様　式　Ｃ－１９、Ｆ－１９、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成 ２８ 年　 ６ 月 １６ 日現在

機関番号：３２６５７
研究種目：若手研究(B)
研究期間：2013～2015
課題番号：２５７３０１５０

研究課題名（和文）人間認知の適応的特性を実装した価値関数の提案と大規模コンピューティングへの応用

研究課題名（英文）Proposal of Value Function Implementing Adaptive Cognitive Properties and Its Application to Large-Scale Computing

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）　　3,200,000円

研究成果の概要（和文）：データ量の増加やロボティクスの進展により、不確実性の下で因果関係を学習しながらより効率的に行動選択を行うことが喫緊の課題である。そこで本研究では、人間の因果関係に関する直感をモデリングした価値関数（LSモデル）の有効性を検討し、またその分析と一般化を進めた。 n本腕バンディット問題、ロボットの運動獲得、モンテカルロ木探索においてその優れたパフォーマンスを示し、また制約の多かった元々のモデルの一般化と、認知的な妥当性についての検証を進めた。

研究成果の概要（英文）：With the increasing amount of data and the progress in robotics, it is one of the urgent issues to establish a more efficient action selection algorithm while it learns causal relationship under uncertainty. In this study, we prove the effectiveness of a value function, the loosely symmetric (LS) model, that models the causal intuition of humans. In the multi-armed bandit problems, robotic action acquisition task, and Monte Carlo tree search, we showed the efficiency realized by the LS model. We also generalized the model loosening the original restrictions based on new analyses, and tested its cognitive validity by a meta-analysis and experiments.

研究分野：認知科学、知能情報学

キーワード：限定合理性　n本腕バンディット問題　満足化　強化学習　計算論的合理性

１．研究開始当初の背景

ビッグデータの普及とロボットの発達につれて自律的に情報を探索・活用するエージェントの必要性が非常に高まっているが、情報活用と探索のジレンマとそれが導く速さと正確さのトレードオフという問題が存在する。一般に仮想的探索が不可能である上に、探索空間が広大なほどエージェント数や探索時間などのリソース制限がより深刻になるため、遺伝的アルゴリズムなどの従来有効な手法は利用しづらい。そのため、できるだけ単純で特定の問題構造に依存しない汎用性と、様々なシステムに容易に組み込める可搬性を持ちながらこのトレードオフによりよく対処できるモデルの開発が喫緊の課題である。

研究代表者はこれまで、認知科学の分野で人間の推論や意思決定に関する計算論的な研究を構成論的アプローチで行い、特にリスクや不確実性に直面した際の人間の認知の偏りやクセ―「認知バイアス」―の適応的な意味を明らかにしてきた。その過程で、人間の不確実性下での、速くかつ正確な効率的行動が人工知能・経営学の満足化 (satisficing)、プロスペクト理論の期待値の信頼性/リスクの考慮、そして行動経済学の相対評価、という 3 つの極めて人間的な行動価値評価の特性によることを解明した。いずれの特性も神経科学でも研究されており、脳の具体的な部位との対応も議論されている。これらを併せ持つ経験ベイズ法の形式を持つある種の条件付き確率、「LS モデル」を研究した [篠原 2007]。

２．研究の目的

手持ちの情報を活用する局所的な最適化と、 探索を行って情報を増やすことの間のトレードオフは、探索空間が広大になればなるほど深刻な困難となる。そこで本研究では、実世界やビッグデータといった巨大な探索空間においてこのトレードオフに対処できる新しいシステムを提案する。この目的のため、進化の過程で情報の活用と探索を両立してきた人間固有の意思決定の仕方、因果関係に関する直感の傾向を再現する LS (loosely symmetric) モデルについて、これが既存のトレードオフの限界を破る有効性を持つ事を、最も基本的な n 本腕バンディット問題で示す。また、より現実的で巨大な問題においてモデルの可搬性と汎用性を示すために、強化学習によるロボット運動学習のシステム、モンテカルロ木探索によるゲーム AI や他メタヒューリスティクスによる探索手法を開発し、その性能を検証する。

３．研究の方法

LS モデルは、元来 2x2 分割表上で定義される。事象 C と E の生起不生起をそれぞれ C と¬C、E と¬E で表現し、これらの組み合わせの生起頻度を表 1 のように a, b, c, d とする。

表 1. 二事象の生起に関する 2x2 分割表

	E	¬E
C	a	b
¬C	c	d

この時、P(E|C)=a/(a+b), P(E|¬C)=c/(c+d) のように条件付き確率に対し、LS は a, b, c, d をそれぞれ a+bd/(b+d), b+ac/(a+c), c+db/(d+b), d+ca/(c+a)に変換したものである[篠原 2007]。LS の性質については [Takahashi 2010] が初等的な分析を行っており、a に加えられる bd/(b+d)のようなある種のスムージング項が LS(E|C)と LS(E|¬C)で共通である点を、ゲシュタルト心理学の図と地の分割に準えて「地の不変性」と呼んでいる。また、振る舞いとしては、LS(E|C)と LS(C|E)、LS(E|C)と LS(¬E|¬C)が高めの正の相関を持つのが特徴的である。また、a+b が c+d よりも遙かに大きい場合は、LS(E|C)が P(E|C)に(これを「焦点化」と呼ぶ)、LS(E|¬C)が 0.5 に(これを「背景化」と呼ぶ)、それぞれ収束する。ここで C と¬C を二つの行動選択肢、E と¬E を報酬 1 (報酬あり)と 0 (報酬なし)とすれば、意志決定 (例えば 2 本腕バンディット問題) において行動価値関数として使用できる。

本研究では具体的に、LS モデルについて、次の項目でその能力を示すとともに、具体的な問題に適用することでモデル自体の分析も進める。研究申請や開始時には (1)–(3) を予定していたが、期間内に(4)と(5)を新しく進めることができた：

(1) n 本腕バンディット問題
(2) ロボットの強化学習による運動獲得
(3) メタヒューリスティクスへの応用
(4) LS モデル自体の分析と一般化
(5) LS モデルの認知的妥当性の検証

以下、項目毎に具体的な方法を述べる。

(1)　n 本腕バンディット問題において LS モデルを価値関数として用い、その特性を調べた。これは、n 個の行動が可能で、それぞれに報酬を与える確率が割り振られている場合に、「後悔」を最小化する最適化問題である。ここで後悔とは期待損失の一種で、最適な行動、つまり報酬確率が最高の行動を最初から最後まで選び続けた場合の期待獲得報酬と、実際に選択した行動の列の期待獲得報酬との差である。この問題については今世紀に入ってから最適に近いアルゴリズムが Auer らによって提唱されているが、LS モデルの成績の方が優れている場合があり、しか

しその条件や理由は分かっていなかった。この点についての解明を行った。また、元来事象 C の在と不在や、二つの行動 A と B のどちらか、のように 2 つの選択肢の評価に限定されていた LS モデルを任意の n 選択肢の評価に用いるため、トーナメント方式の評価法を考案した。更に、LS モデルの分析に基づき、これを拡張したモデルを提案し、そのパフォーマンスを検討した。

(2) 強化学習によるロボットの運動学習に LS モデルを組み込み、バンディット問題よりも現実的な複雑なタスクでの一般性を検証した。LS モデルは行動と報酬の共起頻度で定義され、そのままでは頻度情報を持たない Q 値などの強化学習のスタンダードな関数との互換性がない。そのため、一般的な Q 学習アルゴリズムの上で、ある行動と、それが greedy (主観的に最適) か否かとの共起頻度を考え、それに LS を適用する LS-Q というアルゴリズムを考案した。タスクとしては、鉄棒にぶらさがり、腰部だけが稼働する体操選手の単純化と言える大車輪運動ロボット (Acrobot) の物理シミュレーションと実機テストを行った。報酬は単純に鉄棒と足先の角度に比例して定義され、鉄棒の真上に足先があるときに最大値が与えられる。状態はロボットの鉄棒との角度と角速度、そして姿勢（腰部の角度）の三次元であり、それぞれ離散化されている。これは二重振り子の制御問題であり、連続的な非線形力学系の一様な離散化であるため、マルコフ性を満たさない部分観測的な難しいタスクである。通常はこの個別のタスクに応じた報酬の与え方の工夫や、状態の分割の工夫を行うが、本研究ではそういったチューニングを一切行わず、LS-Q の学習能力自体を評価した。

(3) LS モデルを、モンテカルロ木探索アルゴリズムに組み込み、探索問題でのメタヒューリスティクスへの応用を行った。モンテカルロ木探索 (MCTS) は、囲碁 AI などの成績向上に目覚ましい役割を果たしたが、基本的には木構造上でバンディット問題同様の最適な資源（試行）配置を行い、最適な行動選択を目するものである。本研究では、ゲーム毎の個別の構造や事情に依存しない一般的な結果を調べるため、MCTS の提案で検証のために用いられた抽象的なゲーム木上の探索について LS モデルを MCTS に適用した場合の効果を検証した。

(4) LS モデルの分析と一般化を行った。LS モデルは「対称性バイアス」と「相互排他性バイアス」を緩く持ち合わせる変形条件付き確率で、因果帰納のデータに合い、また同時に 2 本腕バンディット問題のある問題群で優れたパフォーマンスを見せるものとして、提案された [篠原 2007]。

LS の形式については、[4] は経験ベイズ法の枠組みについての分析を行った他、[2] は [Takahashi 2010] が議論した地の不変性と焦点化・背景化の観点に着目し、それらの性質を保ったままで拡張モデルを提案した。

(5) LS モデルは人間の認知の特性に触発されて提案されたものだが、LS の性質が実際に人間認知のそれとどのくらい一致するのかについては予備的な結果しかなかった。
(5-1) そこで、人間の因果関係に関する直感に実際に適合しているのかを、因果帰納の実験と過去の実験結果のメタアナリシスによって検証した。また、因果帰納（統計データから因果関係の有無・強弱を構築する）に留まらず、帰納された因果直感を実際に意志決定に用いているのかについて実験を行って検証した。
(5-2) また、バンディット問題についても、心理学的な研究がまだ少なく、データも入手できないため、実験を行った。

４．研究成果

(1) n 本腕バンディット問題については、まず最も基本的な 2 本腕バンディット問題をパラメトライズして扱い、従来のアルゴリズムよりも優れた結果 (=非常に小さい後悔) を得た [8]。その上で、LS が満足化を行うことを明らかにし、満足化基準が非明示的に 0.5 となっていることを分析した上で、2 本腕と n 本腕(トーナメント形式)で、過満足問題 (全ての行動に満足できる設定)、一意満足状態 (1 つのみの行動が満足できる設定・満足化＝最適化が成り立つ)、非満足状態 (全ての行動について満足ができない設定) それぞれについての結果を示した [4]。それにより、[8]の結果は、一意満足状態を扱っていたことが分かった。

(2) 大車輪運動の獲得にシンプルな強化学習アルゴリズムを用いるロボット運動学習の課題に関しては、まず物理シミュレータにより、提案した LS-Q アルゴリズムの有効性を示した [5]。この段階で有効性については大きく分けて二つが示された。一つは、LS-Q が、Q 学習に比べて学習の初期から後期にいたるまで優れた結果を見せただけでなく、初期のランダムな探索を徐々に止めていった場合に、停滞ループに陥らず、学習を続けられた点である。ランダム探索は強化学習においては従来必須であるが、その探索の度合いの調整の最適化は非常に難しく、経験的なチューニングによる減衰が行われてきたが、LS-Q はこの点で優れていると言える。また、学習率などの Q 学習に備わるパラメータについて、学習の初期から後期までだけでなく、の分割の細かさに合わせたチューニングが不要であった。さらに、これらが、シミュレータと適合させたロボット実機を作成し、現実に有効であることを示すとともに、LS-Q

が停滞ループを抜け出す仕組みや、初回の回転までに要する時間の分布などを詳細に分析した [1]。

(3) 抽象的なゲーム木上の MCTS による探索に LS を組み込んだ LST アルゴリズムをテストした。その結果、MCTS においても、バンディット問題と同様に、満足化基準との関係で探索優先・最適化・知識利用優先を切り替えることが分かった。また最適化を行う場合には UCT などの既存の技術に比べてスケーラビリティにも優れることが示された [3]。

(4) LS モデルの拡張は、それまで 2 本腕バンディット問題など、2 つの選択肢に限られていた行動価値評価を、任意の n 個の選択肢について行えるようにしたことと、満足化の基準を (報酬関数の調整でなく価値関数の形式の中で) R=0.5 から任意の R∈[0,1]に設定できるようにした点である。さらに、満足化基準 R を決めうちではなくオンラインで更新していく有効な方法も提案した [2]。

(5) LS モデルの認知的妥当性
(5-1) まず、LS モデルが人間の因果関係に関する直感を忠実に記述しているのかをメタ分析で検証した。データセットは Hattori & Oaksford, *Cog. Sci,* 31(5):765-814 (2007)と同じものを用いた。その結果、42 の過去の因果帰納モデルの中で最も優れていると示された DFH モデルと互角の記述性能(より高い決定係数と同等に小さい誤差)が見られた。また、因果帰納の実験をいくつか行い、多様な設定、特にこれまでに試されていない設定でも LS がデータに合うかを検証した。この場合にも DFH と同等の結果を示した。また、LS が、満足化基準と同じ 0.5 の値を持つことが、人間の因果直感にとっては無相関・無関係の値に対応することが、2x2 分割表上での相関係数(四分点相関係数φ)との比較によって示された [4]。
(5-2) n 本腕バンディット問題で人間がどのように振る舞うのかについては、通常ソフトマックス法でのモデリングが行われる。ソフトマックス法は、n 行動それぞれに、その平均報酬に応じた選択確率を割り当てるものである。[6] ではソフトマックスでのモデリングが、実験参加者の平均のモデリングには用いられても、参加者個別の振る舞いのモデリングには不適切であることを示した。ただし、バンディット問題では人間は行動方策を序盤、中盤、それ以降で切り替えているようであり、LS のような価値関数のみでのモデリングには限界があった。この点は今後の課題である。

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、LS モデルの単純化である認知的な満足化価値関数の RS [高橋 投稿中][9] や、稀少と仮定される事象の比率あるいは「双条件付確率」として人間の因果関係に関する観察の直感を表現する pARIs [7] の研究も進展した。そのため、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や、満足化的意志決定や因果推論など、より広い分野で本研究の意義が今後示されていくと考えられる。

＜引用文献＞

[篠原 2007] 篠原 修二, 田口 亮, 桂田 浩一, 新田 恒雄: 因果性に基づく信念形成モデルと N 本腕バンディット問題への適用. 人工知能学会論文誌, 22(1), 58‐68. (2007)

[Takahashi 2010] Takahashi, T., Nakano, M., & Shinohara, S.: Cognitive Symmetry: Illogical but Rational Biases. *Symmetry: Culture and Science*, 21(1), 1–3. (2010)

[高橋 投稿中] 高橋 達二, 甲野 佑: 認知的満足化：限定合理性の強化学習における効用. (修正中)

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）

1. Uragami, D., Kohno, Y., Takahashi, T.: Robotic Action Acquisition with Cognitive Biases in Coarse-grained State Space. *BioSystems*, 145, 41–52. (2016) (査読有)
doi:10.1016/j.biosystems.2016.05.007
2. Kohno, Y., Takahashi, T.: A cognitive satisficing strategy for bandit problems, *International Journal of Parallel, Emergent and Distributed Systems*. (published online on Sep. 2, 2015) (査読有)
doi:10.1080/17445760.2015.1075531
3. Oyo, K., Takahashi, T.: Efficacy of a causal value function in game tree search, *International Journal of Parallel, Emergent and Distributed Systems*. (published online on August 10th, 2015) (査読有)
doi:10.1080/17445760.2015.1064918
4. 大用 庫智, 市野 学, 高橋 達二: 緩い対称性を持つ因果的価値関数の認知的妥当性と N 本腕バンディット問題におけるその有効性, 人工知能学会論文誌, 30(2), 403–416. (2015) (査読有)
doi:10.1527/tjsai.30.403
5. Uragami, D., Takahashi, T., Matsuo, Y.: Cognitively inspired reinforcement learning architecture and its application to giant-swing motion control, *BioSystems*, 116, 1–9. (2014) (査読有)
doi:10.1016/j.biosystems.2013.11.002

（備考：研究代表者は 2,3,4 で連絡著者）

〔学会発表〕（計 4 件）

6. Namiki, N., Oyo, K., Takahashi, T.: How Do Humans Handle the Dilemma of Exploration and Exploitation in Sequential Decision Making?, Proceedings of 8th International Conference on Bio-inspired Information and Communications Technologies (formerly BIONETICS). **ボストン**, **マサチューセッツ**, **アメリカ合衆国**, 2014 **年** 12 **月** 3 **日**. (2015) (regular paper)
doi:10.4108/icst.bict.2014.258045
7. **高橋 達二**, **大用 庫智**: **対称性推論のモデルとしての「双条件付確率」と小数サンプルからの因果帰納推論**, **日本認知科学会第** 31 **回大会発表論文集**, O4-3, 141–148. **名古屋大学** (**愛知県名古屋市**), 2014 **年** 9 **月** 20 **日**. (2014)
8. Oyo, K., Takahashi, T.: A cognitively inspired heuristic for two-armed bandit problems: The loosely symmetric (LS) model, *Procedia Computer Science* (*Proceedings of IES 2013: The 17th Asia Pacific Symposium on Intelligent and Evolutionary Systems*), 24, 194–204. **ソウル**, **韓国**, 2013 **年** 11 **月** 8 **日**. (2013)
doi:10.1016/j.procs.2013.10.043
9. Takahashi, T.: The adaptive combo of human cognitive biases – Satisficing, comparative valuation, and risk attitude, JSAI 2013 (2013 **年度人工知能学会全国大会** (**第** 27 **回**)) **予稿集**, 2J1-2. **富山国際会議場**, **富山県富山市**, 2013 **年** 6 **月** 5 **日**. (2013) (**査読無**)

〔図書〕（計 0 件）

〔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：特になし

６．研究組織
(1)**研究代表者**
高橋　達二（Tatsuji Takahashi）
東京電機大学・理工学部・准教授
研究者番号：00514514

(2)**研究分担者**
なし

(3)**連携研究者**
なし