

科学研究費助成事業 研究成果報告書

平成 29 年 6 月 15 日現在

機関番号：82646

研究種目：基盤研究(C) (一般)

研究期間：2014～2016

課題番号：26330267

研究課題名(和文) マルチエージェント系における経験強化型学習XoLの理論と応用

研究課題名(英文) Theory and Applications of Exploitation-oriented Learning XoL in Multi-agent Systems

研究代表者

宮崎 和光 (Miyazaki, Kazuteru)

独立行政法人大学改革支援・学位授与機構・研究開発部・准教授

研究者番号：20282866

交付決定額(研究期間全体)：(直接経費) 3,300,000円

研究成果の概要(和文)：マルチエージェント学習における経験強化型学習XoLに関する発展的研究として、複数のエージェントが同時に学習することによって生じる同時学習問題を回避する手法を提案した。また、マルチエージェント学習における間接報酬の正の効果の解析を行った。具体的には、「間接報酬により不完全知覚問題を軽減する手法」を提案するとともに、「全エージェントが報酬を得るための定理」としてのとりまとめを行った。さらに、Keepawayタスク等への適用を通じて、マルチエージェント学習におけるXoLの有効性を示すための応用例の拡充を図った。これらの成果により、マルチエージェント学習におけるXoLの優位性を強く主張できたと考える。

研究成果の概要(英文)：This research has achieved several progresses about theory and applications of Exploitation-oriented Learning XoL in multi-agent learning. In multi-agent learning, it is important to avoid the concurrent learning problem that occurs when multiple agents learn simultaneously. Firstly, we have proposed a method to avoid the problem. Secondly, we have focused on a positive effect of an indirect reward which is given to the agent that does not receive a reward directly. Especially, we have proposed a method to reduce the perceptual aliasing problem caused by imperfect perception. We have also described the relationship between our previous multi-agent learning theorem and the positive effect. Lastly, we have extended application areas to show the effectiveness of XoL in multi-agent learning through experiments to Keepaway tasks like soccer games. We believe that these results contribute to claim that XoL surpasses traditional reinforcement learning methods in multi-agent learning.

研究分野：機械学習

キーワード：マルチエージェントシステム 経験強化型学習 強化学習 機械学習 人工知能 ソフトコンピューティング

1. 研究開始当初の背景

近年、我が国のロボット技術の進展には目覚ましいものがある。特に、急速な少子高齢化の進行に伴い、介護・子守ロボットなど、人と直接触れ合うことが要求されるロボットの必要性が増し、人の気持ちが理解できる高度な知能(優しさ)を持ったロボットの登場が待ち望まれている。また、先の大震災での原発事故処理用ロボットなど、自律的に判断・行動できるロボット開発の必要性も高まっている。これらのロボットでは、既知の部分は予めプログラミングすることができるが、それ以外の未知な部分は、人との触れ合いや環境との相互作用を通して、試行錯誤的に、機械自らが学習し、与えられた目的(任務)を遂行することが求められる。

そのような「試行錯誤に基づく目的指向の学習」は、現在、「強化学習」において集中的に研究されている。強化学習では、結果の善し悪しのみを表す「報酬」および「罰」という弱い教師信号により学習が進行する。正解を与えることなしに学習できる点がいへん魅力的な枠組みではあるが、キラーアプリケーションと呼べるような応用例が少ないのも事実である。

これには主として次のふたつの点に関係していると考えられる。第一に、学習に時間がかかりすぎる点、第二に、適切な学習をさせるための報酬(罰)の設計指針が存在しない点である。これらはともに、理論的な解析においては問題とされない場合が多いが、手法の応用を考えた場合、深刻な問題を引き起こす。

これに対し、研究代表者らは、報酬(罰)に値を設定せず、あくまで目的達成時(制約違反時)の信号として扱う立場をとっている。また、得られた経験を強く強化することで、試行錯誤回数の軽減も同時に目指している。現在、これらの特徴を満たす接近法として「経験強化型学習 Exploitation-oriented Learning (XoL)」を提唱している。

研究代表者らは、これまで、「複数種類の報酬と罰に対応した XoL 手法」を提案するとともに、応用の際に特に重要となる「報酬と罰の設計指針」を提示した。また、XoL を利用した応用例の拡充を図ってきた。一方、その過程で、複数のエージェントが同時に学習を行う「マルチエージェント学習」の重要性を強く感じてきた。今後、XoL のさらなる適用領域の拡充を図るために、シングルエージェントのみならず、マルチエージェント環境下での研究の進展が特に重要であると考え、本研究課題を着想するに至った。

2. 研究の目的

本研究課題では、複数のエージェントが同時に学習を行う「マルチエージェント学習」における「経験強化型学習 Exploitation-oriented Learning (XoL)」に関する発展的研究を行う。具体的には、

(1) マルチエージェント学習における「同時

学習問題」に対する新たな手法の提案、
(2) マルチエージェント学習における「間接報酬」の正の効果の解析、
および
(3) マルチエージェント学習を用いた応用例の拡充、
の3つを主目的に研究を進める。

3. 研究の方法

本研究課題では、まず平成 26 年度に、複数のエージェントが同時に学習することによって生じる「同時学習問題」を回避する手法を、XoL の一手法である「失敗確率伝播アルゴリズム(EFP)」に組み込む。具体的には、これまでに、同時学習問題の発生を検出する方法として提案されている最小二乗法やポルツマン分布を活用した方法の問題点を指摘し、適切な改良を試みる。

続く平成 27 年度には、かつて研究代表者らが証明した「マルチエージェント環境下での間接報酬の悪影響を防ぐための定理」に対し、間接報酬の正の効果に着目した解析を追加する。具体的な正の効果としては、一般に「不完全知覚問題」として知られる、感覚入力の不完全に起因する困難さの軽減を考える。その上で、間接報酬の正の効果に関する十分条件を見出し、間接報酬に関する定理の取りまとめを目指す。

その後、最終年度にかけて、平成 26 年度および 27 年度に行った成果を踏まえた、XoL を用いたマルチエージェント学習の応用例の拡充を図る。

4. 研究成果

(1) マルチエージェント学習における同時学習問題に対する新たな手法を提案した。

具体的には、平成 26 年度において、XoL における同時学習問題の扱いを整理するとともに、同時学習問題を回避する手法の提案に成功した。

本研究成果は、平成 27 年 3 月に開催された「電気学会 システム研究会 機械学習応用研究の最前線」において、「マルチエージェント学習における 2 次系の必要性に関する研究」という題目で発表するとともに、研究会資料として、その内容を公表した。

そこでは、平成 26 年 11 月に「Procedia Computer Science 誌」に掲載された「The Necessity of a Secondary System in Machine Consciousness」の中で提案した「2 次系 (Secondary System)」と呼ばれる概念を、XoL の一手法である Profit Sharing と組み合わせることで、同時学習問題の解消を図っている。また、同時学習問題が生じる一般的な構造を示すことで、XoL において同時学習問題を取り扱う際の基本方針を明確にした。

(2) マルチエージェント学習における間接報酬の正の効果の解析した。

具体的には、平成 27 年度において、まず

初めに「間接報酬により不完全知覚問題を軽減する手法」を提案した。さらに、以前証明した間接報酬に関する定理との関係を整理し、「全エージェントが報酬を得るための定理」としてのとりまとめを行った。

本研究成果は、平成 28 年 3 月に開催された「電気学会 システム研究会 機械学習応用研究の最前線」において、「マルチエージェント環境における間接報酬に関する一考察」という題目で発表するとともに、研究会資料として、その内容を公表した。

かつて研究代表者らが平成 11 年度に公表した間接報酬に関する定理は、直接報酬を得たエージェント以外のエージェントに間接報酬を分配する際に、間接報酬が悪影響を及ぼさないための定理であった。それは間接報酬の「負の効果」を解析したものである。それに対し、今回、不完全知覚問題の軽減という間接報酬の「正の効果」に貢献する手法が提案できたことは、今後、マルチエージェント学習を進展させる上で、重要な意義をもつと考える。

(3) マルチエージェント学習を用いた応用例を拡充させた。

具体的には、平成 28 年度において、Keepaway Task と呼ばれるサッカーゲームを模したマルチエージェント環境への応用を行った。

本研究成果は、まず初めに、平成 28 年 9 月に「International Workshop on Multiagent Learning: Theory and Applications」において、プロトタイプ的な手法に関する発表を行った。その後、平成 28 年 12 月に開催された「14th European Conference on Multi-Agent Systems」では、先の手法を進展・拡充させた「Proposal of an Action Selection Strategy with Expected Failure Probability and its Evaluation in Multi-agent Reinforcement Learning」という題目の発表を行い、論文を公表した。

また、近年、深層学習が大きな広がりを見せているので、深層学習を利用した応用例として、平成 28 年 7 月に開催された「電気学会 システム研究会」において、ATARI 2600 ゲームを題材とした研究発表を行った。ここでは、Deep Q-Network として知られる既存手法に XoL の一手法である Profit Sharing を組み合わせることで、大幅なる性能の改善を実現した。なお、本研究成果はシングルエージェント環境下のものであるが、現在は、本研究課題の目的であるマルチエージェント環境下への適用を進めている。特に、Palro、Rapiro、LEGO ロボット等の異なる種類のロボットにより構成されるマルチロボット環境下への意識的意思決定システムの実装を中心に研究を進めている。

以上の成果は、それぞれ、国内の学会や国際会議で発表し高い評価を得るとともに、国

際的な学術雑誌に掲載され、広く国際的にアピールすることにも成功している。また、平成 28 年 11 月に開催された「第 6 回知能工学部会研究会『賢さの先端研究会』、第 54 回システム工学部会研究会 機械学習の最先端研究 - 理論および応用研究 - 」においては、招待講演として、本研究課題の成果をまとめて発表した。さらに、平成 28 年 10 月に発行された書籍「これからの強化学習」では、本研究課題で得られた成果等に基づき、「試行錯誤回数の低減を指向した手法：経験強化型学習 XoL」という章を執筆した。

本研究課題で得られた成果は、今後、XoL に関する研究を推進・発展させる上で、中心的な役割を担うものである。また、XoL 以外の強化学習手法を含めたマルチエージェント学習一般に対しても、国内外に与える影響の大きい成果が得られていると考える。

今後は、実ロボットへの適用や、深層学習との組み合わせ等を通じて、さらなる手法の発展・洗練化を行う予定である。

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 5 件)

Kazuteru Miyazaki, Koudai Furukawa and Hiroaki Kobayashi, Proposal of an Action Selection Strategy with Expected Failure Probability and Its Evaluation in Multi-agent Reinforcement Learning, Lecture Notes in Computer Science, 査読有, Vol.10207, pp.1-15, 2017, to appear.

Kazuteru Miyazaki, A Study of an Indirect Reward on Multi-agent Environments, Procedia Computer Science, 査読有, Vol.88, pp.94-101, 2016.

DOI: 10.1016/j.procs.2016.07.411

村岡宏紀、宮崎和光、小林博明、失敗確率伝播アルゴリズム EFP の提案とマルチエージェント環境下での有効性の検証、電気学会論文誌 C、査読有、Vol.136, No.3, pp.273-281, 2016.

DOI: 10.1541/ieejieiss.136.273

Kazuteru Miyazaki and Junichi Takeno, The Necessity of a Secondary System in Machine Consciousness, Procedia Computer Science, 査読有, Vol.41, pp.15-22, 2014.

DOI: 10.1016/j.procs.2014.11.079

[学会発表](計 16 件)

Kazuteru Miyazaki, Koudai Furukawa and Hiroaki Kobayashi, Proposal of an Action Selection Strategy with Expected Failure Probability and its Evaluation in Multi-agent

Reinforcement Learning, 14th European Conference on Multi-Agent Systems, 2016年12月15日, ヴァレンシア(スペイン)

宮崎和光、経験強化型学習 XoL - 強化学習における試行錯誤回数の低減をめざして - (招待講演) 第6回知能工学部会研究会『賢さの先端研究会』、第54回システム工学部会研究会 機械学習の最先端研究 - 理論および応用研究 - 、2016年11月7日、フォーラムミカサ エコ 7F ホール(東京都・千代田区)

Kazuteru Miyazaki, Koudai Furukawa and Hiroaki Kobayashi, Proposal and Evaluation of an Action Selection Strategy with Expected Failure Probability in Multiagent Learning, International Workshop on Multiagent Learning: Theory and Applications, 2016年9月30日, くにびきメッセ(島根県・松江市)

宮崎和光、深層学習を組み込んだ経験強化型学習 XoL: deep Q-network との比較、電気学会 システム研究会「機械学習応用研究の最前線」、2016年7月9日、星陵会館(東京都・千代田区)

宮崎和光、マルチエージェント環境における間接報酬に関する一考察、電気学会 システム研究会「機械学習応用研究の最前線」、2016年3月8日、東京トラック事業健保会館(東京都・千代田区)

Kazuteru Miyazaki, The Necessity of a Secondary System in Multi-agent Learning, The First International Symposium on Swarm Behavior and Bio-Inspired Robotics, 2015年10月29日, 京都大学(京都府・京都市左京区)

宮崎和光、マルチエージェント学習における2次系の必要性に関する研究、電気学会 システム研究会「機械学習応用研究の最前線」、2015年3月11日、青山学院大学 相模原キャンパス(神奈川県・相模原市)

[図書](計1件)

牧野貴樹、澁谷長史、白川真一、浅田稔、麻生英樹、荒井幸代、飯間等、伊藤真、大倉和博、黒江康明、杉本徳和、坪井祐太、銅谷賢治、前田新一、松井藤五郎、南泰浩、宮崎和光、目黒豊美、森村哲郎、森本淳、保田俊行、吉本潤一郎、森北出版、これからの強化学習、2016年、pp.136-147 (総ページ数 320)

[産業財産権]

出願状況(計0件)

取得状況(計0件)

[その他]

ホームページ等

<http://www7b.biglobe.ne.jp/~kazuteru/index.html>

<http://researchmap.jp/read0047214/>

https://www.researchgate.net/profile/Kazuteru_Miyazaki

<https://scholar.google.co.jp/citations?user=JdZXKI4AAAAJ&hl=ja>

6. 研究組織

(1) 研究代表者

宮崎 和光 (MIYAZAKI KAZUTERU)

独立行政法人大学改革・学位授与機構・研究開発部・准教授

研究者番号：20282866