5L-1

非ゼロ和ゲームにおける相手の協調行動を誘う戦略

大橋　資紀†，伊藤　昭†，寺田　和憲†
†岐阜大学工学部応用情報学科

はじめに

我々は利害の対立する状況下でも，必要があれば協調行動の採れる計算機を作ることを目指している．そのためには，計算機にも相手の意図を読み，また相手に意図を伝える能力が必要である．そのようなことを目標に，高得点を獲得するためには協調的行動を必要とする非ゼロ和繰り返しゲームにおいて，人がどのような行動をとるのかを調べることで，意図を読み，伝えることができる計算機に必要な機能についての検討を行った．

同じ非ゼロ和ゲームである繰り返し囚人のジレンマゲーム(IPD)[1]では，人は協調できることが知られている．しかしながら，IPD では協調行動は自明であり，「意図の伝達」が問題とはならない．我々は，相手の行動から意図を読むことがそれほど容易ではない課題として，1・2・5 じゃんけんを考案した．

1・2・5 じゃんけん

1・2・5 じゃんけんは普通のじゃんけんと同じく，2 人のプレイヤが，グー(G)，チョキ(C)，パー(P)のいずれかの手を同時に提示する．勝敗も普通のじゃんけんと同じであるが，勝ったときの手によって図 1 のように異なる点数が与えられる．ただし，負けやあいこのときは 0 点となる．

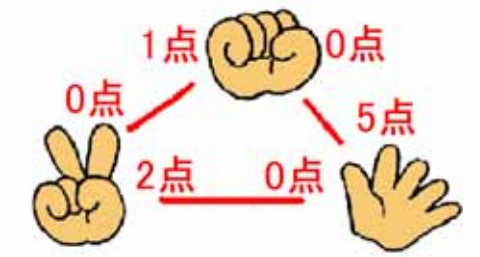

図 1 : 1・2・5 じゃんけんの得点

このゲームでは，ランダムに手を選択した場合は平均 8/9 点である．相互に相手の手の最適戦略となっている Nash 均衡解は，G：C：P を 2/17：10/17：5/17 の確率で選択する混合戦略であり，平均得点は 10/17 点である．これに対して，お互いに G と P を交互に出す戦略を採れば，双方とも平均 2.5 点を得ることができる．ここではこれを協調戦略と呼ぶ．

本研究の目的は，人がこのような協調戦略を発見できるか，また発見できたとして相手を協調戦略に引き込むためにどのような行動を採るのか，採るべきか，ということである．

予備実験

被験者同士で 1・2・5 じゃんけんをプレイしてもらい，人がどのように対戦するのか，協調戦略をとれるのかどうかを調べてみた．実験では，じゃんけんを 100 回戦繰り返し，正の得点に対して点数に比例した報酬を支払った．ただし，1 回毎に-1 点のコストを払うものとした．被験者には，ゲームを行う前にルールを説明した上で，「ゲームの目的は自身の得点をできるだけ高くすることであり，相手との得点の差を争うことではない．」と教示した．この実験では，被験者同士でも，協調を試みた被験者はいたものの協調を実現した組はいなかった．

協調誘導実験

どのような戦略が，相手に協調の意図を伝え，協調行動をとらせることができるのかを調べる．そのために，被験者と協調に誘導しようとする計算機で対戦を行った．実験方法は予備実験と同様であるが，被験者に対して，「相手は，同じような被験者である．」と嘘の教示をした．被験者の対戦相手として用いた計算機のアルゴリズムは以下の 3 種類である

N戦略

・前回相手に P で勝ったら，G を出す．
・前回相手に G で負けたら，P を出す．
・それ以外のときは G，C，P を(2/17:10/17:5/17)の割合(Nash 均衡解)で出す．

NGC 戦略

・以下を除いて N 戦略と同様
・Nash 均衡解が 5 回戦続いたら，それ以降は，相手が P を出すまで G を出し続ける．

NCC 戦略

・前回 G で負けると，引き続く 5 回は，P，G，P，G，P と出す．前回 P で勝つと，引き続く 5

The Strategy which Induce Other Person to Cooperation Action in Non-zero-sum Game
Motoki Ohashi†, Akira Ito†, Kazunori Terada†
†Gifu Univ.　Department of Information Science

回は，G，P，G，P，G と出す．
・上記以外は，Nash 均衡解を出す．
・Nash 均衡解が 6 回戦続いたら相手が P を出すまで G を出し続ける．

実験結果

協調は P，G を交互に出すことで成立する．そこで，P，G の組の 1 試合あたりの平均出現回数を表 2 に示す．GP は計算機が G 被験者が P となった手の組を，また PG はその逆を表す．また PG，GP が 5 回以上繰り返された時を協調の実現と定義するとき，協調の発見者数・実現者数を表 3 に示す．協調が多く実現した戦略ほど，被験者が P で勝つ回数が少なく，G で負ける回数が多い．

表 2 : 協調手の組の出現回数

戦略	GP の出現回数	PG の出現回数
N	11.82	13.45
NGC	12.40	6.40
NCC	23.60	4.80

表 3 : 協調の実現回数

戦略	実験組数	協調 発見者数	協調 実現者数
N	11	5	4
NGC	5	3	1
NCC	5	0	0

Q 学習との対戦

これまでは人を協調に誘導できるかの実験であったが，次に同じアルゴリズムが学習戦略を協調に誘導できるかを調べる．そのため前の実験で用いた戦略と Q 学習戦略との対戦を 10^6 回戦（ステップ）行い，協調までにかかるステップ数を調べた．

Q 学習は，状態 S のとき行動 a をとることの価値（Q 値）を学習する．学習は，ステップ毎に次に示す式によって行う．

$$Q(s_t,a_t)=(1-\alpha)Q(s_t,a_t)+\alpha(r+\gamma\max_a Q(s_{t+1},a))$$

s_t と s_{t+1} は現在のステップ t と次のステップ t+1 のときの状態，a_t はステップ t での行動，r はステップ t での利得である．学習率 α，割引率 γ はパラメータであり，α =0.1，γ =0.9 とする．
行動は次式で求めた確率にしたがって選択する．

$$p_t(a_t)=\frac{\exp(Q_{t-1}(s_t,a_t))}{\sum_a \exp(Q_{t-1}(s_t,a_t))}$$

ただし，これとは別に ε =0.05 の確率でランダムに行動を選択するものとする．

この Q 学習と各戦略との対戦をそれぞれ 100 回行い，協調までにかかったステップ数の分布を求める．ただし，ここで協調までにかかったステップ数とは，各プレイヤが最後にマイナス点からプラス点に転じたステップ数の平均である．10^6 回戦の時点で少なくとも片方のプレイヤがマイナス点の場合，協調が発生しなかったとする．

実験結果を図 2 に示す．NGC 戦略と N 戦略は Q 学習同士の対戦よりも早く協調できている．それでも，ほとんどが 10^4～10^5 ステップかかってしまっている．また，NCC 戦略は被験者との対戦と同様に，協調した組が存在しなかった．

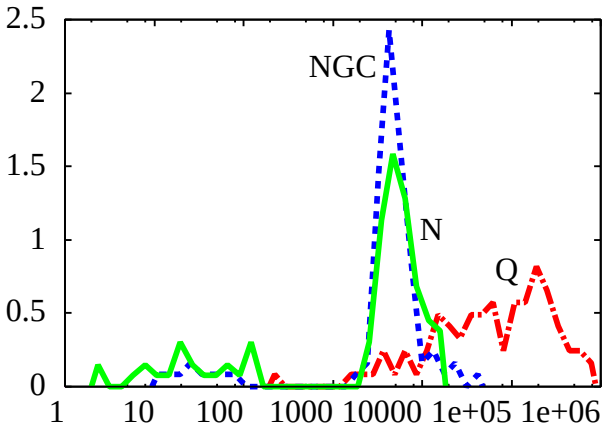

図 2 ： 協調までにかかったステップ数の分布

まとめ

1・2・5 じゃんけんにおいて，人は協調行動を発見できるか，そして，どのような戦略が相手を協調行動に誘導できるかを実験により調べた．その結果，協調を誘導するために必要な条件は，1)相手が協調行動をとらないときは相手に点を与えないこと 2) こちらが得点した時は次に相手に得点させるというような，自己利益追求からは解釈できない行動を取ることで，協調の意図を伝達すること，の二つであることが分かった．

Q 学習は相手の意図を読まずに，統計的に学習するだけのため，協調までに長い時間を要している．しかしながら，上記 2)で述べたような行動を正しく解釈できる相手モデルを計算機が獲得できればよいわけで，そのようなアルゴリズムを開発することが一つの目標となる．

参考文献

[1] Axelrod,R.: *The Evolution of Cooperation*, Basic Books Inc., (1984), 松田裕之訳, つきあい方の科学,HBJ 出版局,(1987).