

繰り返し対戦型じゃんけんにおける戦略の特徴

牧野 泰裕* 西野 順二** 小高 知宏** 小倉 久和**

A Characteristic of Strategy for the Repeated Janken Game

Yasuhiro MAKINO, Junji NISHINO, Tomohiro ODAKA and Hisakazu OGURA

(Received Feb. 28, 1997)

In this paper, we investigate of the repeated janken game to estimate characteristic of game agents. In the repeated janken game, players repeat single "janken" reiteratively. The repeated janken game has strategy, for examples, random strategy, fixed strategy, repeated strategy, history-based strategy that uses history of opponent hands. We analyze superiority of strategy by computer simulations. As a result, we show superiority of history-based strategy.

1 はじめに

本研究では、チェス、オセロ等よりも、比較的単純な規則で、勝負が容易に分かるじゃんけんを対象として、対戦型ゲームにおける戦略知識獲得を試みる [1][2]。戦略の獲得のためには、どのような戦略も表現できるような戦略の表現方法が必要である。本研究は対戦ゲームとして、繰り返しじゃんけんをとりあげ、その戦略の表現と、特性を調べることを目的としている [3]。

さて、じゃんけんは、相手の手を見てから、自分の手を決めるようなものでなく、1回限りの対戦では、相手の手と自分の手を同時に出して勝負を決めるために、戦略は存在しない。そこで、今回はじゃんけんに戦略を取り入れようとするために、じゃんけんを繰り返して行う。じゃんけんを何回も繰り返して行っていくなかで、1回ごとの勝負の結果や、相手の手の履歴などを用いる事で、勝負における戦略というものを考える事ができる。そのために、じゃんけんを行う回数や、勝敗に対する得点、といったルールを決めて、繰り返し対戦型のじゃんけんゲームを設定する。

各々に戦略を持ったプレーヤーを集めて、そのプレーヤー同士が各自の戦略にしたがって、繰り返し対戦を行なっていく、総当たりの試合を行なっていく、その試合のたびに得られた得点の合計により、

*大学院情報工学専攻

**工学部情報工学科

プレーヤー、つまり戦略の順位が決められる。戦略には、次の2章で示すようなものを用いることにした。今回は、繰り返し対戦型じゃんけんゲームにおいて、用意した戦略の中で、高得点を得るのは、どんな戦略であるかを調べた。3章では、それらの戦略について、他の戦略と比較してそのような結果を得られたのか、特徴を調べる。その結果から、どのような戦略が有効であるかを考え、繰り返し対戦型のじゃんけんというものの戦略をどのように実現するかを検討する。

2 繰り返し対戦型じゃんけんゲームの設定

2.1 対戦方法・ルール

繰り返し対戦型じゃんけんゲームの対戦は、全てコンピューター上で行なう。以降プレーヤーとは、プログラムされた戦略のことを指す。じゃんけんは、2人のプレーヤー同士で、一般に知られているじゃんけんのルールに従い行なう。各々のプレーヤーは、それぞれに自分自身の戦略を持ち、その戦略に従ってじゃんけんをする。じゃんけん1回の勝負ごとに、以下の表からその結果に応じた得点が、プレーヤーに与えられる。

A \ B	g	t	p
g	0, 0	1, -1	-1, 1
t	-1, 1	0, 0	1, -1
p	1, -1	-1, 1	0, 0

g : グー
 t : チョキ
 p : パー

Aがチョキ、Bがパーの時、
 Aが1点、Bが-1点を得ることを示す。

g
 t ← p
 ← 矢印の方が強い
 gはtより強い
 tはpより強い
 pはgより強い

図 1: じゃんけんの利得表

対戦は、以下のように、コンピューター上でシミュレートする。

1. じゃんけんを行なうプレーヤーエージェントを用意する。各プレーヤーは、それぞれの戦略を持ち、その戦略に従った行動をする。
2. 各勝負は1対1で行なう。プレーヤーは、1回のじゃんけんごとに、勝ち、負け、引き分けのそれぞれに応じて図1のような得点を得る。
3. プレーヤーは、1試合について同じ相手と続けて n 回のじゃんけんをする。(引き分けの場合も1回とする。)
4. コンピューター上で行なうため、自分自身を含めたプレーヤー全員と試合を行う。
5. 1、2、を総あたりで行い、全試合を終了した時点で、それぞれの対戦結果を集計して、勝ち数から負け数を引いたものを、総合得点とする。

結果から、高得点順にプレーヤーの持っている戦略を、成績表にしてまとめる。

2.2 各種戦略の定義

以下に、今回検討した5種39個の戦略について、詳しく説明する。

2.2.1 でたため戦略：1個

1回ごとに出す手を決定する場合に、ランダムに3種類の中から選択する戦略である。

ある手番上に出す自分の手 h_t は、グー、チョキ、パーをそれぞれ、g、t、p とすると、次のように表せる。

$$h_t = \text{random}(g, t, p) \quad \text{random は } g, t, p \text{ を等確率でランダムに選択する。}$$

2.2.2 一筋戦略：3個

g、t、p の中で1つだけを選択し、試合中ずっと同じ手を出し続けて、他の2種類は出さない戦略である。一筋戦略は、

$$h_t = g \quad (\text{g一筋戦略})$$

$$h_t = t \quad (\text{t一筋戦略})$$

$$h_t = p \quad (\text{p一筋戦略})$$

の3種類の場合のいずれか1つである。

図2の例は、g (グー) 一筋戦略の対戦の様子であり、相手の手に関係なく、最初からずっとグーを出し続けているのが分かる。

2.2.3 ものまね戦略：4個

相手の数手前の手と同じ手を、自分の手として出す。

1手前の手と同じ手をまねる場合は、自分の手の系列 $(h_t, h_{t-1}, h_{t-2}, \dots)$ 相手の手の系列 $(o_t, o_{t-1}, o_{t-2}, \dots)$ とすると、相手の手の1手前を真似る、ものまね1戦略の場合は、

$$h_t = o_{t-1}$$

となる。1手前を真似る、ものまね1戦略の例を図3に示す。

2手、3手前のものまね戦略の場合は、

$$h_t = o_{t-2}, h_t = o_{t-3}$$

と表せる。

勝負の回数	1	2	3	4	5	...
相手の手	g	t	p	g	t	...
自分の手	g	g	g	g	g	...

図2: g一筋戦略の例

勝負の回数	1	2	3	4	5	...
相手の手	g	t	p	g	t	...
自分の手		g	t	p	g	t

図3: ものまね1戦略の例

2.2.4 場合分け戦略：27 個

自分の一つ前の勝負の結果が、勝ち、引き分け、負けのそれぞれの場合の時、その次の手を 3 種類のいずれかの手にあらかじめ決めておく戦略であり、以下のように表す。

$$h_t = f(h_{t-1}, o_{t-1})$$

$$f(h_{t-1}, o_{t-1}) = \begin{cases} h^1, & (h_{t-1}, o_{t-1}) \in \{(g, t), (t, p), (p, g)\} \quad (\text{勝ち}) \\ h^2, & (h_{t-1}, o_{t-1}) \in \{(g, g), (t, t), (p, p)\} \quad (\text{引分}) \\ h^3, & (h_{t-1}, o_{t-1}) \in \{(t, g), (g, p), (p, t)\} \quad (\text{負け}) \end{cases}$$

ただし、 h^1, h^2, h^3 は、勝ち、引き分け、負けの場合に出す手であり、それぞれ g, t, p のいずれかに設定しておく。その組合せにより以下の図 4 のような 27 種類の戦略がある。

	h^1	h^2	h^3	
	勝ち	引き分け	負け	
g^3 戦略:	g	g	g	} 27 種類
$g^2 t$ 戦略:	g	g	t	
$g^2 p$ 戦略:	g	g	p	
⋮	⋮	⋮	⋮	
p^3 戦略:	p	p	p	

図 4: 場合分け戦略

例えば、勝った場合に g 、引き分けた場合に t 、負けた場合に p を出す、場合分け戦略である $g t p$ 戦略の対戦は以下の図 5 のようになる。

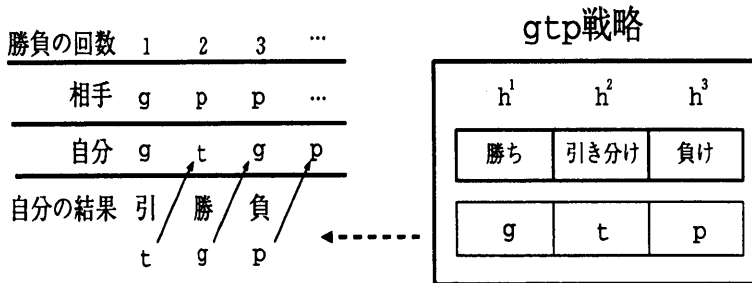


図 5: 場合分け $g t p$ 戦略の処理

2.2.5 履歴学習型戦略：4 個

過去の相手の手の系列を、ある長さの履歴パターンの出現回数として記憶しておき、その履歴から、次の相手の手を予測し、それに勝つような手を出す戦略。

長さが 2 の履歴を用いる場合の履歴学習型戦略は、次のように書ける。

$$h_t = f(o_{t-2}, o_{t-1}, s_t)$$

o_{t-2}, o_{t-1} は相手の1手前、2手前の手であり、この2つと、相手が次に出すと考えられる手とを組み合わせた長さ3の履歴のパターンを用いる。ベクトル s_t はプレーヤーの現在の状態を表す。状態の要素 w_{XYZ} は、過去に相手が手X、Yの次に手Zを出した回数を示す。

$$s_t = (w_{ggg}, w_{ggt}, \dots, w_{ppp})$$

$$s_1 = (0, 0, \dots, 0)$$

実際に出す手は、相手の過去の履歴の中から o_{t-2}, o_{t-1} の2つに相手の次の手と考えられる g、t、p の3つをそれぞれに組み合わせた長さ3のパターンの出現回数、 $w_{o_{t-2}, o_{t-1}, g}, w_{o_{t-2}, o_{t-1}, t}, w_{o_{t-2}, o_{t-1}, p}$ が最も大きい手が、予想される相手の手である。この場合に予測される相手の手に勝つ手を出す。

状態 s_t のそれぞれの要素は、次のように更新する。

$$w_{o_{t-3}, o_{t-2}, o_{t-1}}^t = w_{o_{t-3}, o_{t-2}, o_{t-1}}^{t-1}$$

例えば、 $(o_{t-3}, o_{t-2}, o_{t-1}) = (g, g, g)$ とし、 s_{t-1} に対し、状態 s_t は以下ようになる。

$$s_{t-1} = (1, 0, \dots)$$

$$s_t = (2, 0, \dots)$$

このとき、状態 s_t の要素 w_{ggg} がカウントされる。図6に履歴2の場合の例を示す。履歴3、履歴4、履歴5の場合も同様である。

履歴2 戦略の場合

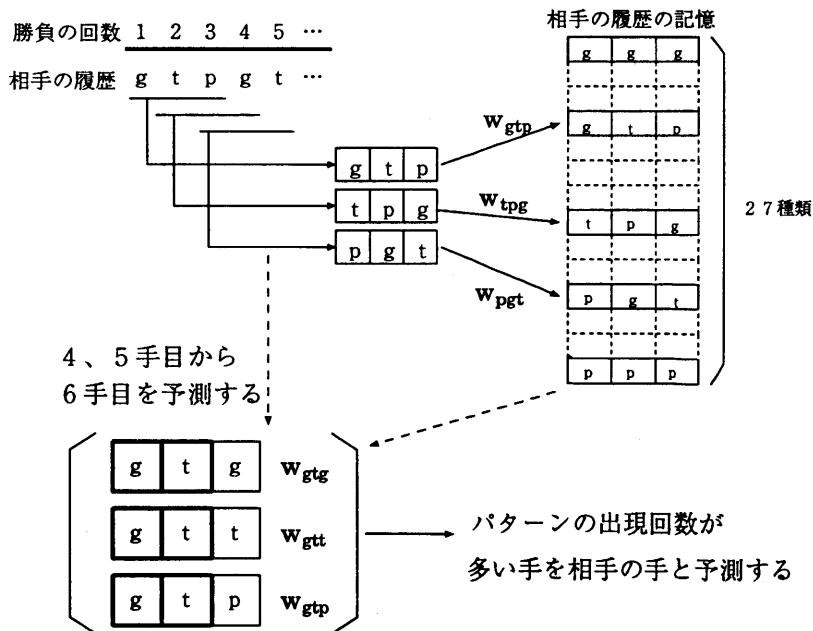


図6: 履歴学習型戦略

3 対戦結果と考察

まず総あたりの対戦を行い、その結果から履歴学習型戦略についてのみの対戦を行った。その結果を、それぞれ3.1、3.2で説明する。

3.1 総あたりの結果

用意した全ての戦略が他の全ての戦略と行なった1試合ごとの結果を、全試合について、勝ち数、負け数、引き分けの数を集計した結果を、以下の表に示す。総合得点は、得点表より、勝ち数と、負け数の差により得られる。

以下に、1試合につき $n = 1000$ 回、5000回、10000回ずつのじゃんけんを、総プレイヤー数39個での総あたりの結果のうち上位10位までを、表1~3に示す。

順位	戦略	勝ち	負け	引き分け	総合得点 (勝数-負数)
1	履歴4戦略	32085	3328	3587	28757
2	履歴3戦略	32087	3427	3486	28660
3	履歴5戦略	31649	3692	3659	27957
4	履歴2戦略	31192	3505	4303	27687
5	でたため戦略	13044	12960	12996	84
6	g g t 戦略	9839	10917	18244	-1078
7	t t p 戦略	9757	10880	18363	-1123
8	p p g 戦略	9818	11217	17965	-1399
9	t g p 戦略	8554	10834	19612	-2280
10	g t p 戦略	10002	12364	16634	-2362

表 1: 1 試合に 1000 回のじゃんけんを行った結果

順位	戦略	勝ち	負け	引き分け	総合得点 (勝数-負数)
1	履歴5戦略	166836	14201	13963	152635
2	履歴4戦略	164216	14729	16055	149487
3	履歴3戦略	163833	16168	14999	147665
4	履歴2戦略	157915	16261	20824	141654
5	でたため戦略	64687	65368	64945	-681
6	t t p 戦略	48668	54914	91418	-6246
7	g g t 戦略	48684	54955	91361	-6271
8	p p g 戦略	48844	56160	89996	-7316
9	t g t 戦略	44995	57449	92556	-12454
10	p t p 戦略	50592	63090	81318	-12498

表 2: 1 試合に 5000 回のじゃんけんを行った結果

順位	戦略	勝ち	負け	引き分け	総合得点 (勝数-負数)
1	履歴5戦略	335647	27316	27037	308331
2	履歴4戦略	329728	29205	31067	300523
3	履歴3戦略	328864	31660	29476	297204
4	履歴2戦略	316508	32199	41293	284309
5	でたため戦略	130139	130406	129455	-267
6	g g t 戦略	97290	109996	182714	-12706
7	t t p 戦略	97222	109933	182845	-12711
8	p p g 戦略	97536	112371	180093	-14835
9	p t p 戦略	101295	126181	162524	-24886
10	g p g 戦略	101244	126160	162596	-24916

表 3: 1 試合に 10000 回のじゃんけんを行った結果

1 試合の繰り返し回数を変えて、試合を行なった結果、どの場合においても履歴学習型戦略が上位を独占することになった。履歴学習型戦略のなかでも、履歴の長さによって順位に差があるものの、それ以外の戦略との得点の差が大きい。これは、用意した戦略の中では、履歴学習型戦略が他の全ての戦略に対して、勝ち越しているためである。この結果から考えられることは、履歴戦略は最終的に相手の過去の全ての対戦内容を履歴として用いているのに対して、他の戦略においては、ただか数手前の手について、その勝負の結果や、内容だけしか用いていない。そのために、相手の過去の手の情報を用いている場合と、そうでない場合の違いが結果にあらわれたものと考えられる。相手の過去の手をどれくらいまで用いるか、すなわち記憶しておくかによって、対戦結果にかなりの差がでた。

対戦結果から、それぞれの戦略についての関係が、次の図7ように示される。ここでは、履歴学習型戦略の4種、ものまね戦略4種、場合分け戦略27種、でたらめ戦略1種、一筋戦略3種を一つのグループとして、それぞれのグループの間で対戦を行った。その結果、矢印のものとグループに対して、先のグループが、数字の数だけ勝ち越している。

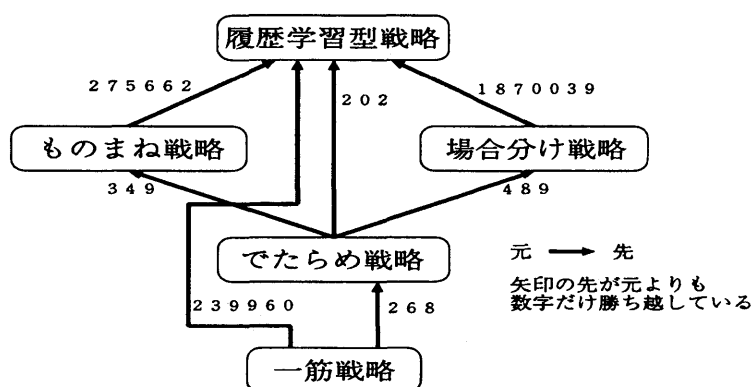


図7: 各戦略グループ間の優劣関係

でたらめ戦略、一筋戦略、場合分け戦略、ものまね戦略、履歴学習型戦略の戦略の組ごとに、対戦の結果をみると、履歴学習型戦略が他の戦略に対して勝ち越していることがわかる。その履歴学習型戦略は、でたらめ戦略のような戦略には、相手の手が予測しにくいために、あまり勝ち越せなかった。履歴学習型戦略以外の関係については、一筋戦略がでたらめ戦略に負け越している、ものまね戦略や、場合分け戦略に対しては、引き分けであり、一筋戦略が一番結果が悪かった。これは、相手の手の情報を用いず、ただ同じ手を出し続けるため当然である。ものまね戦略と、場合分け戦略とは引き分けである。これは、場合分け戦略の出す手を、ものまね戦略がまねるので、ずっと負け続ける場合や、勝ち続ける場合があるためであると考えられる。このふたつは、でたらめ戦略に対して引き分けではないけれど、少しだけ勝ち越している。これは、でたらめ戦略は、相手の手に関係なく手を決めていたために、そのために少しだけものまね戦略や、場合分け戦略に負け越したと考えられる。これは、履歴学習型戦略は、でたらめ戦略のように、ランダムに出す手を選択して、まんべんなく3種類の手を出すような戦略には、予測ができないからである。

履歴学習型戦略は、相手の手の系列の全てを記憶しているために、数手前の相手の手の系列しか用いていない、ものまね戦略、場合分け戦略に対して勝ち越すことができた。ものまね戦略は、相手の手の系列を使っているが、その時の勝負の結果は考えていない。場合分け戦略の場合は、相手との1回の勝負の結果は考えているが、その場合の相手の手については、考えていない。どちらも、数手分しか相手の手の系列を用いていないために、相手に勝ち越せなかった。相手の手の系列、相手との1回ごとの勝負の結果の利用方法の違いにより、各戦略の間の勝負の結果の差に現われた。

3.2 履歴学習型戦略同士の結果

前節で示した結果より、他の戦略に勝ち越して、高得点をあげた履歴学習型戦略が有効な戦略と考えられる。そこで、ここでは履歴学習型戦略同士の対戦についてさらに検討する。

1試合につき10000回ずつじゃんけんを繰り返した場合の、履歴の長さが2~5の4種類の履歴学習型戦略同士の結果を以下の図に示す。図8は、最初に出す手を同じにした場合の結果であり、図9は、最初に出す手が違うものについての結果である。

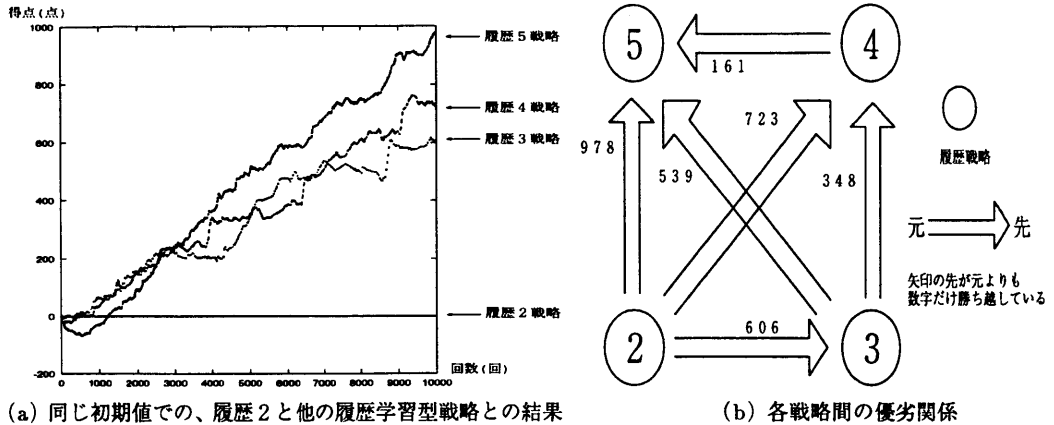


図8:同じ初期値での、履歴2と他の履歴学習型戦略との結果 (a)と各戦略間の優劣関係 (b)

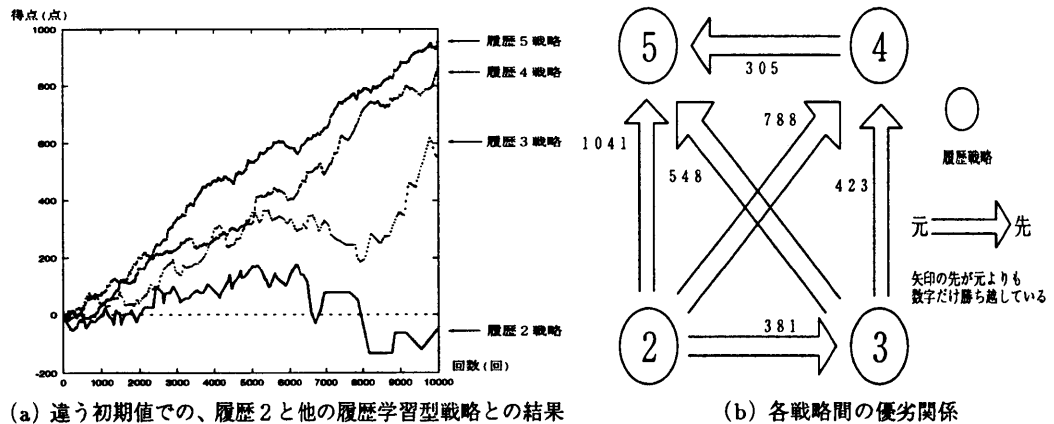


図9:違う初期値での、履歴2と他の履歴学習型戦略との結果 (a)と各戦略間の優劣関係 (b)

履歴学習型戦略同士の対戦をするとき、履歴を用いる長さ、初期値が同じもの同士の戦略は、結果はずっと引き分けが続いてしまう。これは、同じ手の履歴になってしまうためである。そこで、最初に出す手を変えて行なったところ、同じ長さの履歴を用いた戦略同士の対戦であっても、最初に出す手によってずっと引き分けが続くようなことはなくなった。このことから、履歴学習型戦略は履歴を用いるために相手の最初に出す手によって、勝負の結果も変わってくる。履歴の長さによる比較のために、履歴のパターンが少なくとも1度は出るような回数までじゃんけんを繰り返した結果、履歴が長い戦略が短いものよりも良い結果がえられた。

履歴学習型の戦略同士の対戦では、履歴が長い方が短いものより強いという結果になった。しかし、履歴の用いる長さを長くすればするほど、それだけ繰り返す回数を多くする必要が生じ、対戦回数によっては必ずしも履歴の長さが長いものが有利とはいえない。そして、履歴に偏りがでないようにするため

に、初期値についても考えなければならない。実験で用いた履歴学習型の戦略は、履歴2～5の長さのもので、試合の一番最初に出す手もあらかじめ設定してあるものを使っていたが、履歴の長さ、最初に出す手といったものを動的に変更できるような戦略について検討が必要である。

4 おわりに

本研究では、繰り返し対戦型じゃんけんゲームというものを設定することで、じゃんけんに戦略を取り入れることを考え、そこで有効な戦略の獲得と、その戦略の特徴についての検討を行った。繰り返し対戦型ゲームを行なった結果、履歴学習型の戦略が有効であった。このことは履歴学習型戦略は、相手の全ての手の履歴を用いて相手の手を予測できているために、他の用意した戦略よりも良い結果が得られたと考えることができる。

3章の対戦の結果からみられるように、繰り返し対戦型じゃんけんゲームにおいて、戦略によっての差がある。それは、用意した戦略により、相手の手の履歴を利用するものと、しないものがあったためである。なかでも、履歴学習型戦略は、相手の全ての手の履歴を用いているのに比べ、他の戦略は数手前の履歴だけしか用いていないという差が今回の様な結果になった要因である。また、用意した戦略の種類も数も少なかったために、かたよりがみられた。相手の手の履歴、1回ごとの勝負の結果などを用いる方法により、今回のような、戦略の間の差が現われた。このようなことから、繰り返し対戦型じゃんけんゲームにおいて、相手の手を用いて、次の手を予測するような戦略が有効であると考えられる。

今後は、対戦型ゲームにおける戦略知識の獲得のためには、繰り返し対戦型じゃんけんゲームの結果から、履歴学習型戦略のような相手の手を学習して、次の手を予測するような戦略も獲得できる戦略表現が必要であると考えられから、今回の実験で用いた履歴学習型戦略は、履歴の長さ、試合の最初に出す手の設定といったものをあらかじめある値に設定してあり、どのような値の組合せが有効であるかははっきりとはわからない。また、相手の手の系列、勝負の結果の用いかたにより、様々な戦略が考えられる。そういった戦略を表現するためには、履歴学習型戦略や一筋戦略、その他の戦略を自己生成のモデルを用いて表現することを目指す。相手の手を学習して、それから相手の手の予測をするための方法としては、GAなどを用いるなどが考えられる [4]。今後としては、繰り返し対戦型じゃんけんゲームにおいて、戦略知識の獲得を試みるのが課題である。

参考文献

- [1] モートン・D・デービス：「ゲームの理論」, 講談社 (1994)
- [2] 西山 賢一：「勝つためのゲーム理論」, 講談社 (1995)
- [3] 牧野 泰裕, 西野順二, 小高 知宏, 小倉 久和：「対戦型ゲームにおける戦略知識獲得の試み」, 平成8年度電気関係学会北陸支部連合大会講演論文集 p393, (1996)
- [4] 久保 長徳, 西野 順二, 小高 知宏, 小倉 久和：「遺伝的アルゴリズムによる知識獲得」, 平成7年度福井大学工学部情報工学科 修士論文, (1996)

