように、単純な不感化・不均質化では、各エージェントおよび系全体にとっての理想からずれた状態に学習が進んでしまう。系が安定化した段階で不感化・不均質化を低減出来れば、最終的な収束を理想状態に持っていくことが可能となり、たとえオープンシステムで利己のエージェントが参加したとしても、各エージェントが不利益を被る事なく、適応することが可能となる。

## 3. 研究の方法

本提案ではマルチエージェントの系全体

の状態に応じて、各エージェントの学習の不感化・不均質化を調整する方法を探る。具体的には、以下の2つの側面から研究を進める。

- エージェントの各学習・行動選択パラメータと学習の感度および系全体への影響を調べ、系の状態をモニターしながらパラメータを調整する方法を開発する。例えば、強化学習におけるステップサイズパラメータや行動選択における温度パラメータとエージェントの環境としての系全体との関係を導き出す。その結果に基づき、これらのパラメータに關係する系の状態をモニター・推定する方法を構築し、状態に応じてパラメータを調整して系が安定するように学習を不感化する手法を確立する。
- エージェントが学習の際に利用する情報のうち、他のエージェントの行動選択に関する情報の収集・配信方法について、配信される情報と系の状態の関係を調べ、エージェントの学習や行動選択が適切に不均質化するような情報配信の仕組みを構築する。

## 4. 研究成果

2009年度は、これまで開発を進めてきた手法の一般化を行い、理論的な裏づけのある枠組みの構築を行った。

まず学習・行動選択パラメータについては、強化学習のステップサイズパラメータを周りの環境に合わせて調整する方法を構築し、Recursive Adaptation of Step Size Parameter (RASP) と名づけて具体的な学習アルゴリズムを定式化した。この方法は強化学習で用いる指数移動平均(EMA)を再帰的に求めることで、ステップサイズパラメータによる学習対象値の高次導関数を効率よく求めるというものであり、数学的に裏づけされた汎用性の高い手法となっている。このため、さまざまな学習課題に適用でき、実際に金融データや気象データなどを用いた適応実験を進め、これらの系の分析に用いることができることがわかってきている。また、学習性能についても、いくつかの数値実験を行い、Optimal Step Size Algorithm などの既存手法より適切にパラメータ学習ができることが示された。

また行動選択については、デマンドバスなどのシミュレーションを行う環境を整備し、具体的事例からエージェントの行動選択やそれによる系の最適化・サービス安定化手法などを探る実験を進めた。これについては今後実験を重ね、公共交通など規模の大きい社会システムでの系の安定分析の例として用いることができるよう、整備を進める予定である。

2010年度は、前年度開発した Recursive

Adaptation of Step Size Parameter (RASP) の手法を推し進め、高次の導関数を用いて最適ステップサイズをニュートン法により求める方法を構築し、それによりマルチエージェント系においてどのように系を安定化・円滑化するかを確認した。

RASP を用いると高次の導関数が自動的に計算できるため、これを利用してニュートン法を導出することができる。これにより、これまで山登り法で徐々に変化させてきたステップサイズパラメータを一気に変更することが可能となる。この方法を適用した Rapid RASP by Newton 法 (RRASP-N) は、環境の変化に迅速に追従することができ、同時にノイズに対する頑健性も担保できる。この性質は、多様な行動原理を持つエージェントが多数共存する環境において、それらのエージェントの挙動をうまく吸収する働きを發揮でき、系全体の性質を安定化させることができる。これを確認するために、共通資源を奪い合うマルチエージェント系を構成し、学習を行うエージェントとそれ以外にランダムに行動するエージェントを混在させる実験を行った。その結果、学習エージェントの存在により適切なステップサイズパラメータが動的に選ばれ、系全体の安定性向上に貢献していることが確認できた。これらの現象は理論的にも確認でき、多様な応用が期待できる。さらに、ラプラス変換・フーリエ変換との類似性に基づき、周期的に変化する環境について安定的な動作を学習する方法についても開発に着手した。

また、情報配信の計に与える影響に関しては、避難における人流シミュレーションにおける実際の人間の行動のモデル化を試み、学習モデルとの関連について検討を行った。

2011 年度は、これまで得た結果を元に、動的な環境におけるマルチエージェントの群としての挙動の分析を進め、系の制御を可能とする情報の収集・配信方法について検討を行なった。

まず、これまで進めてきた再帰的ステップサイズパラメータ調整法 (RASP) を株取引市場に適用し、系全体の特徴をパラメータ化してエージェントの行動の制御に用いることを試みた。ここでは株価の変化を環境からの報酬と扱い、RASP を用いて系の特徴を抽出し、株の売り買いにどの程度適用可能かを検証した。

また、平行して、エージェント行動を制御する温度パラメータについて、エージェントが学習を進める系全体の変化の大きさとの関係を数値解析的に分析し、その結果に従ってパラメータを制御し、系全体の状態を表す情報を獲得する方法を構築した。動的環境下におけるエージェントの行動学習では、行動選択の方針、特に Exploration と Exploitation

の割合 (以下 E-E ratio と呼ぶ) を制御する温度パラメータをどう決めるのが重要な問題となる。特にマルチエージェント環境下では各エージェントの行動選択が他エージェントの学習環境に直接影響するため、E-E ratio は十分に注意して選ぶ必要がある。それに対し、二人のエージェントが強化学習を行う場合の Exploration と Exploitation の割合 (E-E ratio) と行動選択が正解となる確率の関係を調べ、正解確率を最大化する E-E ratio を求める方法を提案した。この方法では、変化する利得行列に追従するための Exploration の確保と、相互の学習に与える影響の低減のための Exploitation のトレードオフを、行動選択の正解確率最大化という確率で解く方法となっている。この最大化を求めるにあたっては、エージェント相互の行動選択の対称性を仮定しており、エージェントの社会性につながる前提を用いている。これにより、突発的に変化する環境においても、温度パラメータを適切に制御する方法を確立することができ、学習および系全体を安定させる枠組みを構築することができた。

さらには、震災における災害救助や避難行動および情報配信の調査を行ない、情報がエージェント群に与える影響についてのデータの蓄積を行なった。

## 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 4 件)

- ① Itsuki Noda, "Adaption of Step Size Parameter Using Newton's Method", AGENTS IN PRINCIPLE, AGENTS IN PRACTICE、査読有、Vol. 7047/2011、2011、pp. 349-360、DOI:10.1007/978-3-642-25044-6\_28
- ② Itsuki Noda, "Adaptation of Step Size Parameter to Minimize Exponential Moving Average of Square Error by Newton's Method", Proc. of Adaptive and Learning Agents、査読有、-、2010、pp. M-2-1、DOI:なし
- ③ 野田五十樹、松原仁、"ロボカップ 12 年"、人工知能学会誌、査読無、Vol. 25、2010、183-188、DOI:なし
- ④ Itsuki Noda, "Recursive Adaptation of Step Size Parameter for Non-stationary Environments"、Principles of Practice in Multi-Agent Systems、査読有、Vol. 1、2009、pp. 525-533、DOI:10.1007/978-3-642-11161-7\_38

〔学会発表〕(計17件)

- ① 野田五十樹、"災害救助マルチエージェントシミュレーションの可能性"、FUN-AI 2012、2012.03.03、大沼
- ② 野田五十樹、"動的環境における Exploitation / Exploration 比率の制御"、JAWS2011、2011.10.27、熱海市
- ③ 松井宏樹、林慶樹、野田五十樹、"再帰的ステップサイズパラメータ調整法による株取引におけるボリュームカーブの推定"、人工知能学会全国大会、2011.06.02、盛岡市
- ④ 野田五十樹、Kim Hyun-Tae、"マルチエージェント学習下における温度パラメータの調節手法"、人工知能学会全国大会、2011.06.01、盛岡市
- ⑤ 野田五十樹、"周期的環境に対するフーリエ混合強化学習法"、情報処理学会全国大会、2011.03.04、東京
- ⑥ 野田五十樹、"複素数指数移動平均を用いた強化学習による周期的環境への適応"、JAWS 2010、2010.10.28、富良野
- ⑦ Itsuki Noda、"Rapid Recursive Adaptation of Step Size Parameter by Newton's Method"、PRICAI 2010、2010.08.30、Daegu, Korea
- ⑧ 松井宏樹、林慶樹、野田五十樹、尹熙元、"再帰的ステップサイズパラメータ調整法を用いた機械学習による金融データの分析"、人工知能学会全国大会、2010.6.11、長崎市
- ⑨ 野田五十樹、"マルチエージェント環境下における強化学習のステップサイズパラメータの適応"、人工知能学会全国大会、2010.6.9、長崎市
- ⑩ 野田五十樹、"ニュートン法による強化学習ステップサイズパラメータの調整法"、情報処理学会全国大会、2010.03.09、東京
- ⑪ 野田五十樹、"強化学習と社会の安定化"、FUN-AI 2010、2010.02.22、函館
- ⑫ 野田五十樹、"指数的移動平均 2 乗誤差の最小化によるステップサイズパラメータの調整法"、JAWS-2009、2009.10.28、蔵王
- ⑬ Itsuki Noda、"Adaptation of Step Size Parameter for Non-Stationary Environments by Recursive Exponential Moving Average"、ECML 2009 LNIID Workshop、2009.09.07、Bled, Slovenia
- ⑭ 松井宏樹、野田五十樹、尹熙元、"機械学習における再帰的ステップサイズパラメータ調整法を用いた価格変動の分析"、人工知能学会全国大会、2009.06.19、高松市
- ⑮ 舟山和男、吉村忍、野田五十樹、藤井秀樹、狩野宏和、"デマンド型交通導入

に関する仮想社会実験"、人工知能学会全国大会、2009.06.19、高松市

- ⑯ 野田五十樹、"強化学習エージェントと報酬頻度に関する考察"、人工知能学会全国大会、2009.06.17、高松市
- ⑰ Itsuki Noda、"Recursive Adaptation of Step Size Parameter for Unstable Environments"、ALA-2009、2009.05.12、Budapest, Hungary

〔図書〕(計1件)

- ① Itsuki Noda、Springer、"Recursive Adaptation of Step Size Parameter for Non-Stationary Environments" in Adaptive Learning Agents、2010、74-90

## 6. 研究組織

### (1) 研究代表者

野田 五十樹 (NODA ITSUKI)

独立行政法人産業技術総合研究所 サービス工学研究センター サービス設計支援技術研究チーム 研究チーム長

研究者番号：40357744

### (2) 研究分担者

山下 倫央 (YAMASHITA TOMOHISA)

独立行政法人産業技術総合研究所 サービス工学研究センター サービス設計支援技術研究チーム 研究員

研究者番号：50415759