様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　２ 年　 ７ 月　 ７ 日現在

機関番号：３２６５７
研究種目：若手研究(A)
研究期間：2017～2019
課題番号：１７Ｈ０４６９６

研究課題名（和文）限定合理性を備えた深層強化学習理論の展開

研究課題名（英文）Developing a theory of deep reinforcement learning equipped with bounded rationality

研究代表者
高橋　達二（Takahashi, Tatsuji）

東京電機大学・理工学部・准教授

研究者番号：００５１４５１４

交付決定額（研究期間全体）：（直接経費）　　18,200,000円

研究成果の概要（和文）：実世界で活動する人間、動物、ロボットは、知覚の能力・情報処理の速度と容量・行動の効果、の三点それぞれにおいて制約のある状況で、各々のゴールの達成を目指して合理的（限定合理的）に学習・行動を行う。本研究はそれが「最適化」の代替案としての「満足化」という探索・意思決定の方策により可能になっていると仮定し、満足化に新しい実装を与え、工学的に有用なアルゴリズムとして世界で初めて確立するとともに、その性質について数学的に明らかにした。またそのアルゴリズムを、強化学習の分野において様々なタスクに適用し、最も基本的なバンディット問題や、一般的な強化学習タスクにおいてその有効性を示した。

研究成果の学術的意義や社会的意義
人間や動物の扱う、試行錯誤を伴う自律的な学習のロジックの重要な一端を明らかにした。特に、なぜ人間や動物が競争と「対抗模倣」により効率的なパフォーマンスの向上を見せるのかについて機械論的な説明を与えた。さらに、数学的に効率性を証明するとともに、様々な状況で効率性を示した。また、資本主義や市場の観点から、競争や対抗模倣の効率性と、表裏一体であるその危険性についても論じた。

研究成果の概要（英文）：The real world agents such as human beings or animals learn and act in some (bounded) rational way toward their respective goals. The learning and acting are under severe restrictions as for perception, information processing, and actuation. In this project, we hypothesize that the efficient learning and acting are enabled by exploiting the search and decision-making policy called "satisficing" that was proposed as an alternative of optimization. We gave a new implementation (RS) of satisficing, establishing it as a useful algorithm, and we proved its efficiency. We applied RS to various tasks in reinforcement learning and showed its efficiency, including the most basic bandit problems and general tabular and non-tabular MDPs.

研究分野：認知科学

キーワード：限定合理性　強化学習　満足化　社会学習　弱教示的学習　判定問題　仮説検証　試行錯誤

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９（共通）

１．研究開始当初の背景

囲碁やビデオゲームなどで人を上回る性能を見せている試行錯誤学習の理論とアルゴリズムである「人工強化学習」は多大な成功を見せている。しかしその学習には無数の致命的な失敗（＝死）が必要である。このことが、強化学習がゲームやシミュレーション上でしか有効なパフォーマンスを示せない理由となっている。この学習の仕方は、個体が学習する動物というよりは、大量の個体集合で解決を図る昆虫に近い。それに比して人間や動物の試行錯誤学習には、厳しい知覚的・認知的・行動的制約を課されているにも関わらず、少数の試行から有効な行動系列を獲得できるなど、より強力な側面が数多くある。

２．研究の目的

そこで本研究では、人間や動物の「限定合理的」（制約の下で効果的・合理的な結果をもたらす）な特性を分析し、それを実装することで、強化学習に人間や動物の「自然強化学習」の強みを付与することを目的とする。

そのために、まず限定合理性を実装するモデル・アルゴリズムを確立する。また、それを用いて、一般的な強化学習タスクに適用するためのアルゴリズムを開発する。さらに、深層学習と組み合わせ、より広範なタスクに対応できる深層強化学習にも応用する。

本研究では、限定合理性について、特に「満足化 satisficing」という探索・意思決定手法にフォーカスする。満足化とは、エージェントがある基準以上の行動を探索し、それが見つかり次第探索を打ち切る、といった方策である。対比される概念は最適化であり、最適化の場合は全ての行動の中から最も優れた行動が見つかるまで、探索を続ける。

３．研究の方法

本研究では「リスク依存満足化価値関数 (risk-sensitive satisficing value function) 」の RS モデルを扱った。これは RS が非常に効率的な満足化を実装し、限定合理性を備えているためである。RS モデルは単純な形式をしている (高橋, 甲野, 浦上, 2016) ものの、その性質は必ずしも明らかではなかった。そこで、強化学習の最も基本的なタスクのクラスである K 本腕バンディット問題における RS の性質を分析する。

また、RS を強化学習一般に適用するために、大局基準値変換法 (GRC: global reference conversion) を定義し、それにより強化学習全般での満足化の効率的な実現を目する。さらに、深層学習と組み合わせて深層強化学習においてビデオゲームなどの複雑なタスクでの性能を検証する。

複雑な環境での効率的な学習と行動のためには、手持ちの情報から新しい情報を引き出す推論や、環境における因果関係を効率的に推定し、それを活用する因果推論が必要である。そのため、人間の推論と因果推論についての調査、研究も行う。

４．研究成果

RS の分析に関しては、バンディット問題において、(1) RS が（基準値が不確実であるとしても）必ず満足化をすること、(2) その場合の性能が最適であり、性能の下限が求められること（リグレット（期待損失）の上限が計算できる）、 (3) RS の行動価値の変化の期待値が常に一定で

あり、そのことにより (1) と (2) の合理的な性質が直感的にも理解できること、を示した。また、(1') 満足化の基準が最適な行動と二番目に最適な行動の間にありさえすれば、RS が極度に効率的な最適化を実現することも示した (Tamatsukuri & Takahashi, 2019)。

また、満足化基準を自律的に獲得する方法についての提案を行った。これにより、バンディット問題で最もパフォーマンスの高いアルゴリズムと見なされることの多い Thompson sampling と同等の成績を実現している (甲野 & 高橋, 2018)。更に、非定常環境においての RS の有効性を示した (花安 et al., 2019; 齋藤 & 高橋, 2019)。

RS の強化学習一般への応用に関しては、上述の GRC により効率的に実現されている (Shinriki et al., 2020; 其田 & 高橋, 2019; 其田, 神谷 & 高橋, 2019; 佐鳥 et al., 2019; 齋藤 & 高橋, 2019; 佐鳥 et al., 2018; 其田 et al., 2018; 牛田 et al., 2017。ただしパラメータの設定が難しいため、その自律的な設定アルゴリズムを開発した (佐鳥 et al., 2020)。また深層強化学習については、擬似カウントを用いた探索ボーナスと、オートエンコーダーを用いた手法の二つにおいて適用に成功している (佐鳥, 吉田, 神谷, 高橋, 2019)。

この強化学習一般への応用の過程で、社会学習における満足化と RS の有効性について明らかになった (其田 & 高橋, 2019; Shinriki et al., 2020; 其田, 神谷 & 高橋, 2019)。すなわち、一見設定が難しいと見える満足化の基準であるが、これが他エージェントから得られるとすれば、効率的なグループでの学習を行うことができる。この点に関して、工学、哲学、理学、経済学の点からの分析を行った (高橋, 2019)。

複雑な環境での効率的な学習と行動のためには、手持ちの情報から新しい情報を引き出す推論や、環境における因果関係を効率的に推定し、それを活用する因果推論が必要である。そのため、人間の推論と因果推論についての調査、研究も行った (Shinohara et al., 2020, 2020; Baratgin et al., 2018; Nakamura et al., 2018; Hattori et al., 2017; 宝田 & 高橋, 2019; Kamiya & Takahashi 2018; Yokokawa et al., 2018; Takahashi et al., 2020)。

全体として、人間や動物の効率的な学習と推論を単純な理論とアルゴリズムにより表現し、またその工学的有用性を示した。また、限定合理性と社会性の繋がりを明らかにした。今後は 2020 年度より科研費基盤 B (20H04259)「内発的動機付けと社会性の統合による自然強化学習の実現」として同様の研究を継続していく。

＜引用文献＞

高橋 達二, 甲野 佑, 浦上 大輔: 認知的満足化|限定合理性の強化学習における効用, 人工知能学会論文誌, 31, 6, AI30-M-1-11 (2016). doi:10.1527/tjsai.AI30-M

５．主な発表論文等

〔雑誌論文〕　計7件（うち査読付論文　7件／うち国際共著　3件／うちオープンアクセス　5件）

１．著者名 Baratgin Jean、Politzer Guy、Over David E.、Takahashi Tatsuji	４．巻 9
２．論文標題 The Psychology of Uncertainty and Three-Valued Truth Tables	５．発行年 2018年
３．雑誌名 Frontiers in Psychology	６．最初と最後の頁 1479
掲載論文のDOI（デジタルオブジェクト識別子） 10.3389/fpsyg.2018.01479	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

１．著者名 Tamatsukuri Akihiro、Takahashi Tatsuji	４．巻 180
２．論文標題 Guaranteed satisficing and finite regret: Analysis of a cognitive satisficing value function	５．発行年 2019年
３．雑誌名 Biosystems	６．最初と最後の頁 46～53
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.biosystems.2019.02.009	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

１．著者名 Shinohara Shuji、Manome Nobuhito、Suzuki Kouta、Chung Ung-il、Takahashi Tatsuji、Okamoto Hiroshi、Gunji Yukio Pegio、Nakajima Yoshihiro、Mitsuyoshi Shunji	４．巻 15
２．論文標題 A new method of Bayesian causal inference in non-stationary environments	５．発行年 2020年
３．雑誌名 PLOS ONE	６．最初と最後の頁 e0233559
掲載論文のDOI（デジタルオブジェクト識別子） 10.1371/journal.pone.0233559	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

１．著者名 Shinohara Shuji、Manome Nobuhito、Suzuki Kouta、Chung Ung-il、Takahashi Tatsuji、Gunji Pegio-Yukio、Nakajima Yoshihiro、Mitsuyoshi Shunji	４．巻 190
２．論文標題 Extended Bayesian inference incorporating symmetry bias	５．発行年 2020年
３．雑誌名 Biosystems	６．最初と最後の頁 104104～104104
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.biosystems.2020.104104	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

１．著者名 Nakamura Hiroko、Shao Jing、Baratgin Jean、Over David E.、Takahashi Tatsuji、Yama Hiroshi	４．巻 9
２．論文標題 Understanding Conditionals in the East: A Replication Study of Politzer et al. (2010) With Easterners	５．発行年 2018年
３．雑誌名 Frontiers in Psychology	６．最初と最後の頁 505
掲載論文のDOI（デジタルオブジェクト識別子） 10.3389/fpsyg.2018.00505	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

１．著者名 Hattori Ikuko、Hattori Masasi、Over David E.、Takahashi Tatsuji、Baratgin Jean	４．巻 23
２．論文標題 Dual frames for causal induction: the normative and the heuristic	５．発行年 2017年
３．雑誌名 Thinking & Reasoning	６．最初と最後の頁 292～317
掲載論文のDOI（デジタルオブジェクト識別子） 10.1080/13546783.2017.1316314	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

１．著者名 Shinriki Moto、Wakabayashi Hiroaki、Kono Yu、Takahashi Tatsuji	４．巻 1128
２．論文標題 Flexibility of Emulation Learning from Pioneers in Nonstationary Environments	５．発行年 2020年
３．雑誌名 Advances in Artificial Intelligence. JSAI 2019. Advances in Intelligent Systems and Computing	６．最初と最後の頁 90～101
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/978-3-030-39878-1_9	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

〔学会発表〕　計17件（うち招待講演　0件／うち国際学会　2件）

１．発表者名 佐鳥 玖仁朗，吉田 豊，山岸 健太，牛田 有哉，神谷 匠，高橋 達二
２．発表標題 満足化原理の強化学習全般への適用に向けて
３．学会等名 2018年度人工知能学会全国大会（第32回）(JSAI 2018)
４．発表年 2018年

１．発表者名
玉造 晃弘，高橋 達二

２．発表標題
認知的満足化価値関数の分析

３．学会等名
2018年度人工知能学会全国大会（第32回）(JSAI 2018)

４．発表年
2018年

１．発表者名
甲野 佑，高橋 達二

２．発表標題
満足化を通じた最適な自律的探索

３．学会等名
2018年度人工知能学会全国大会（第32回）(JSAI 2018)

４．発表年
2018年

１．発表者名
其田 憲明，神谷 匠，甲野 佑，高橋 達二

２．発表標題
満足化基準値共有を用いた社会的強化学習

３．学会等名
2018年度人工知能学会全国大会（第32回）(JSAI 2018)

４．発表年
2018年

１．発表者名
高橋 達二，大用 庫智，玉造 晃弘，横川 純貴

２．発表標題
稀少性仮定の下での非独立性の判断としての人間の観察的因果推論

３．学会等名
2017年度人工知能学会全国大会（第31回）(JSAI 2017)

４．発表年
2017年

１．発表者名
牛田 有哉，甲野 佑，高橋 達二

２．発表標題
生存を目的とする満足化強化学習

３．学会等名
2017年度人工知能学会全国大会（第31回）(JSAI 2017)

４．発表年
2017年

１．発表者名
Kamiya, T., Takahashi, T.

２．発表標題
Are word learning biases based on symmetry in cognition?

３．学会等名
The Twenty-Third International Symposium on Artificial Life and Robotics 2018 (AROB 23rd 2018)（国際学会）

４．発表年
2018年

１．発表者名
Yokokawa, J., Oyo, K., Takahashi, T.

２．発表標題
Causal induction under rarity and small data

３．学会等名
The Twenty-Third International Symposium on Artificial Life and Robotics 2018 (AROB 23rd 2018)（国際学会）

４．発表年
2018年

１．発表者名
佐鳥玖仁朗，神谷匠，高橋 達二

２．発表標題
弱教示的強化学習における探索割合の自律調整

３．学会等名
情報処理学会第82回全国大会（IPSJ 2020）

４．発表年
2020年

１．発表者名
其田憲明，高橋 達二

２．発表標題
満足化と記録共有による対抗模倣の強化学習的モデリング

３．学会等名
日本認知科学会第36回大会

４．発表年
2019年

１．発表者名
宝田 悠，高橋 達二

２．発表標題
少数データからの観察的因果帰納

３．学会等名
2019年度人工知能学会全国大会(第33回) (JSAI 2019)

４．発表年
2019年

１．発表者名
其田 憲明，神谷 匠，高橋 達二

２．発表標題
大局基準値共有による社会的強化学習

３．学会等名
2019年度人工知能学会全国大会(第33回) (JSAI 2019)

４．発表年
2019年

１．発表者名
花安 勇人，齋藤 建志，吉井 佑輝，甲野 佑，高橋 達二

２．発表標題
非定常環境における認知的満足化価値関数の適応性能

３．学会等名
2019年度人工知能学会全国大会(第33回) (JSAI 2019)

４．発表年
2019年

１．発表者名
佐鳥 玖仁朗，吉田 豊，神谷 匠，高橋 達二

２．発表標題
深層満足化強化学習に向けて

３．学会等名
2019年度人工知能学会全国大会(第33回)（JSAI 2019）

４．発表年
2019年

１．発表者名
齋藤 建志，高橋 達二

２．発表標題
認知的満足化による環境変化への適応

３．学会等名
情報処理学会第81回全国大会（IPSJ 2019）

４．発表年
2019年

１．発表者名
齋藤 建志，高橋 達二

２．発表標題
認知的満足化アルゴリズムの木探索への応用

３．学会等名
情報処理学会第81回全国大会（IPSJ 2019）

４．発表年
2019年

１．発表者名
牛田 有哉，甲野 佑，高橋 達二

２．発表標題
強化学習と満足化による素早い行動系列の獲得

３．学会等名
情報処理学会第79回全国大会（IPSJ 2017）

４．発表年
2017年

〔図書〕　計3件

1．著者名 Shira Elqayam, Igor Douven, Jonathan St B. T. Evans, Nicole Cruz	4．発行年 2020年
2．出版社 Routledge	5．総ページ数 274
3．書名 Logic and Uncertainty in the Human Mind	

1．著者名 Mykel J. Kochenderfer、繁桝 算男、本村 陽一、麻生 英樹、大西 正輝、河本 満、古野 公紀、繁桝 算男、高橋 達二、中島 秀之、宮澤 芳光、本村 陽一、山崎 啓介	4．発行年 2020年
2．出版社 共立出版	5．総ページ数 434
3．書名 不確定性下の意思決定	

1．著者名 郡司ペギオ幸夫，吉川浩満，山本貴光，久木田水生，高橋達二	4．発行年 2019年
2．出版社 一般社団法人 大学出版部協会	5．総ページ数 40
3．書名 大学出版 119号	

〔産業財産権〕

〔その他〕

—

6．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考