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　５ 年　 ６ 月 １７ 日現在

機関番号：３２６５７
研究種目：基盤研究(B)（一般）
研究期間：2020～2022
課題番号：２０Ｈ０４２５９

研究課題名（和文）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

研究課題名（英文）Natural reinforcement learning integrating intrinsic motivation and sociality

研究代表者
高橋　達二（Takahashi, Tatsuji）

東京電機大学・理工学部・教授

研究者番号：００５１４５１４

交付決定額（研究期間全体）：（直接経費）　　13,700,000円

研究成果の概要（和文）：本研究では、報酬、動機づけ、計算理論的な問題定式化、そして社会性の観点から、強化学習理論の見直しを行い、人間や動物の扱う「自然強化学習」の長所を強化学習アルゴリズムに採り入れた。成果として、理論的には主観リグレット概念による、限定合理性・意思決定・採餌理論の統合に成功した。産業的な応用も行った他、不確実性の下でのエミュレーション的な社会学習の原理を定式化した。マウスに関しては本研究の理論を一般化しうる興味深い結果を得た。

研究成果の学術的意義や社会的意義
人間や動物がどのように不確実な環境において学習しているかについての知見を深めました。これは今後、教育、訓練、社会活動などをどのように行うべきかについて指針を与える可能性があります。また、ChatGPTなどが人間と対話できるようにするために肝要な強化学習技術について、学習の目標を定めれば、それに向かって非常に効率的に学習を行えるようになりました。これは、生成AI、ゲーム技術、ロボット制御などにおいて広範な応用を得る可能性があります。

研究成果の概要（英文）：aIn this project, we have formalized the mechanisms and merits of the natural reinforcement learning that humans and animals do. The formalization was done reconsidering the concepts of reward, motivation, task formalization (in terms of theory of computation), and sociality. Theoretically, we succeeded in a unification of bounded rationality, decision-making, and foraging theories from the notion of subjective regret. Some industrial applications were done and a principle of social learning under uncertainty was formulated. We also found that mice adaptively control the (asymmetric) learning rates under uncertainty, according to the environments that they face. It leads to a generalization of our theory.

研究分野：計算論的認知科学

キーワード：強化学習　満足化　限定合理性　動物実験　機械学習

１．研究開始当初の背景

人間や動物が実際に行う、試行錯誤を通じた効果的な行動の獲得である「自然強化学習」を参考に、人工知能技術としての「人工強化学習」が開発された。近年は深層学習を組み合わせた深層強化学習技術が発展し、囲碁や多様なビデオゲームのプレイなどで人間を上回るパフォーマンスを見せた。その後、強化学習の最先端のベンチマークタスクはチームプレイを含んだゲーム(Dota2)など、協力や競争、分業などの社会性が必要なものにシフトしてきており、社会的な動機付けの理論化が必要である。社会性を扱うためには、自然強化学習に固有な、報酬レベルに対するある基準（希求水準）という形での、他個体のパフォーマンスレベルへの参照が必要である。そこで本研究は、次の二つの問いに答える：

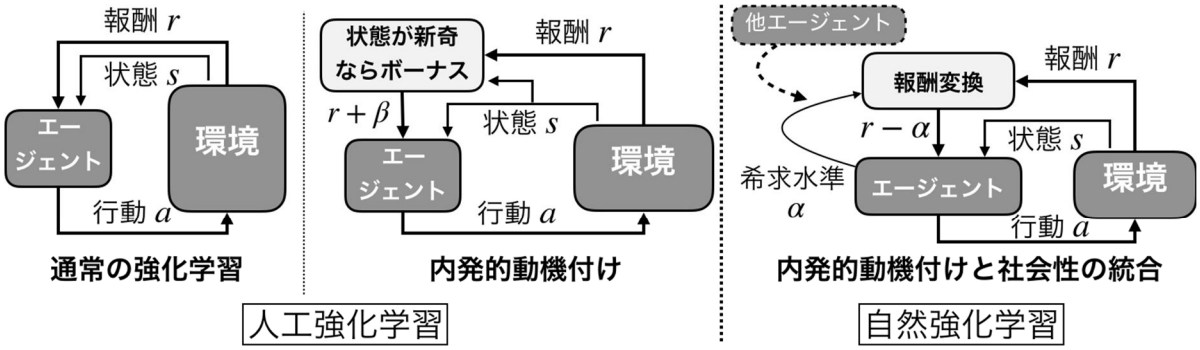
1. どうすれば自然強化学習とその特徴である社会的な動機付けをモデリングできるか
2. 自然強化学習・社会的な動機付けにはどのようなメリットがあるのか

２．研究の目的

自然強化学習と人工強化学習の間にある様々な違いのうち、本研究では動機付け概念の分析を行い、社会性を導入するとともに、強化学習の根本概念である報酬と、タスクの問題としての定式化を刷新し、新しい強化学習の理論とアルゴリズムを開発する。

人工強化学習においては、環境がエージェントに与える報酬という外発的動機付けによってのみ学習が進行するが、それだけでは、報酬がまばら(スパース)にしか与えられない現実的なタスクには全く対処できない。そこで近年、人間や動物の好奇心を模し、新奇な状態の訪問にボーナス β を与える内発的動機付けが探索領域を広げることに成功しているが、広大な状態行動空間では不要な探索を膨大に行って環境把握を続けるという問題がある。

それに対して人間や動物の自然強化学習においては、ある基準(希求水準) α により自ら内発的動機付けが調整され、自律的に探索範囲が拡大縮小される。希求水準と比べて現状がより低い、つまり不満足な場合は新奇な行動を好んで探索し（リスク選好、好奇心)、現状に満足な場合は保守的に探索を控える（リスク回避、負の好奇心)。希求水準 α はエージェント自らの必要（生存のための食料など）だけでなく他エージェントとの関係（分業や協力、あるいは対抗関係といった社会的要因）にも依存し、内発的動機付けに基準を与える（内発的動機付けと社会性の統合)。つまり α 決定の再帰的プロセスが社会的入力を持つ（下図)。



※ 報酬への足し算 r+β と引き算 r-α は全く異なる操作である。引き算は r=α の場合に変換後の報酬を0とし、報酬の原点を設定する。r-α>0 なら「十分」、r-α<0 なら「不十分」という意味付け（弱い教示）が与えられ、探索範囲が自律的に調整される。

また自然強化学習においては、強化学習の根本である報酬や、環境探索を制御する動機付け、そして計算論的な問題の定式化の三点が変容する。この点を下表にまとめる。

	人工強化学習 既存の強化学習アルゴリズム	**自然強化学習** 本研究で実現する人間・動物の強化学習
動機付け	外発的：報酬 内発的：リスク選好（好奇心）	外発的：報酬、**競争** (対抗模倣) 内発的：リスク選好 or **リスク回避、生存・満足**
報酬	純粋な**評価的**フィードバック （報酬の値自体に意味**なし**）	**弱い教示的**フィードバックに**変換** （報酬の値自体に意味**あり**）
問題定式化	**最適化**問題 純粋に**ボトムアップ**に進行	**決定**問題 (decision problem) としても **トップダウン**制約 (一種の仮説)を調整しながら進行

３．研究の方法

自然強化学習の特性を実現するアイディアとして Simon が提唱した満足化(satisficing)がある。これは最適化の代替案であり、最適化が損失関数の最小化を追求して探索し続けるのに対し、満足化はある基準(希求水準)より優れた行動が見つかればそれで探索を打ち切る行動方針である。例えば狩猟採集者は、食物の獲得量の最大化の追求というよりは、自分や家族が飢えないという根本的な最低基準の達成や、最近の平均や昨年の同じ季節よりも多い狩猟・採集の達成を目標とし、達成後は食物以外の他の次元での(メイティングや安全の確保などの)探索に力を注ぐであろう。

満足化のアイディアは行動の記述モデルとしては妥当とされながら、一般性のあるモデリングはされてなかった。これに関しこれまで科研費 2 件を受けて研究代表者が開発した満足化価値関数 RS (risk-sensitive satisficing) がある [高橋, 甲野 & 浦上, 2016]。状態 s と行動 a について、状態行動価値を報酬平均 Q(s,a)，それまでに s で a を実施した回数を n(s,a)，生存に必要な食物の単位量といった基準を α として、s で a を行うことの価値を

$$RS(s, a) = n(s, a) \cdot [Q(s, a) - \alpha]$$

と定義する。行動の価値 Q(s,a) が、行動価値と基準 α の差 [Q(s, a) - α] に変換されており、基準以上であれば正、基準以下なら負の値を取り、この符号が十分か不十分かを表現している。また、RS は n(s, a) 回分の報酬 ri と基準 α との差の合計であり、

$$RS(s, a) = (r1 - \alpha) + (r2 - \alpha) + \cdots + (rn - \alpha).$$

つまり、報酬＝評価的フィードバックが、弱教示的フィードバック (基準以下か以上かという情報をその符号としてコード）に変換されている。報酬 r という外発的動機付けに対し、α がエージェントの内的目標として内発的動機付けを左右する。RS は α 以上の価値をもつ行動が存在すればそれを最適効率で発見できることが証明でき[Tamatsukuri & Takahashi, 2019]、強化学習タスクを最適化問題ではなく決定問題として解くことができる。RS は、α を各状態に割り振ることで強化学習全般に適用でき、また「擬似カウント技術」などと組み合わせて深層強化学習でも有効性が示せている(科研費若手研究(A)2017-2019 年度での成果)。

本研究では、この RS モデルを自然強化学習の中心的なモデルとして、その分析と応用を行う。

４．研究成果

RS に関しては、多数の基礎的・応用的な結果が得られ、学会発表などを行ったが、ここではその中から特に 3 つについてのみ触れる。

RS は上記のように決定論的な形を持っており、実際の行動データとの比較が難しかった。これに関し、RS を softmax 形式の確率分布として表現し、またリスク対応型採餌行動 (risk-sensitive foraging) との定性的な一致を確認した [Kamiya & Takahashi 2022]。これにより、今後の行動データの分析が可能となった。また、自然強化学習の、プロスペクト理論や採餌行動理論などとの関係を明確化する一歩を踏み出すことができた。

実験的には、マウスが不確実性の下で適応的に学習率を制御していることが判明し、これは本研究の理論を一般化しうる興味深い結果である [Ohta et al. 2021]。具体的には、マウスは 5 本腕バンディット問題を行った。満足化のアイディアを用いて導いたモデルにより、マウスの実際の行動選択履歴を、既存モデル(23 種類)と比べて最も良く予測できている。マウスは強化学習を担う基底核を持つ一方、記号的推論を行う大脳皮質を発達させておらず純粋な自然強化学習の行動データを提供するため理論の検証能力が高い。

また、RS と類似の形をし、ある前提の下で相互に変換可能な pARIs モデルが、人間の効率的な因果推論（因果探索）を可能にすることを、メタ分析とシミュレーションで確かめた [Higuchi, Oyo, and Takahashi 2023]。自然強化学習と因果推論の関係は、人間や動物の学習という意味では重要性が高く、この点は今後の研究の価値がある。

全体として、当初の目標を果たし、多方面でアイディアの有効性を示したほか、今後の広範な現象説明や広い分野での応用を準備する結果が得られた。意思決定のプロスペクト理論、神経科学で着目されてきた採餌行動理論、社会学習などとの自然強化学習の関連性も具体的に明らかになってきた。さらに限定合理性やそのアップデートと目され、心・脳・AI を統合的に理解する枠組みと期待される計算論的合理性といった広い文脈で本研究の意義が今後示されていくと考えられる。

５．主な発表論文等

〔雑誌論文〕　計9件（うち査読付論文　9件／うち国際共著　0件／うちオープンアクセス　7件）

１．著者名 Higuchi Kohki、Oyo Kuratomo、Takahashi Tatsuji	４．巻 225
２．論文標題 Causal intuition in the indefinite world: Meta-analysis and simulations	５．発行年 2023年
３．雑誌名 Biosystems	６．最初と最後の頁 104842～104842
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.biosystems.2023.104842	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 Kamiya Takumi、Takahashi Tatsuji	４．巻 213
２．論文標題 Softsatisficing: Risk-sensitive softmax action selection	５．発行年 2022年
３．雑誌名 Biosystems	６．最初と最後の頁 104633～104633
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.biosystems.2022.104633	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 Fukuchi Yosuke、Osawa Masahiko、Yamakawa Hiroshi、Takahashi Tatsuji、Imai Michita	４．巻 9
２．論文標題 Conveying Intention by Motions With Awareness of Information Asymmetry	５．発行年 2022年
３．雑誌名 Frontiers in Robotics and AI	６．最初と最後の頁 783863
掲載論文のDOI（デジタルオブジェクト識別子） 10.3389/frobt.2022.783863	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 Ohta Hiroyuki、Satori Kuniaki、Takarada Yu、Arake Masashi、Ishizuka Toshiaki、Morimoto Yuji、Takahashi Tatsuji	４．巻 143
２．論文標題 The asymmetric learning rates of murine exploratory behavior in sparse reward environments	５．発行年 2021年
３．雑誌名 Neural Networks	６．最初と最後の頁 218～229
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.neunet.2021.05.030	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 －

1．著者名 Manome Nobuhito、Shinohara Shuji、Takahashi Tatsuji、Chen Yu、Chung Ung-il	4．巻 11
2．論文標題 Self-incremental learning vector quantization with human cognitive biases	5．発行年 2021年
3．雑誌名 Scientific Reports	6．最初と最後の頁 3910
掲載論文のDOI（デジタルオブジェクト識別子） 10.1038/s41598-021-83182-4	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 池田 駿介、布山 美慕、西郷 甲矢人、高橋 達二	4．巻 28
2．論文標題 不定自然変換理論に基づく比喩理解モデルの計算論的実装の試み	5．発行年 2021年
3．雑誌名 認知科学	6．最初と最後の頁 39～56
掲載論文のDOI（デジタルオブジェクト識別子） 10.11225/cs.2020.065	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 Fuyama Miho、Saigo Hayato、Takahashi Tatsuji	4．巻 197
2．論文標題 A category theoretic approach to metaphor comprehension: Theory of indeterminate natural transformation	5．発行年 2020年
3．雑誌名 Biosystems	6．最初と最後の頁 104213～104213
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.biosystems.2020.104213	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0件

〔図書〕　計1件

1．著者名 Takahashi Tatsuji、Oyo Kuratomo、Tamatsukuri Akihiro、Higuchi Kohki	4．発行年 2020年
2．出版社 Routledge	5．総ページ数 20
3．書名 Logic and Uncertainty in the Human Mind	

〔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
研究分担者	甲野　佑 (Kono Yu) (10870313)	東京電機大学・理工学部・講師 (32657)	
研究分担者	玉造　晃弘 (Tamatsukuri Akihiro) (10876361)	東京電機大学・理工学部・研究員 (32657)	
研究分担者	太田　宏之 (Ohta Hiroyuki) (20535190)	防衛医科大学校（医学教育部医学科進学課程及び専門課程、動物実験施設、共同利用研究施設、病院並びに防衛・薬理学・講師 (82406)	

6．研究組織（つづき）

	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
研究分担者	浦上　大輔 (Uragami Daisuke) (40458196)	日本大学・生産工学部・教授 (32665)	
研究分担者	大用　庫智 (Oyo Kuratomo) (60755685)	関西学院大学・総合政策学部・講師 (34504)	

7．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

8．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関