4H1-3

交代取引ゲームにおける他者識別規則の進化

Evolution of Identification in Mutual Trading Game

大澤 博隆[*1]
Hirotaka OSAWA

今井 倫太[*2]
Michita IMAI

[*1] 筑波大学
University of Tsukuba

[*2] 慶應義塾大学
Keio University

Acquisition of the opponent's model and achieving mutual trust with others are notable traits of humankind's intelligence. Achieving mutual trust is a big challenge for artificial intelligences, and it is a key factor in trading. However, how players observe each others' behaviors and how they achieve mutual trust are not fully known. In this study, we researched the growth of a mutual trust protocol in a trading game in a human-based simulation. We designed and implemented web-based multi-player trading game based on the refusable iterative Anti-Max Prisoner's Dilemma game (rAMPD). In the game, each agent's strategy is described by a automaton and periodically modified by human players. We conducted a long-term human-based multi-agent simulation using this trading game for approximately one month and observed how the agents' automata changed. Analyses of the high-ranking agents' automata and introspective reports by the human players revealed that the mutual trust protocol is achieved by using the initial trade as a signal for mutual recognition.

1. はじめに

交渉ゲームのマルチエージェントシミュレーションにおける重要な課題の一つは、各エージェントが相手の意図推定を行い、信頼出来る相手を見つけ出さなければ解決しない種類の課題である。意図推定は社会を生きる人間が得意とする能力の一つであり、社会脳仮説の観点からは、脳の発達の主要因であると考えられており、知的なエージェントが必要となる[Byrne & Whiten, 1989]。

特に、相互協調ではなく、各エージェントが交互に大きな得と小さな損を立場を変え繰り返すことで、利益が最大化される繰り返し交渉ゲームにおいては、交渉相手が信頼出来る相手であるかどうか、交渉拒絶の可能性を含め、各エージェントの意図推定能力が大きく必要となってくる。Fisher と Shapiro は実際の交渉教育の際に、腕相撲ゲームを用いて、時間的な間隔をおいた利益交換が有効であることを示している[Fisher & Shapiro, 2005]。勝者が得点を得る腕相撲ゲームでは、両者が全力を上げてリソースを消費するのではなく、交互に腕を倒し合うことで全体の利益が最大化する。問題は、自分が手を倒した時、次に相手が手を倒すエージェントであるかどうか判断することである。

交渉ゲームにおける典型的モデルである Iterative prisoner's dilemma(IPD) game は、相互協調を行なうことで集団の利益が最大化するように設計されたゲームであり、このように時間差のある交渉をモデル化しない[Axelrod, 1984]。IPD は各エージェントの戦略が遺伝的に埋め込まれた生物における集団のシミュレーションには有効である[Le & Boyd, 2007]。一方で、他者の意図推定を行なう、人間のような知的なエージェントの交渉ゲームの集団シミュレーションにはあまり役立たない。相互協調でなく交互の手で利益が最大化される Iterative Prisoner's Dillema Anti-Max Prisoner's Dilemma (rAMPD)も Angeline らによって定式化されているが、予め決められたパターンから交互の戦略を選択するのみであり、複雑な意図の推定や、意図推定に依る交渉拒絶などがどのように創発してくるかを分析できていない[Angeline, 1994]。

本研究ではエージェントシミュレーションの代わりに、実際に数十人の人間同士で一ヶ月交代取引ゲームを繰り返し行い、どのようにして戦略が発達するか、相手の認証戦略が生まれるか、などといった戦略の検討を行った。

論文の構成は以下のとおりである。2 章では、本研究で行う交代取引ゲームをモデル化する。3 章ではでは実際に行った対人型のゲームの設計と実施結果について述べ、4 章で結果を考察する。5 章で、結論について述べる。

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

2. 交代取引ゲームのモデル

本研究で行われる交代取引ゲームは、Robert Axelrod の行った繰り返し囚人のジレンマゲームを基盤として考える[Axelrod, 1984]。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。本利得表で囚人のジレンマが発生する条件は、式 1 のとおりである。また、協調よりも交互の取引が有効である、交代取引ゲームが発生しうる条件は、式 2 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	$(A:b, B:a)$	$(A:d, B:d)$

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

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

また我々は、それぞれのエージェントの交渉拒絶の選択肢を付け加えた。もしどちらかのエージェントが交渉拒絶を選んだ場合、交渉は中断され、二度と再開されない。我々は Axelrod の使用した報酬テーブル (a = 5, b = 0, c = 3, d = 1)を編集元として使用した。なぜなら、これがもっとも著名な報酬テーブルだからである。交渉拒絶の価値を高めるため、我々は4つの値の平均を 0 に近づけた。そのため、a,b,d から 2 を引き、c から 3 を引いた。本研究で使用した rAMPD のテーブルは(a = 3, b = -2, c

= 0, d = -1)となる。4 つの定数の平均は 0 であり、かつ式 2 の条件を満たす。c の値が 0 であることは、Fisher と Shapiro の腕相撲ゲームにおいて、お互いが協力的である事、つまりお互い力を抜く行為が利益を生み出さない、という事象に対応する。

この rAMPD では、各エージェントが総当りで交渉を行う。我々は交渉のタイミングを 100 回と制限した。総当り戦はある期間を置いて行われ、各エージェントの戦略を記述する人間参加者が、期間の間にエージェントの戦略を改良することができる。

2.1 関連研究

相手が信頼出来るかどうかをエージェントが判断するシミュレーションとして、事前のメッセージ交換によるコミュニケーションを分析した cheap talk game というモデル化[Crawford & Sobel, 1982]や、エージェントの局在性を用いたエージェント同士のクラスタリング、という研究が行われている[Wärneryd, 1991][Nowak & May, 1992]。しかしながら、一般的な自由取引では局所性に依存せず、両者の同意があって初めて交渉が成立し、どちらか片方が拒絶を行えば、それ以降の交渉は不成立となりうる。人間はこれらの区別がなくとも相手の意図を推定し、交渉の refusal が可能である。このように、従来のモデルでは意図を持った集団の交渉生成という課題のシミュレーションが難しかった。

3. 複数ユーザーを用いた交代取引ゲームの実行

我々は交渉拒絶条件が他者識別を促し、知能の軍拡競争をもたらすのではないか、と仮説を立てた。そこで我々は、本仮説を検証するため、GA のシミュレーションの代わりに、人間同士にオートマトンの戦略を継続的に進化させ、どのような戦略の発達が行われるかを観察することにした。筆者らの過去の研究では、有限状態オートマトンに関する授業課題として、囚人のジレンマゲームを元にした検討を行なっている[大澤 & 今井, 2007]。本研究ではこのゲーム実装を元に、利得表を 2 章で記述した形に変更した。

交代取引ゲームは web ページのゲームという形で設計した。各参加者は web のフォーム上から自身の戦略を記述し、更新し、前回の試合結果を観察することが可能である。ゲーム全体は AJAX プログラムとして実装されており、web ページ上のインタフェースは JavaScript で記述し、オートマトンの計算と試合のプログラムは、サーバサイドで、Perl で記述した。また、本ゲームの参加者に動機を与えるためのカバーストーリーを作成した。カバーストーリーの設計に際しては、直感的に理解がしやすくなることを考え、食べ物による利得表の説明を行った。

それぞれの参加者は各エージェントの戦略を有限状態オートマトンで記述した。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。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}, {1,1,0}} は常に裏切りを行い、相手に裏切られたら交渉を中止する搾取戦略を意味する。{{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。

本実験は有限状態オートマトンに関する授業の授業参加者の一部が任意参加し、授業の得点を追加するという形をとった。結果として、100 人以上の授業参加者の一部が参加した。応募開始から 2 週間後に人数を締め切った。最終的に、74 人がゲームに参加した。期間は 2012/5/29 12:30 から 2012/6/26 8:30 までの 29 日間とした。更新のタイミングは 1 日 4 回、8:30, 12:30, 16:30, 20:30 とした。これは授業時間に重ならない時刻であり、各参加者が更新前後に結果を確認し、自分のエージェントの戦略改善を円滑に行うことを期待できる。合計更新数は 28 * 4 の 112 回となる。

各参加者は前回の更新時間に行われた自分と他エージェントとの試合結果を呼び出し、その結果を参照しながら、自分のエージェントのオートマトン戦略を改善できる。他のエージェント同士の試合結果を知ることはできない。これは、エージェントシミュレーションにおける各エージェントの立場を模している。6/26 8:30 に行われる最終試合結果の成績上位者に授業得点が追加されることが、応募前に発表された。1 点以上を取った授業参加者を 16 のグループに均等に分け、上位グループから順に 20 点から 5 点までの点数を追加した。応募開始時には具体的な得点基準が発表されておらず、具体的な得点法は募集締め切りと同時に発表された。なお、本実験はオートマトンに関する授業課題の復習を兼ねている。そのため、実験に関する倫理的な問題は発生しないと考えられる。

3.1 実施結果

各参加者の得点と分布を図 1 に示す。図 1 より、上位のオートマトンに置いて状態数が爆発的に増えたことがわかる。74 人中、上位 68 人(A01-A68 まで)が 1 点以上の利得を得た。これらの参加者に授業追加点が与えられた。

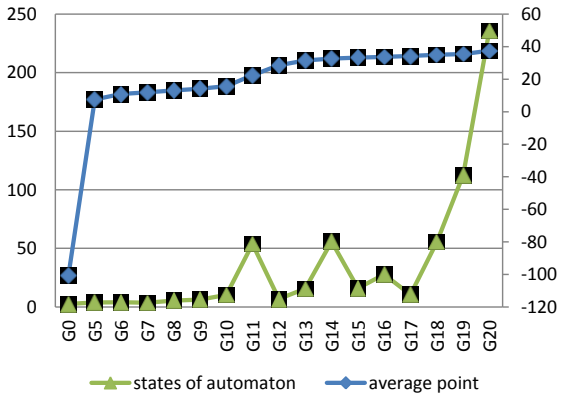

図 1 オートマトンの平均状態数と平均得点。左側の軸の数字がオートマトンの状態数を表し、右側の軸の数字が平均得点を表す。下軸の数字は 16 個の特典グループと平均が 0 以下のグループ、合わせて 17 個のグループを表す。

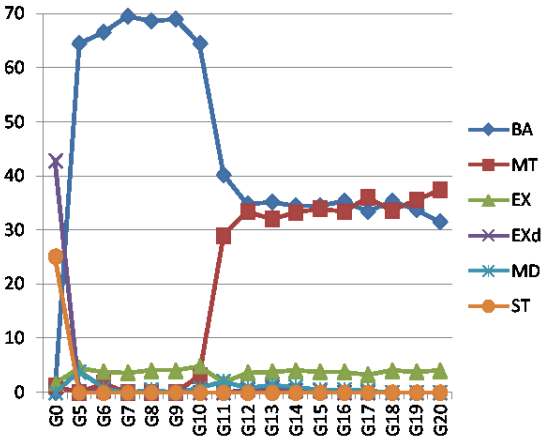

図 2 各グループにおける戦略の分類

各エージェントの交渉について、以下のように分類した。お互いの点数が 50±10 点を得た場合、相互連携(MC/Mutual Cooperation)ができたとみなし、片方が 50 以上、片方が-50 以下の場合に搾取(EX/Exploit)が起きたとみなす。また、両者が-50 以上の場合に、相互破壊(MD/ Mutual Destruction)がおきたとみなす。両者が±15 以下の場合には停滞(ST/ Stagnation)がおきたと分類した。ST は両者がずっと wait を選んだ場合、どちらかが早い段階で拒絶を選んだ、などの場合が考えられる。各戦略の分類を図 2 に示す。

4. 考察

全エージェントの状態の平均数は 33.7、総合得点が 1 以上であったエージェントの状態の平均数は 36.5 であった。いくつかのエージェントは 100 から 200 前後の状態数を保持しているが、これは最大 100 回区間の有限回繰り返しの囚人のジレンマゲームであったため、相互連携状態に移行したエージェントのいくつかが、100 回目に take や拒絶を選択したためであると考えられる。実際に、100 以上の状態を持つエージェントからは最終状態チェックのアルゴリズムが確認された。したがって、実際の有効状態数はもっと少ないと考えられる(例えば A03 のようなエージェントは、12 個の状態数しか保持していない)。また、A02 のエージェントは 600 以上の状態数を持っているが、これはオートマトンのプログラムを自動生成したため、と考えられる。ただし、付録 B の表を見る限り、100 回目に相互連携を止め、take や拒絶を選択する、という行動は、相互連携から得られる利得よりも少ないため、実際には点数順に影響を与えなかったと考えられる。点数順は、相互連携できたエージェントの数によって決定された。上位をとったエージェントはオートマトンの数に現れないものの、相手を見分け、複数の戦略を切り替えるという振る舞いを行っていた。

上位のエージェントの戦略の概略を以下に示す。まず、自身が Take のとき相手が Wait を取り、自身が Wait のとき相手が Take を取る場合、相互連携を達成した。一方で、相手が Take を狙ってくる攻撃的エージェントである場合には、交渉を拒絶する。また、相手が Wait しか行わず、Take に対して Take を返してこない愚者エージェントの場合、相互連携をやめ搾取を行う。相手がこの3種類のどれに当てはまるか、初期の交渉結果から見分け、この3通りの戦略を適切に切り替えられたエージェントが得点を得た。また、相互連携可能な相手を攻撃的エージェントと見間違える、愚者エージェントを相互連携可能者や攻撃者などと見間違えた場合に、得点は下がる傾向にあったことが、各試合結果及び参加者のアンケートより判明した。

エージェントの戦略は、初手が wait か take か、という大まかなに2種に分かれ、内訳は初手 wait のエージェントが 17 体、初手 take のエージェントが 57 体であった。初手 take が多いのは、利得の確定が後手に回るのを嫌ってのことだと考えられる。一方で、初手 wait の場合は、初手 take の相手と相互連携に入りやすくなる、という利点があるため、初手 wait のエージェントがなくなることがなかった、と推定できる。

以下、相互連携・搾取・拒絶の 3 条件について、各エージェントの戦略を比較しながら詳しく考察する。

4.1 相互連携行動

A01 から A41、および A61 が相互連携を行うことが確認できた。順位が高いエージェントほど相互連携を行えるエージェントを増やしており、効率の良い他者識別規則の適切な実装が、得点を増加させた主要因となったと考えられる。代表的な戦略として 2 位を取ったエージェント A02 の戦略を図 3 に示す。

相互連携の間隔は wait, take が同数であれば良いため、理論上は wait, take, wait, take...の交互、wait, wait, take, take...の交互、といった複数の相互連携方法が考えられる。しかしながら、実際に確認されたのは wait, take が交互に繰り返される、最も短い相互連携のペアが全ての相互連携ペアにおいて発生していた。

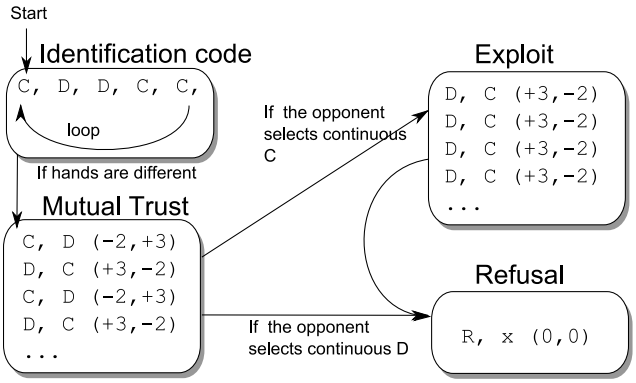

図 3 代表的な協調戦略のメタオートマトン（エージェント A02）

4.2 搾取行動

搾取はほとんどのエージェントで確認できた。A01 から A71 までの 71 体のエージェントが搾取アルゴリズムを持っていた。また、搾取を受けた非搾取エージェントは、A71 から A74 までの 4 対のエージェントである(付録 B、注 1)。

搾取戦略への移行条件としては、複数回の連続した wait を相手が選択した場合、が挙げられる。A10 を作成した参加者はゲーム後のレポートの中で、4 回連続で wait 出会った場合に搾取行動に移行する、と記述されている。また、搾取行動を行ったうち反撃を受けた場合には、拒絶を行うエージェントが多かった。

上位エージェントが搾取戦略を保持した理由として、下位 3 体の参加者(A72-A74)が参加から終了後までオートマトンを変更しなかったことが考えられる。これらのエージェントは、初期オートマトンのままであり、搾取に対する反撃アルゴリズムを持たなかった。また、A71(初期状態 2、戦略{(2,0,4), (2,1,4), (4,0,1), (4,1,1), (1,0,2), (1,1,2)})のように、3 回に 1 回 Take を行うだけで、相手の行動に反応して Take を行わないエージェントも存在した。これらのエージェントはすべて Take を行って来るエージェントに弱かった。これらのエージェントの存在が、弱い相手から搾取する抜け目の無い行動を他のエージェントに獲得させる要因になったと考えられる。

また、これ以外 7 体のエージェントでも非搾取が観察された(A23, A28, A32, A33, A61, A67, A69)。これらのエージェントはオートマトン戦略の中にバグを持っており、wait 手で固定となってしまう状態が見られた。攻撃側のエージェントはこれらの欠陥を突くように戦略を書き換えたものと思われる。

4.3 拒絶行動

本研究では双方が take を出した時の値がマイナスになっているため、協力できない相手を判定し拒絶行動を行うことが重要である。上位のエージェントは拒絶行動を持っており、A01 から A41 まで、相互連携を可能とするエージェントはすべて拒絶行動に対応することができていた。

拒絶行動を選択する基準としては、連続した複数回の take が相手から返ってきた場合に、拒絶を選択するエージェントが

多かった。例えば A02 のエージェントは自動生成で複数の状態遷移図を重畳させているため、3 回の take が相手から返された場合に、拒絶を行うように記述している。

一方で、1～2 回の take 行動に対しては、すぐに拒絶せず猶予期間を置くエージェントも多かった。これは、協力できる相手を逃さないための戦略と考えられる。

4.4 自己の信号と他者の認識

本ゲーム課題では、相手と同じオートマトンを選択すると、毎回出る手が同じになってしまうため、必ず交渉が失敗する、という特徴がある(これは参加者同士で戦略を共有する、というチート戦略に対する防止行為としても機能している)。

上位のエージェントは、相手と違う手を出した段階で相互連携に入れるかどうかを模索する例が多かった。そのため、相手と違う手をいかにして出せるかが重要とされた。

相手と違う手を出すために上位エージェントでもっともよく見られた戦略は、自身を定義付ける固有の行動パターンを繰り返し行い、それが相手と違ってきた場合に相互連携を模索した。例えば A02 は、wait, take, take, wait, wait という 5 回の行動を 1 組、A04 は wait,take,take の 3 回行動を 1 組とし、これを繰り返して発信することで信号としている。74 体のエージェントを識別するのに必要な信号は 2^7=128 通りであるが、識別戦略を考えていたのは全体の半数弱である。これは 2^5=32 に近く、5 個の組みより大きなセットを信号としたエージェントがなかったことに符合する。

取引ゲームにおいて、相手にとって認識しやすい信号が自然発生的に発生した、という現象は、人間の意図推定とコミュニケーション手段の発達、という課題に対し、示唆を与えるものと考えられる。

4.5 本研究の貢献範囲と改善点

今回の課題ではコンピュータ上のシミュレーションの代わりに、人間同士が各エージェントのオートマトンを記述して戦略の改良を行う。このような人間を最適解を求める計算手段として用い、得られた結果を分析してモデルを立てる手法はヒューマンエージェントインタラクション分野においてエージェントのプロトコルを探る手法としても有効である[9]。本研究の行った人間同士のゲームによるモデル取得は、このような計算手法がマルチエージェントシミュレーションの分野においても有効に働くことを示唆する。

ただし、このような人間を計算資源として利用する方法では、人間側の動機の設計を慎重に行う必要があると考えられる。本課題では 100 回以上の更新可能なタイミングが存在するが、実際に大きく更新が行われたのは開始時と終了間際の一週間であり、シミュレーションと異なり、すべての参加者が継続して改善を行なっていったわけではない。特に、更新を行わなかった下位3体のエージェントはゲームバランスに大きな影響を与えたため、ゲーム後のレポートでは参加者からの不満が見られた。また、倫理的な問題に触れないように、人間にとってゲーム参加が益あるものとして注意し、設計しなければならない。また、今回のゲームでは、わざと自分のエージェントの戦略を弱くしておき、相手に読まれない、などの対策も見られた。ゲーム後の参加者の意見では、途中で 1 位になったものが狙われて不利だった、というコメントも見られた。このような偽装を防ぎ、継続した改善を行わせるためには、最終結果を参加者の利益とするのではなく、途中の更新時の順位を利益にするのが望ましいと考えられる。

また今回は、100 回区間の有限回繰り返しの囚人のジレンマゲームとして人間同士の試合を行った。有限回繰り返しのゲームは最善手が異なるため、繰り返し回数を施行毎に不定にして実行するのが望ましいと考えられる。

5. 結論

本研究では人間集団同士の交代取引ゲームを行い、初手の交渉結果を情報として他者を識別する戦略が獲得されることを確認した。

参考文献

[Angeline 1994] Angeline, P.J: An Alternate Interpretation of the Iterated Prisoner's Dilemma and the Evolution of Non-Mutual Cooperation. *Proceedings of 4th artificial life conference* (pp. 353–358). 1994.

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

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

[Crawford 1982] Crawford, V. P., & Sobel, J. : Strategic Information Transmission. *Econometrica*, *50*(6), 1431 – 1451. 1982

[Fisher 2005] Fisher, R., & Shapiro, D: *Beyond Reason: Using Emotions as You Negotiate* (p. 256). Viking Adult, 2005.

[Le 2007] Le, S., & Boyd, R: Evolutionary dynamics of the continuous iterated prisoner's dilemma. *Journal of theoretical biology*, *245*(2), 258–67, 2007

[Nowak 1992] Nowak, M. A., & May, R. M: Evolutionary games and spatial chaos. *Nature*, *359*(6398), 826–829, 1992.

[Wärneryd 1991] Wärneryd, K: Evolutionary stability in unanimity games with cheap talk. *Economics Letters*, *36*(4), 375–378, 1991.

[大澤博隆 2007] 大澤博隆、今井倫太: 人間同士による繰り返し型囚人のジレンマゲームの実施と結果. *Game Programming Workshop* (pp. 188–194), 2007.