「価値観」の形成：エージェントによる構成論的研究

研究ノート

森　山　甲　一*

Forming "Personal Values": Constructive Approach by Using Artificial Agents

Key Words：Reinforcement Learning, Reward Structures, Evolution, Genetic Algorithm, Game Theory

1　はじめに

我々は，毎日様々な意思決定を行っている．例えば，昼食に何を食べようか，今度の休みはどこに行こうか，などと複数の選択肢から選択することを日々行っているだろう．ある人の意思決定がその人にとって望ましい状況を導くならば，その意思決定はその人にとって合理的であると言うことができる．例えば，*A* さんが「昼食にハンバーガーを食べよう」と意思決定をする．*A* さんがその時にハンバーガーを食べたいと思っていれば，食べることが満足感という望ましい状況を導くので，合理的である．次に，*A* さんを観察している人を考えてみよう．ただし，この観察者は肉を食べない菜食主義者だとする．もし，この観察者が，世の中の全ての人は自分と同じ菜食主義者だと信じていれば，*A* さんを見て，「何でハンバーガーなんて食べるのだろう」と感じる，つまり *A* さんの意思決定は非合理的であると感じるかもしれない．

すなわち，個々の人間にはそれぞれの価値観が存在し，自分の価値観に従って合理的に行動している人を，別の価値観を持つ他者は非合理的だと感じるかもしれない．このような価値観の違いは相手がどう振る舞うかの予測を困難にし，個人のレベルから国家間や宗教など国家を超えたレベルまでの様々な争いを生み出す．上の例は馬鹿げたものに思えるかもしれないが，例えば，ハンバーガーを鯨肉と置き換えたらどうだろうか．あるいは，同性同士の婚姻の是非といった意思決定ならどうだろうか．

争いなど，複数の意思決定個体間の相互作用を扱う分野にゲーム理論がある．ゲーム理論では，金銭に代表される数値化された利得を最大化することが全個体にとって望ましいとの共通の仮定の下での意思決定を扱ってきた．ところが，ゲーム理論の結果と人による被験者実験の結果とが異なることが知られている．この結果の乖離には様々な原因が考えられているが，必ずしも人類全体が守銭奴なわけではなく，個人ごとに異なる価値観を持つためとは考えられないだろうか．

筆者らはこのような考えに基づき，個体の価値観がどのように形成されるかを，コンピュータ上に構築した仮想的な意思決定個体（以下エージェントと呼ぶ）を用いて構成論的に解明しようと試みている．本稿では，その中でもエージェントの「価値観」[1] を進化計算を用いて自動的に獲得する筆者らの研究 [1] について簡単に紹介する．

2　学習するエージェントと「価値観」

人間は遺伝子に組み込まれた設計図にのみ従って行動するのではなく，生活の中で学習することにより行動を変更する能力を持っている．従って，エージェントにも学習による行動変更能力を持たせることにする．ここでは，行動に対して環境から与えられる報酬に基づき，それをより増大させる行動を選択するように学習を行う強化学習 [4] を用いることにする．強化学習も前述したゲーム理論などと同様に，環境から与えられる数値化された報酬（の期待値）を最大化することが望ましいとの仮定に基づい

[1] 本稿で扱う「価値観」は学術用語としての価値観とは異なる可能性があるため，念のためカギカッコ付きで表記する．

* Koichi MORIYAMA
1975年12月生
東京工業大学　大学院情報理工学研究科
計算工学専攻　博士後期課程（2003年）
現在、大阪大学　産業科学研究所　第1研究部門（情報・量子科学系）　知能アーキテクチャ研究分野　助教　博士（工学）
人工知能・マルチエージェントシステム
TEL：06-6879-8426
FAX：06-6879-8428
E-mail：koichi@ai.sanken.osaka-u.ac.jp

ている．ところが，報酬を入力として状況の望ましさを出力とする「価値観」メカニズムをエージェントに導入し，エージェントはその出力（以下効用と呼ぶ）を用いて強化学習をすることにすれば，このエージェントは自身の「価値観」に基づく価値判断の下で，効用を増大させる行動を選択するように学習を行うことになる．

では，どのようなメカニズムを「価値観」として導入すれば良いだろうか．エージェントを設計する立場からは任意の「価値観」を作り込むことが可能だが，我々人間の価値観は進化により得られたものであると思われるため，遺伝的アルゴリズムを用いた進化計算によって「価値観」を進化により獲得することを試みた．遺伝的アルゴリズム [3] は生物の進化を模したアルゴリズムであり，複数の遺伝子からなる染色体（解候補）を複数用いて，それらの選択，交叉，突然変異を繰り返し行うことにより，適合度（求める解への合致度）を最大化する染色体を求めるものである．

3　エージェントの「価値観」の進化

筆者らは，エージェントが囚人のジレンマゲーム [2] と呼ばれるゲーム理論の分野で非常に有名なゲームを行ったとき，どのような「価値観」が現れるかを進化計算手法を用いて調査した．このゲームの例を表1に示す．相手の選択に関わらず D を選択する方が利得が大きいが，両者が C を選択すると，両者が D を選択して得られる利得よりも大きな利得が得られるゲームである．

ここでは，「価値観」メカニズムとして報酬（＝利得）r を引数とする関数

$$u(r) = ar^3 + br^2 + cr + d$$

を導入した．エージェントはこの出力である効用 $u(r)$ を用いて強化学習を行うものとし，各係数 a, b, c, d（ただし $-10 \leq a, b, c, d \leq 10$）を遺伝子，

表1：囚人のジレンマゲームの例：プレイヤー P_1 が行から，P_2 が列からそれぞれ行動 C，D を選び，その組合せのところに記述されている数値（左：P_1，右：P_2）が各々の利得となる．

$P_1 \backslash P_2$	C	D
C	3, 3	0, 5
D	5, 0	1, 1

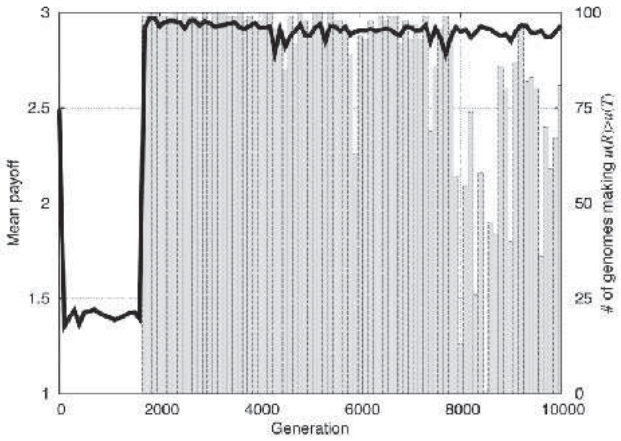

図1：ある実験における1ゲームあたり平均利得（折れ線グラフ：左目盛り）および $u(3) > u(5)$ となった個体数（棒グラフ：右目盛り）

その配列を染色体とした遺伝的アルゴリズムで係数を進化させた．まず，係数をランダム値としたエージェントを100個体用意し，総当たりで1000回ずつ囚人のジレンマゲームを行わせ，行動戦略を学習させた．そして，エージェントが得た報酬 r の合計を適合度として係数を進化させた．行動戦略を係数の進化ごとに初期化することで係数だけが次の世代へ伝わる．これを10000回（世代）繰り返し，最終的にどのような係数が得られたかを100回調査した．

100回の実験のうち83回で10000世代時の100個体の1ゲームあたり平均利得が2.7以上となった．$(5+0)/2 = 2.5$ から，両者が C を選択するケースがないとこの値にはならないことに注意されたい．

ある1回の実験の各世代における100個体の平均利得を図1の折れ線グラフに示す．図から，数十世代ほどの比較的短い時間において急激に平均利得が上昇する一種の相転移が起こっていることが分かる．

この実験における10000世代時の適合度最大の染色体が表現する関数 $u(r)$ は

$$u(r) = -1.18073r^3 + 7.81789r^2 - 4.44985r - 10$$

となった．$u(3) = 15.13175$ および $u(5) = 15.60675$ とわずかに $u(3) < u(5)$ となっている．また，$u(0) < u(1) \ll u(3)$ となっており，効用の順序関係も表1と同一，すなわちこのエージェントの効用も囚人のジレンマゲームになっていることが分かる．

図1の棒グラフは $u(3) > u(5)$ となった個体数を表している．これは，両者 C の時の効用が自分だ

け D の時の効用よりも大きいことを意味しており，C を選択しやすくする効果がある．図から，相転移時にほぼ全ての個体が $u(3) > u(5)$ となり，それから増減していることが分かる．

4　おわりに

本稿では，意思決定個体の価値観がどのように形成されるかをエージェントを用いて構成論的に解明する研究として，エージェントの「価値観」を進化計算により獲得する研究 [1] を紹介した．環境から与えられる報酬を状況の望ましさとして解釈する「価値観」メカニズム $u(r)$ を導入し，その係数を遺伝的アルゴリズムを用いて求めた．100 回実験を繰り返したところ，うち 83 回で両者 C をもたらすことが分かった．また，それが生じる際には一種の相転移が起こることを示した．

紙面の都合もあり簡単にしか本研究を紹介出来なかったため，詳しくは文献 [1] をご覧頂きたい．他にも非常に興味深い現象が観察されているが，なぜこのような結果が生じるのかについては不明な点が多い．特に，もっとも重要な相転移前後の挙動についてはまだ分からないことばかりである．

本研究はまだ基礎の基礎の段階であり，「はじめに」に述べたような社会的問題への適用ははるか先の話である．しかし，そのような問題の根本的原因を探る研究は必要かつ重要であり，今後の進展が望まれる．

参考文献

[1] K. Moriyama, S. Kurihara, and M. Numao. Evolving Subjective Utilities: Prisoner's Dilemma Game Examples. In *Proc. 10th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pp. 233−240, IFAAMAS, Richland, SC, 2011.

[2] W. Poundstone. *Prisoner's Dilemma*. Doubleday, New York, 1992.（松浦 訳．囚人のジレンマ．青土社，東京，1995）．

[3] 坂和，田中．遺伝的アルゴリズム．ソフトコンピューティングシリーズ 1. 朝倉書店，東京，1995.

[4] R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA, 1998.（三上，皆川 訳．強化学習．森北出版，東京，2000）．