2M1-1

知能の軍拡競争:報酬の遅延によるエージェント同士の戦略の複雑化

Process of Enlarging Complexity of Agent using Delayed Reward Game

大澤 博隆*1
Hirotaka OSAWA

*1 筑波大学
University of Tsukuba

Social brain theory hypothesizes that the human brain becomes larger through evolution mainly because of reading others' intentions in society. Reading opponents' intentions and cooperating with them or outsmarting them results in an intelligence arms race. The author simulates the evolution of intelligence arms race during long-term using finite state automatons.

1. はじめに

他者の意図推定を行う心の理論は、人間の持つ最も複雑な知能プロセスの一つである [Hyatt 10]。社会脳仮説では、人間がお互いにコミュニケーションを取ることでこのような意図推定プロセスを身につけたのではないかと推定している [Bryne 89]。被食者と捕食者間の軍拡競争と同じように、お互いに意図を読み合うことを要請される環境では、知能に対する軍拡競争が発生することが期待できる [Flinn 05]。あるエージェントが他のエージェントより賢ければ、他のエージェントを欺き多くの利益を手に入れられる。このようなテーマはマルチエージェントシステムと人工知能にとって共に興味深い課題である。他者の意図を読み解くシステムを解明することで、ユーザの意図を読んだ人工知能システムを作ることが可能になるだろう。

筆者はこれまで、Anti-Max Prisonner's Dillemma game (AMPD)と呼ばれる繰り返し囚人のジレンマゲームのバリエーションを用いて、このような知能が発生するプロセスを確認してきた。AMPD を用いた人間間のシミュレーションでは、上位の戦略が人間によって複雑化するプロセスを発見できた [Osawa 13]。また、遺伝的プログラミングによって、100 サンプル、1500 世代にわたるコンピュータシミュレーションを行った結果、このような知能の複雑化が、普通の繰り返し囚人のジレンマゲームの場合には発生せず、AMPD の場合のみに発生することを突き止めた [Osawa 14]。

本研究では、AMPD を用いたシミュレーションによって、この複雑化が長期的にどういった過程をたどるか、10000 世代にわたるシミュレーション結果から検討する。

2. ゲームのモデル

本研究で行われる交代取引ゲームでは、Robert Axelrod の行った繰り返し囚人のジレンマゲーム[Axelrod 84]を Angeline が改良した、Anti-Max Prisoner's Dilemma game (AMPD)を使用する [Angeline 94]。AMPD では、相互協調でなく交互の手で利益が最大化される

戦略型ゲームにおける一般的な利得表は表 1 の形となる。wait と take は、それぞれ Axelrod のゲームにおける coop と betray に対応する。本利得表で繰り返し囚人のジレンマが発生する条件は、式 1 のとおりである。この場合には、両者が協調を繰り返すほうが、交互に裏切り返すよりも利得が大きい。両戦略が同じ利得を得る場合は式 2 であり、これは Multi-Max Prisoner's Dilemma game (MMPD)と呼ばれる。また、協調よりも交互の取引が有効である。AMPD ゲームが発生しうる条件は、式 3 のとおりである。

表 1 : 交代取引ゲームの一般的利得表

B \ A	Cooperate	Defect
Cooperate	$(A:c,B:c)$	$(A:a,B:b)$
Defect	$(A:b,B:a)$	$(A:d,B:d)$

$$a>c>d>b,\quad a+b<2c \quad (1)$$

$$a>c>d>b,\quad a+b=2c \quad (2)$$

$$a>c>d>b,\quad a+b>2c \quad (3)$$

AMPD 条件では交互に Defect を繰り返すことで両者の利益が最大化される。問題は、こうした相互裏切りが簡単に発生しないことである。交互裏切りが達成されるためには、相手と違う手を出し、攻撃を行う必要があるが、攻撃を行いすぎて、相手のエージェントに自分が搾取を行おうとしている、と誤認されてしまってはいけない。また、相手のエージェントに自分が協力的すぎて、搾取するに値する対象である、と思われてもいけない。

本研究では、AMPD 条件として、$a=7$、$b=-3$、$c=1$、$d=-1$という条件のゲームを採用した[Osawa 14]。

3. シミュレーション条件

各エージェントの戦略は有限状態オートマトンで記述される。各オートマトンは数字の状態を持っており、偶数への推移が協力戦略、奇数への移動が裏切り戦略を意味する。代表的な戦略を図 1 に示す。

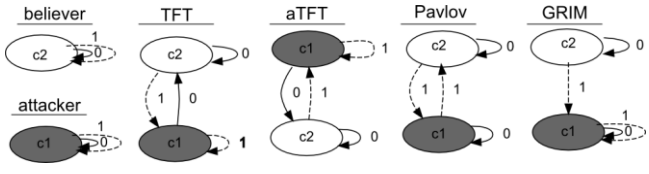

図 1 オートマトンによる戦略奇術の例

戦略の推移は 3 つの数字の組で記述される。最初の数字が現在状態、次の数字が相手の手(0 が協調、1 が裏切り)を意味し、3 番目の数字が次に推移する状態を意味する。例えば{{2}, {2,0,2}, {2,1,2}}は常に協調を行う戦略を意味し、{{1}, {1, 0, 1}, {1,1,1}} は常に裏切りを行う戦略を意味する。{{2}, {2,0,2}, {2,1,1}, {1,0,2}, {1,1,1}}は IPD においてよく見られるしっぺ返し戦略(TFT)を意味する。aTFT は TFT の開始時が裏切りの TFT であり、Pavlov は相手が裏切りを返した場合に協調と裏切

連絡先:大澤博隆, システム情報系知能機能工学域, 〒305-8573 茨城県つくば市天王台 1-1-1 , osawa@iit.tsukuba.ac.jp

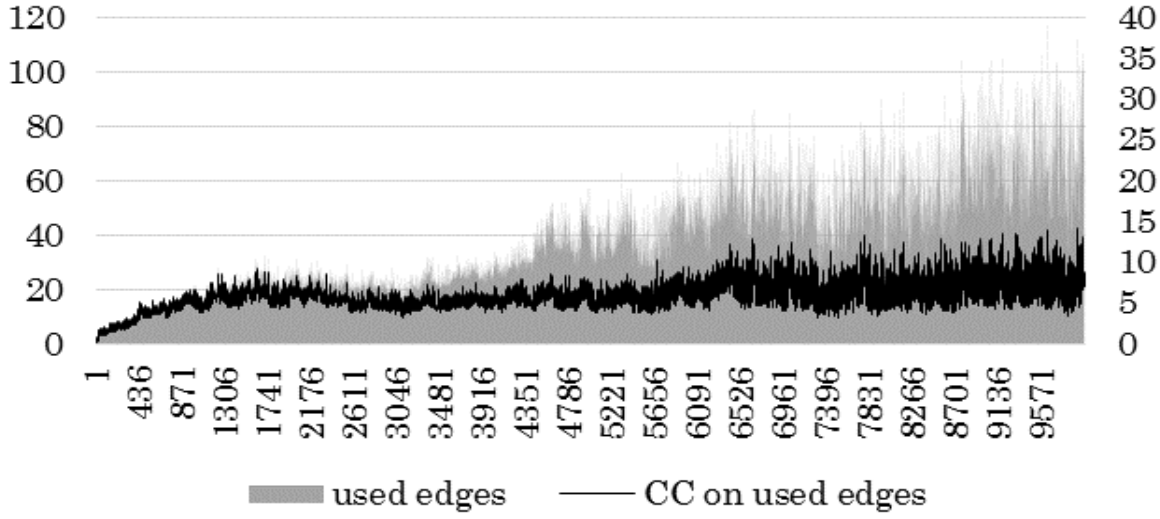

図 2 10000 世代にわたるシミュレーション結果。グレー領域および Y 軸左の数字が使用された枝の平均数を表し、黒線及び Y 軸右の数字が使用された枝の循環的複雑度(CC)を表す。

りを反転させるエージェントである。GRIM 戦略は、一度裏切られるまで強調し、一度裏切られた場合に全て裏切りで返す戦略である。

オートマトンの戦略は遺伝的プログラミング(GP)によって世代ごとに進化を行う。今回のシミュレーションでは、50 エージェントが 10000 世代にわたって進化を行った。それぞれのエージェントは各世代ごとに、他 49 エージェントに対する総当り戦を行う。エージェントの成績は総当り戦の結果得られた総スコア数で決定する。各ラウンドの終了時、最もランクの低いエージェントが処刑される。つぎに、47 位~49 位の 3 体のエージェントが突然変異を起こす。最後に、1 位と 2 位のエージェントが交差によって新しいエージェントを生み出し、これが足される。

突然変異のプロセスは以下の 3 つに分かれている。10%の確率で、オートマトンのノードがひとつ選択され、その状態が反転する。80%の確率で、オートマトンの枝が一つ選択され、その枝の指し示す先が他のノードになる。10%の確率で、新しいノードがエージェントに追加される。新しく付け加えられたノードは、次の枝の突然変異によってオートマトンに繋がれる可能性がある。

交差のプロセスでは、2 つのエージェントのうちどちらか一つが選ばれ、そのエージェントのオートマトンのツリーの一部が、もう一つのエージェントのオートマトンのツリーの一部に置き換えられる。置き換えられるポイントはランダムに選択される。置き換えられたノードのツリーのナンバーが元のナンバーと重複する場合には、重複しない形にノードが置き換えられる。もし孤立したノードや枝が突然変異や交差によって発生した場合、これらは将来的な突然変異で繋がれる可能性を残し、保存される。

初期のオートマトンの多様性は、GP でより良い結果を導くために重要である。本研究ではゲーム理論でよく選ばれるシンプルな戦略を含んだ戦略群を初期 50 体のエージェントとしてランダムに配置した。それぞれのエージェントは 1 または 2 のノードをもち、これらの開始ノードと枝はランダムに配置される。損結果として、32 体のランダムなエージェントが生成され、これらが初期 50 体としてランダムに配置される。

評価方法として、使用されたエージェントの平均スコア、戦略の枝数、循環的複雑度(cyclomatic complexity) [McCabe 76]の3点を比較した。樹冠的複雑度(CC)は、大まかには各戦略の分岐の複雑さを表すパラメータである。本研究で使用するオートマトンには終了状態が存在しないため、循環的複雑度は、使用されたユニークな枝数から、使用されたノード数を引く、という形で計算した。例えば図 1 の believer や attacker の CC は 0、TFT や aTFT、Pavlov の CC は 2、GRIM の CC は 1 となる。また、遺伝的プログラミングの手法では、実際に使用されない戦略の枝数も大きく増加していってしまう。そのため、循環的複雑度を数えるにあたっては、そのラウンドの総当り戦で一度以上使われたノード、枝のみを使用して計算した。そのラウンドで使用されていない枝であっても、将来的な戦略に対する潜在的な頑健さ、対応力を示している可能性はある。しかしながら今回は、使用された枝のみを数えることで、その試合で確実に使用された枝のみを計算している。

4. シミュレーション結果

事前のテストとして、世代数を 1500 世代とした AMPD 条件のシミュレーションを 100 条件行った。そのうち 43 条件の場合にエージェントのノード数が成長することを確認した。これらの条件のうち、1500 世代目の平均枝数が中央値となった条件を元に 10000 世代にわたる長期シミュレーションを行った。

最初の 500 世代で平均スコアは 2 に近づいた。これは、AMPD 条件における最も高いスコアが取れる相互裏切りの条件に該当する。

図 2 にシミュレーションの結果を示す。図2のグレーの領域は、使用された枝の平均数、図 2 の黒い線が使用された枝の循環的複雑度の平均数を示す。最初の 1500 世代間では使用された枝と使用された枝の CC がともに増加していくことが見て取れる。本結果は筆者が以前行ったシミュレーション結果と同様である[5]。また、1500~10000 世代の間、使用された枝の平均数が増えていくことがわかる。一方で、循環的複雑度は 1500 世代まで順調に増加するものの、それ以降の増加はより緩やかになる。

5. 考察

AMPD ゲームにおいて、エージェント同士が同じ戦略を持っている場合、そのエージェント同士は相互裏切りの状態を保つことができない。従って、AMPD ゲームは各エージェントがお互

いにユニークな戦略を持つように働きかけることを示唆する。人間同士のシミュレーションによる、著者らの以前の実験結果では、各人間が戦略として、最初にお互いにユニークな信号を送り合う初期フェーズが存在したことがわかっている [大澤 13]。50 体のエージェントがユニークな信号を持つためには、最低で 6bit のユニークな分岐が必要である。これより、循環的複雑度は 6 以上が望ましいことになる。本研究で得られた使用された循環的複雑度の平均値は、この条件を満たしているといえる。

また、図 2 における、使用された枝の循環的複雑度の平均値と標準偏差を、1000 世代ごとに取ったものを図 3 に示す。

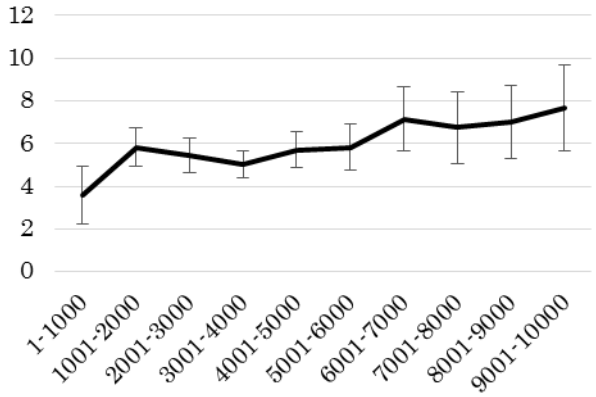

図 3 1000 世代ごとの循環的複雑度の平均と標準偏差

図 3 の値より、使用された枝の数は世代を経る毎に緩やかに増加していることがわかる。AMPD 条件で緩やかな枝の増加が使用された枝に限っても起こっているということは、エージェントの知能を示す戦略の増加が、単純に遺伝的プログラミングによって必要のない分が生み出されるだけではなく、必要のある形で使われていることがわかる。これは、AMPD 条件が長期にわたって意味のある知能の増加を促す、ということを示唆する。

また、標準偏差の値が世代を経る毎に増加していることも見て取れる。これは、各ラウンドにおける、使用された循環的複雑度の平均値のぶれが大きくなっていくことを意味している。実際に図 2 のように、使用された循環的複雑度は世代を経る毎に増減が激しくなっており、4~12 までの値に合わせて変化する。使用されているエージェントは同一の物が多く、同一のエージェントが世代の各ラウンドで、たくさんのノードを使用したり、少数のノードを使用したり、というように対応を変化させていることがわかる。ここから、エージェント集団が複雑な分岐を要するような条件と、単純な分岐で済む条件の間を交互に移動していることが推測できる。

また、図 3 に表示されているのはあくまで使用された循環的複雑度である。ここから推測できることとして、エージェントの持つ全体的な戦略オートマトンが、複雑な分岐を要する条件と、単純な分岐で済む条件のどちらにも対応した、頑健な戦略を持ち合わせているのではないか、ということが推測できる

6. 結論

本研究では有限状態オートマトンを使った遺伝的プログラミングの手法で、軍拡競争を後押しする利得条件を調べた。特に本研究ではこれまでの著者の結果を受け継ぎ、長期にわたって、どのように戦略が複雑化するかを調べた。結果、報酬が遅延される利得表において、エージェントの内部状態の複雑化・軍拡競争が発生すること、こうした内部状態の増加が長期にわたって継続することがわかった。

7. 謝辞

本研究は独立行政法人科学技術振興機構の戦略的創造研究推進事業（さきがけ）領域「情報環境と人」の援助を受け行われました。

参考文献

[Hiatt 10] L. M. Hiatt and J. G. Trafton, “A Cognitive Model of Theory of Mind,” in *International Conference on Cognitive Modeling*, 2010, pp. 91–96.

[Bryne 89] R. W. Byrne and A. Whiten, *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans*. Oxford University Press, USA, 1989.

[Flinn 05] M. V. Flinn, D. C. Geary, and C. V. Ward, “Ecological dominance, social competition, and coalitionary arms races,” *Evol. Hum. Behav.*, vol. 26, no. 1, pp. 10–46, Jan. 2005.

[Osawa 13] H. Osawa and M. Imai, “Evolution of Mutual Trust Protocol in Human-based Multi-Agent Simulation,” in *12th European Conference on Artificial Life*, 2013, pp. 692–697.

[Osawa 14] H. Osawa, “Intelligence Arms Race: Delayed Reward Increases Complexity of Agent Strategies,” in *International Conference on Autonomous Agents*, 2014, p. (accepted).

[Axelrod 84] R. Axelrod, *The Evolution of Cooperation*. Basic Books, 1984.

[Angeline 94] P. J. Angeline, “An Alternate Interpretation of the Iterated Prisoner’s Dilemma and the Evolution of Non-Mutual Cooperation,” in *Proceedings of 4th artificial life conference*, 1994, pp. 353–358.

[McCabe 76] T. J. McCabe, “A Complexity Measure,” *IEEE Trans. Softw. Eng.*, vol. SE-2, no. 4, pp. 308–320, 1976.

[大澤 13] 大澤博隆 and 今井倫太, “交代取引ゲームにおける他者識別規則の進化,” in 人工知能学会全国大会, 2013, pp. 4H1–3.