

アルゴリズム理解とエージェンシー知覚

Algorithm Understanding and Agency Perception

寺田和憲^{*1} 山田誠二^{*2} 伊藤昭^{*1}
 Kazunori Terada Seiji Yamada Akira Ito

^{*1}岐阜大学 Gifu University
^{*2}国立情報学研究所／総合研究大学院大学／東京工業大学
 National Institute of Informatics, SOKENDAI, Tokyo Institute of Technology

Agents are implemented by software programs. This implies that the behavior of agents is described by a finite set of rules. However, human feel agency form the behavior of a software program. In the present study, we conducted an experiment with participants to investigate the cognitive process of higher level abstraction (algorithm understanding) performed in a human-computer collaboration task. The most recently used (MRU) algorithm, known to be one of the simplest adaptive algorithms, and probabilistic MRU algorithm were used to test the human capability to understand an algorithm. The experimental results showed that inductive reasoning with deterministic bias and Markov bias played key role to correctly understand the MRU algorithm. The results also showed that deductive reasoning was used to understand algorithms when participants rely on prior knowledge, and that there was a case in which the algorithm, even known to be the simplest one, was never understood.

1. はじめに

エージェントはソフトウェアプログラムによって実現される。プログラムなので、有限のルールで記述されている。しかし、人はそのようなソフトウェアプログラムからエージェンシー(行為主体性, 人らしさ, 社会性, 意図性)を感じる。1997年5月, 当時のチェスの世界チャンピオンカスパロフはIBMのコンピュータディープブルーとの対戦に敗れた。この対戦は人間の世界チャンピオンが初めてコンピュータに敗れた歴史的対戦として知られている。この対戦の第一局の終盤にディープブルーはある手を打った。その一手は常套的な手ではなく, その一手後にディープブルー自ら投了するほど無意味なものであった。しかしこの手に対して, カスパロフは, その直感に反する手は優れた知性の証であると結論づけた [Silver 12]。ところが, 実際にはこの一手はプログラムがバグによって手を選択することができず, 単にランダムに打った結果生成されたものであったとされている。カスパロフは一般的なチェスプログラムがどのようなアルゴリズムで動作しているかを熟知していた。そして, 実際にディープブルーはその一手以外はコンピュータらしい手を打っていた。アルゴリズムに従った戦術を使っている限りはディープブルーの振舞いはカスパロフの想定範囲内だったのであろう。しかし, 例えランダムな手であったとしても, その一手が想定外のものであったために, カスパロフはコンピュータに対してエージェンシーのようなものを感じたのではないかと思われる。

我々はこれまでに, システムが定型性を逸脱することが, エージェンシーを知覚させる鍵になることを論考と実験によって示してきた [寺田 11, 寺田 12]。しかし, アルゴリズム理解がエージェンシー知覚へとどのように発展するかについては明らかになっていない。そこで, 本研究ではまず, エージェントの振舞いがアルゴリズムとしてどのように理解されるかを明らかにすることを目標として, 最もシンプルなインタラクティブかつ適応的アルゴリズムであるMRU (Most Recently Used) アルゴリズムがどのように理解されるかを調べる実験を行った。

連絡先: 寺田和憲, 岐阜大学工学部電気電子・情報工学科, 〒501-1193 岐阜市柳戸 1-1, 058-293-2792, terada@info.gifu-u.ac.jp

本研究では, 適応的な振舞いが有効に働く協調タスクである協調記号合わせゲームをタスクとして用いる。このゲームは二人のプレーヤによって行われる繰り返しゲームである。プレーヤは各ラウンドにおいて, 記号セットの中から種類の記号を選択し, 同時に見せ合う。記号セットは例えば, ♠, ♠, ♠, ♠である。記号が一致していれば両方のプレーヤに得点が与えられ, 一致しなければ得点は与えられない。本研究においては片方のプレーヤはコンピュータである。

観測した時系列からコンピュータのアルゴリズムを同定することは不良設定逆問題である。なぜなら, 人はコンピュータの現在の手がどれぐらいの長さの履歴に依存しているかを知るすべがなく, また, 異なるアルゴリズムが同じような履歴を生成する可能性があるからである。従って, 我々は, 人がシステムのアルゴリズムを同定するためにシステムの振舞いを探索する際に, 何らかのバイアスを用いて探索範囲を制限するのではないかと考える。

2. 実験方法

我々はMRUアルゴリズムがどのように理解されるかを調べるために参加者実験を行った (詳細は [Terada 13] を参照のこと)。タスクとして50ラウンドの繰り返し協調記号合わせゲームを用いた。

MRU (Most Recently Used) アルゴリズム ([Findlater 04]) は最も単純な適応アルゴリズムとして知られる。これはアダプティブメニューの「よく使われるメニュー」欄やかな漢字変換の候補の提示において, 最も最近使われたものを上位に表示するアルゴリズムである。類似のアルゴリズムとしてMFU (Most Frequently Used) アルゴリズムがある。これは, 過去の使用頻度の高いものを上位に表示するアルゴリズムである。

実験はWebブラウザ上で行われた。実験参加者は各ラウンドで10秒以内にハート, クラブ, ダイヤモンドのいずれかのマークをクリックして自分の記号を決定するように求められた。画面上には双方の過去5ラウンドの選択が表示された。

参加者に対して, 記号が一致していればコンピュータ, 参加者双方の得点となり, 一致しなければ双方ともに得点を得られない, 最終的な獲得ポイントが高くなるように記号を選択する

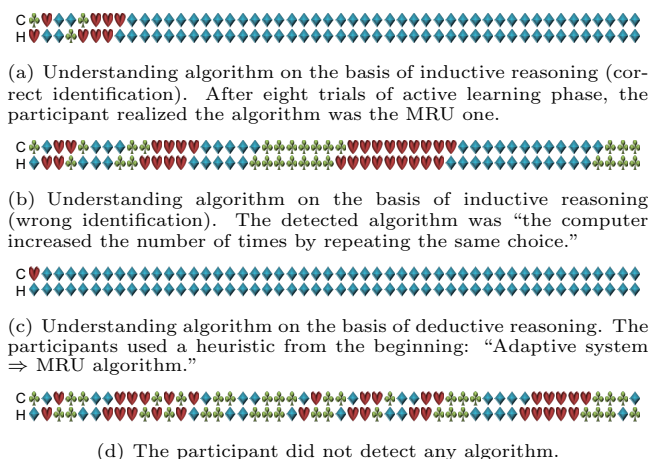


図 1: Examples of typical behavioral pattern in the D condition. C: computer, H: human.

ことを指示した。

実験には 25 人が参加した。参加者には、この実験の目的はオンラインゲームシステムのユーザビリティを調査する目的であると伝えられた。また、コンピュータは協調的に振舞うと伝えられた。また、得点に応じた価格帯のコンピュータ周辺機器が得られると伝えられた。

3. 実験結果と考察

実験結果を分析したところ、MRU アルゴリズム理解には帰納的、演繹的の 2 種類の方法が用いられたことが分かった。

76%の参加者 (19 人) は帰納的にアルゴリズムを同定した。そのうち 15 人は 10 ラウンド以内に最適戦略すなわち同一手を連続する方法を発見した。この 15 人の参加者のうち 13 人がコンピュータのアルゴリズムを MRU であると答え、一人が戦略なし、一人が同一記号戦略であると答えた。このような参加者の典型的な最初の 10 手を図 1(a) に示す。このタイプの参加者はゲームの最初で統計モデルを構築しそのモデルに従ってコンピュータの戦略を同定したと考えられる。しかしながら、理論的には、相手の戦略を統計に基づいて同定するために無限回の試行を必要とするために、10 ラウンド以内で相手の戦略を同定するためには何らかのバイアスが用いられたと考えられる。ここで用いられたバイアスは決定論バイアスとマルコフ性バイアスであると考えられる。もし、アルゴリズムが決定論的であると仮定するならば、参加者は最低 3 回の試行で MRU アルゴリズムを同定することができる。また、マルコフ性バイアスがなければ、参加者は 1 ラウンド前だけでなく、それより以前の履歴も考慮に入れなければならない、アルゴリズム同定のためにさらなる時間が必要となる。

このゲームに対しては MRU アルゴリズム以外のアルゴリズムの帰属も可能である。帰納的にアルゴリズムを同定した参加者のうち 4 人は MRU 以外のアルゴリズムであると理解していた。例えばある参加者は「コンピュータは単に同じ記号を出し続けていた」と答えた。また別の参加者は「コンピュータは二つの記号を交互に出していた」と答えた。さらに「 $\heartsuit\spadesuit\heartsuit\heartsuit$ のように同一記号を連続する回数を段階的に増やしていった」という複雑なアルゴリズムを読みとって答えた参加者もいた (図 1(b))。

16%(4 人) の参加者は演繹的にアルゴリズムを同定した。彼

らは選択する記号を第一ラウンドで固定し、50 ラウンド全て同じ記号を選択した (具体例は図 1(c) 参照)。しかし、それらの参加者全てがコンピュータのアルゴリズムを MRU であると記述した。一切の探索をせずにこのようなアルゴリズム理解がなされる理由は彼らが **トップダウン適応バイアス** を持っていたからにほかならない。すなわち、教示によってコンピュータが協調的に振舞うと知らされていたために、「協調的に振舞うコンピュータは人の選択に従順に対応するものだ」という論理的な推論を行い、それを探索によって確かめることなく最後まで信じてゲームを行ったと思われる。実際にコンピュータは彼/彼女らの仮説を覆す選択を行わなかったため、彼/彼女らが自信の信念に疑念を挟む余地はなかったものと思われる。

8%(2 人) の参加者は、50 ラウンドを経てもいかなるアルゴリズムも理解できなかった (図 1(d))。

4. おわりに

意図とは心的に表象される目的である。人を含む生物の短期的、長期的目的は自己複製と自己保存という根源的な目的のサブゴールと考えることができる。そして、知能はそれらの目的を達成するために獲得された汎用的な手段である。チェスの目的は相手のキングをチェックメイトすることであり、プレーヤに対して明示的に与えられている。本研究における実験でもプレーヤ双方には「協調して高得点を得る」という目的が共有されていた。すなわち、目的が既知な状況での実験であった。アルゴリズムを同定することはそれだけでは振舞いを制御している法則を知ることには貢献せず、目的そのものの同定には直接寄与しない。

通常の人同士のインタラクションの中では相手の意図もアルゴリズムも未知な場合が多々ある。しかし、人は瞬時に意図もアルゴリズムも同定し、うまくその状況を乗り切っている。アルゴリズム理解から意図理解へと至るために何らかの飛躍や抽象化が行われるのか。それとも、そもそもアルゴリズム理解と意図理解に因果関係が存在しないのか。それらの同定プロセスをモデル化することが今後の我々の課題である。

参考文献

- [Findlater 04] Findlater, L. and McGrenere, J.: A comparison of static, adaptive, and adaptable menus, in *Proceedings of the 11th international conference on Intelligent user interfaces*, pp. 89–96 (2004)
- [Silver 12] Silver, N.: *The Signal and the Noise: Why So Many Predictions Fail-but Some Don't*, The penguin press (2012)
- [Terada 13] Terada, K., Yamada, S., and Ito, A.: An Experimental Investigation of Adaptive Algorithm Understanding, in *CogSci 2013: Proceedings of the 35th annual meeting of the cognitive science society* (2013)
- [寺田 11] 寺田 和憲, 伊藤 昭: 人間はロボットに騙されるか?—ロボットの意外な振舞いは意図帰属の原因となる—, 日本ロボット学会誌, Vol. 29, No. 5, pp. 43–52 (2011)
- [寺田 12] 寺田 和憲, 山田 誠二, 伊藤 昭: ボーナス付きマッチングペニーゲームにおける人間からエージェントへの適応プロセスの解明, 人工知能学会論文誌, Vol. 27, No. 2, pp. 73–81 (2012)