2X−7

フットステップにおける効率的な学習手法

南雲夏彦
（ 神奈川大学　理学部 ）

１．まえがき

アブストラクトボードゲーム「フットステップ」の戦略を習得する手法を比較することを通して、駆引きの要素を含むゲームの効率的な学習について検討する。

２．二人用ゼロ和ゲームの分類

二人用ゼロ和ゲームは、各プレイヤーの所有する情報量により、完全情報ゲームと不完全情報ゲームに分類される。完全情報ゲームは、各プレイヤーがゲームに関するすべての情報を所有しているゲームである。これらのゲームは、確定ゲームと不確定ゲームに分類される。確定ゲームは、少なくとも一方のプレイヤーに避敗戦略が存在する。即ち、どんな局面からでも、双方のプレイヤーが最善を尽くすと、常に一方のプレイヤーが勝つかまたは引き分けになる。一方、不確定ゲームでは、各々の局面において最善手（少なくともｎ％の勝率を保証する手）が存在することが証明されている。確定ゲームは不確定ゲームの特異なケースと考えることもできる。

一方、不完全情報ゲームは、一部の情報しか各プレイヤーが所有していないゲームである。局面における最善手が一つに決まらず、混合戦略が最善になることが多い。

３．拡張フットステップのルール

①各競技者にはビッド用にｎポイントの持ち点が与えられる。
②フットステップ盤（図１）のセンターライン上の交点にマーカを配置する。
③各々の競技者は、同時に、１ポイント以上自分の持ち点以下まで、任意な数のビッドをおこなう。
④相手の競技者より大きいビッドをした競技者は、マーカを相手側に１コマ進める。これで１ターンが終わる。

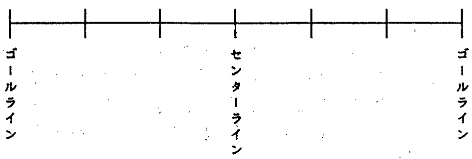


図1　（簡易化）フットステップ盤

⑤両方の競技者の持ち点が無くなるまでビッドを繰り返す。途中で持ち点をすべて使いきった競技者は、自動的に、それ以後のビッドは０ポイントになる。０ポイントをビッドできるのはこの場合に限られる。
⑥マーカをセンターラーンからｋコマ進めて相手のゴールラインに到達させるか、両方の競技者の持ち点が無くなれば、ゲームは終了する。ゲームが終了したとき、マーカをゴールラインまで進めれば２点勝ち、相手の陣地内にマーカがあれば１点勝ちになる。マーカがちょうどセンターライン上にあれば引き分けである。

４．局面と戦術の表記法

フットステップの局面は、「そのときの自分の持ち点（ｘ）」および「そのときの相手の持ち点（ｙ）」および「マーカの位置（ｚ）」のみによって表記できる。過去の履歴やターン数とは無関係である。このときに、ある確率分布に基ずいて、１以上ｘ以下までビッドする戦略を混合戦略と呼ぶ。

$$\sum_{i=1}^{x} P(i \mid x, y, z) = 1$$

混合戦略の特異なケースとして、確率分布によらず、ｘ，ｙ，ｚの関数として自分のビッドを決定するを方法がある。これを単純戦略と呼ぶ。

$$i = F(x, y, z)$$

５．混合戦略の優位性

標準型のフットステップはｎ＝50：ｋ＝４であるが、今回はｎ＝20：ｋ＝３の簡易化フットステップについて検討する。単純戦略の第ｍターンのビッドをf_mとするとき、単純戦略に対抗する戦略の第ｍターンのビッドをb_mとして、簡易化フットステップでは、（標準型のフットステップでも）単純戦略が必ず２点負けになることを示す。

（１）$f_1 \geqq 5$のとき

$b_1 = 1$　$b_i = f_i + 1$ $(2 \leqq i \leqq 5)$　にて対抗戦略の２点勝ち

On Optimal Learning Method Using the Game of Footstep

Natsuhiko Nagumo
Department of Information and Computer Science, Kanagawa Univ.

（２）$f_1 < 5$：$f_2 \geqq 5$のとき
$b_1 = f_1 + 1$　$b_2 = 1$　$b_i = f_i + 1$（$3 \leqq i \leqq 5$）　にて対抗戦略の２点勝ち
（３）$f_1 < 5$：$f_2 < 5$：$f_3 \geqq 5$のとき
$b_1 = f_1 + 1$　$b_2 = f_2 + 1$　$b_3 = 1$　$b_i = f_i + 1$（$4 \leqq i \leqq 5$）　にて対抗戦略の２点勝ち
（４）$f_1 < 5$：$f_2 < 5$：$f_3 < 5$のとき
$b_1 = f_1 + 1$　$b_2 = f_2 + 1$　$b_3 = f_3 + 1$
にて対抗戦略の２点勝ち

６．ランダムプレーヤーの学習

フットステップにおけるランダムプレイヤーとは、自分の持ち点からベット可能な各々の個数を等確率でビッドするプレイヤーである。即ち、自分の持ち点がｘのときに、$1/x$の確率で$1 \sim x$個をビッドするプレイヤーのことである。

どんな学習機能を持つランダムプレイヤーが、ランダムプレイヤーを相手にしたときに効率的に戦略を学習できるか検討した。

・奨励戦略

１回のゲームが終了する毎に、勝ったゲームについて、実際に行なわなかったビッドの確率を一定の割合で減少させて、減少させた分だけ、実際に行なったビッドの確率を増加させる戦略。即ち、自分の持ち点がｘ、相手の持ち点がｙ、マーカーの位置がｚの局面で、ｉのビッドを行ない、結果的に勝ったとすると、ｃを減衰係数（$0 \leqq c < 1$）として、

$1 \leqq k\ (\neq i) \leqq x$なるｋに対して、

$$P_{new}(k \mid x, y, z) = c \cdot P_{old}(k \mid x, y, z)$$

$$P_{new}(i \mid x, y, z) = (1 - c) + c \cdot P_{old}(i \mid x, y, z)$$

・叱責戦略

１回のゲームが終了する毎に、負けたゲームについて、実際に行なったビッドの確率を一定の割合で減少させて、減少させた分だけ、実際に行なわなかったビッドの確率を増加させる戦略。即ち、自分の持ち点がｘ、相手の持ち点がｙ、マーカーの位置がｚの局面で、ｉのビッドを行ない、結果的に負けたとすると、ｃを減衰係数（$0 \leqq c < 1$）として、

$$P_{new}(i \mid x, y, z) = c \cdot P_{old}(i \mid x, y, z)$$

$1 \leqq k\ (\neq i) \leqq x$なるｋに対して、

$$P_{new}(k \mid x, y, z) = P_{old}(k \mid x, y, z) + (1 - c) \cdot P_{old}(i \mid x, y, z) / (x - 1)$$

・飴鞭戦略

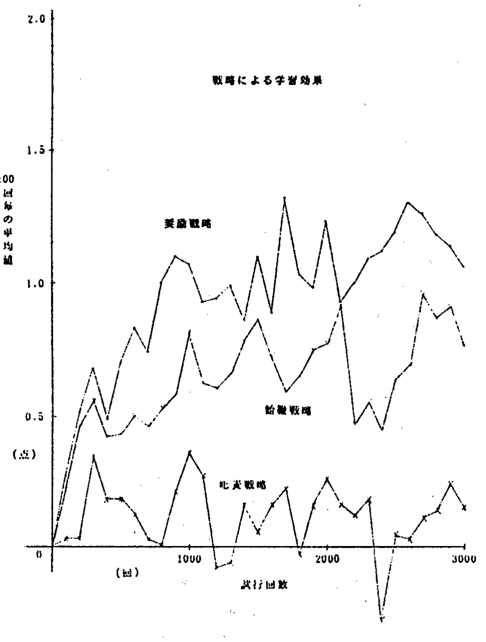
１回のゲームが終了する毎に、勝負結果によって、奨励戦略と叱責戦略の両方を用いて学習する戦略。但し、引き分けのときには何の学習も行なわない。

７．戦略の修正

パラメーターの数が多すぎて、少ない試行回数では有効な実験結果が得られなかったので、取りあえずマーカの位置を考慮しない確率分布を考えて実験を行なった。

８．結果

ｃ＝0.9として実験を行ない、それぞれの戦略について100回毎に得点の平均値を計算した。その結果、このケースにおいては奨励戦略が一番優れており、飴鞭戦略がその次で、叱責戦略が最も劣っていることが判った。（図２）



９．むすび

ランダムプレイヤーを相手にしたときに、どこまで得点の期待値を上昇させることが可能か検討する。さらに、自己学習の有効性を検討し、人間の上級者の戦略との相違点を比較検討する。

参考文献（１）南雲夏彦：フットステップとスーパーパズ,共立出版社,bit,Vol.23,No.5, pp.99-101,1991.