

ミニマックス戦略と相手モデル探索の混合戦略： tictactoe によるセミランダムと勝負手探索を題材として

梶原 羊一郎

飯田 弘之

cs6501@cs.inf.shizuoka.ac.jp iida@cs.inf.shizuoka.ac.jp

静岡大学情報学部

概要

本稿はミニマックス戦略と相手モデル探索の混合戦略を論じる。そのために、tictactoe のゲームを用いて、セミランダムプレイとそれに対する勝負手探索の融合を検討する。セミランダムプレイの手法は先読みが導入されたランダムプレイであり、ゲームの終了状態だけが考慮される。この手法により、実力の異なる様々なプレイヤーのモデル化が可能になる。次に、セミランダムプレイに対する勝負手探索を提案する。tictactoe を用いた自動プレイによる実験で、提案する勝負手探索の有効性を確認する。同時に、勝負手探索を限られたタイミングで用いると仮定した状況で、どのように戦略を制御すべきかを検討する。これらの結果が、一般の二人零和ゲームにおいて、ミニマックス戦略と相手モデル探索の混合戦略を構築する際に、有意義な示唆を与える。

A Mixed Strategy of Minimax Strategy and Opponent-Model Search: a case study using a semi-random play and its speculative play on tictactoe

Yoichiro Kajihara and Hiroyuki Iida

Department of Computer Science, Shizuoka University

ABSTRACT

This paper explores the control of strategies for game playing: the minimax strategy and the opponent-model search or more generally a speculative strategy. It first describes a search strategy for game playing called *semi-random self-play* that is a search strategy using random decisions and a look-ahead search for end goals only, by which a model of players with different strength may be obtained. Some experiments of semi-random self-play are performed using the simple two-person perfect-information game TICTACTOE. In the experiments a computer program plays against itself many games using semi-random self-play with various search depths. We then explore the use of speculative play against semi-random self-play. This search strategy is a variant of opponent-model search, and is applied against a player using semi-random self-play. We implemented such a speculative play in a simple TICTACTOE program. Experiments performed confirm the effectiveness of the proposed speculative strategy. Moreover, we consider which situation to apply the speculative strategy if one player is allowed to use such speculative strategy on some limited times in a game.

1 セミランダムプレイとは？

セミランダムプレイは、ランダムプレイをより一般化したアイデアである [6][7].

定義 1 セミランダムプレイとは、与えられた深さの先読み探索を行う探索戦略であり、以下の **R1** および **R2** に従う.

R1: 与えられた局面での全可能手を生成し候補手リストへ入れよ. 次に、与えられた深さのゲーム木を作る. このとき、相手の応手に関係なく必然的に勝てる指し手が見つければ、その指し手を選択せよ. そうでなければ、**R2** へ.

R2: 相手の適切な選択により必然的に負けとなる指し手の全部を候補手リストから除外せよ. これによって、候補手リストが空集合となるならば、除外する前のリストからランダムに指し手を選択せよ. そうでなければ、いま得られた候補手リストからランダムに指し手を選択せよ.

探索の深さが 0 の場合、セミランダムプレイはランダムプレイと等価である. セミランダムプレイのルール (**R1** と **R2**) に基づき、先読み深さの関数としてレベルの異なる様々なプレイヤーのモデル化が可能になる. このようにしてモデル化されたプレイヤーを先読みの深さに対応させて、**P1**, **P2**, あるいは一般的に **Pi** と表す. 先読みの深さとレーティングの関係を示唆するコンピュータチェスでの実験 [8] が示すように、探索の深さが大きくなるにしたがって、セミランダムプレイのレベルは向上する.

1.1 セミランダムプレイの実験

TicTacToe のコンピュータプログラムを用いたセミランダムプレイの実験を以下のように行った.

- コンピュータプログラムは、標準的なミニマックスアルゴリズムを用いて実装される. 局面評価は、ゲームの終了局面 (勝ち, 負け, 引き分け, あるいは不明) を判定するだけである.
- 各コンピュータプログラム **Pi** は、与えられた探索の深さ i をもって、**R1** と **R2** の規則にしたがって先読み探索を実行する. このとき、局面評価は上述したものだけである.
- ランダムな選択のために標準的な C の関数 `rand()` を用いる.

- セミランダムによる自動対戦を 10,000 回試行する. すなわち、**Pi** 対 **Pj** ($0 \leq i, j \leq 6$) の組合せだけ対戦を実行する.
- プログラムは C (Microsoft Visual C++) で記述され、Windows 98 上で実行した.

1.2 実験結果と考察

実験結果を表 1 に示す. 参考のため、先読みの深さが同じプレイヤー同士による対戦の結果を図 1) に示す.

これらの結果から、次のような知見を得る.

- **P6** はパーフェクトプレイヤーで、**P0** は超初心者にそれぞれ相当する.
- 先読みの深さが大きいほどセミランダムプレイヤーの実力は向上する. ただし、いくつかの例外もある. 例えば表からわかるように、**P2** 対 **P5** では 27.30% 負けだが、**P2** 対 **P6** では 16.40% の負けであり、**P2** との対戦結果からみる限りでは、**P6** の方が **P5** よりも強いように思える. われわれはこの現象を次のように解釈する. **P6** はその試合の結果に関する知識を **P5** より多く持っているため、リスク (負ける危険性) のある指し手を選択しない. ゆえに、**P2** のミスを期待するような指し手を好んで選択しようとしな. 実際、負け率 (勝率 + 引き分け率) を強さの指標とすると、例外なく、先読みの深さが大きいほど強い.
- 図 1 から、試合結果の割合 (勝ち, 引き分け) は探索の深さの関数として変動している. われわれはこの現象を次のように解釈する. 探索プロセスの性質として、先読みの深さが奇数の場合、プログラム **Pi** は、葉局面で自分の勝ちを判断する. 一方、先読みの深さが偶数の場合、プログラム **Pi** は、葉局面で自分の負けを判断する. ゆえに、探索の深さが奇数から偶数に変化するときに勝率は上がり、偶数から奇数に変化するときに勝率は下がる.

これらの分析は、意図したように、セミランダムプレイによってレベルの異なるプレイヤーのモデル化が達成できたことを示す.

2 セミランダムプレイ相手の勝負手探索

先に二人ゲームを題材として相手モデル探索を提案した [3, 2]. その後、相手モデルのファクターと

表 1: TicTAcToE を用いたセミランダムプレイの結果 (Pi と Pj の対戦).

W\B	P0	P1	P2	P3	P4	P5	P6
P0	59.06	39.88	8.77	9.00	4.64	4.33	0
	12.74	8.15	22.20	19.98	21.15	20.32	22.63
	28.20	51.97	69.03	71.02	74.21	75.35	77.37
P1	81.59	67.81	13.84	14.36	7.58	7.13	0
	6.18	4.45	20.37	17.86	19.98	19.32	22.63
	12.23	27.74	65.79	67.78	72.44	73.55	77.37
P2	89.28	88.68	31.45	31.13	18.40	16.73	0
	9.07	8.78	51.26	44.82	58.60	55.97	83.60
	1.65	2.54	17.29	24.05	23.00	27.30	16.40
P3	93.93	93.55	52.61	52.52	30.80	29.78	0
	4.95	4.79	34.59	30.45	49.19	46.60	83.60
	1.12	1.66	12.80	17.03	20.01	23.62	16.40
P4	93.97	93.56	52.67	52.24	30.43	29.68	0
	5.09	5.09	37.32	34.37	54.49	50.80	88.67
	0.94	1.35	10.01	13.39	15.08	19.52	11.33
P5	96.51	96.47	76.62	76.53	67.74	67.67	0
	3.09	2.98	20.62	20.30	28.56	26.61	88.67
	0.40	0.55	2.76	3.17	3.70	5.72	11.33
P6	96.60	96.60	77.62	77.62	67.80	67.80	0
	3.40	3.40	22.38	22.38	32.20	32.20	100
	0	0	0	0	0	0	0

それぞれの数字は、試合結果の比率を表す。上段が勝ち、中段が引き分け、下段が負けである。‘W’は先手、‘B’は後手を示す。

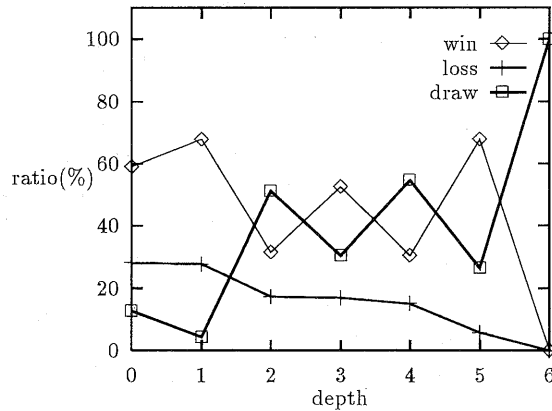


図 1: 先読みの深さが同じセミランダムプレイヤー同士の対戦結果

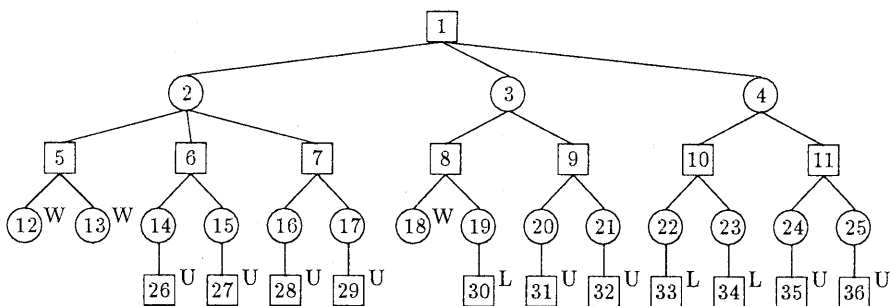


図 2: セミランダム勝負手探索の例.

‘W’, ‘L’, and ‘U’ はそれぞれ、先手の勝ち、負け、引き分けの局面を表す.

して、評価関数と先読みの深さの両方を考慮した (D, d) -勝負手探索が実行される. すなわち、プレイヤー X OM search を提案した. これは相手モデルに基づく勝負手探索に他ならない [1]. 相手モデル探索や (D, d) -OM search は、相手プレイヤーに関する十分な知識 (評価関数、先読み深さなど) を持った場合の勝負手探索とみなすことができる. しかし、実際のゲームプレイでは、正確な相手モデルを獲得できることはまず期待できない. 相手に関する仮想的なモデルを持つに過ぎない. その結果としてリスクを負うことになる.

セミランダムプレイに対する勝負手探索を検討する. 相手がセミランダムプレイを用いるならば、まさに実戦のように、十分な相手モデルを獲得できない状況 (相手はランダムに意志決定する) である. このような状況でも相手モデル探索のような勝負手探索は可能だろうか.

本節で、われわれはセミランダムプレイを行う相手のミスに期待する勝負手探索を提案する. ただし、相手の先読みの深さは自分より小さいと仮定する. それから、セミランダムプレイに対する勝負手探索を実装し、提案する勝負手探索とセミランダムプレイの自動対戦の実験について述べる.

2.1 勝負手探索の設計

先行研究 [7] において、セミランダムプレイに対する勝負手探索を提案した. その勝負手探索は、ゲーム木中の全ノードに対する自分の勝ち局面ノードの割合を最大化することで実現した. 本稿では、先に提案した勝負手探索より洗練されたバージョンを提案する. 以下にその定義を述べる.

定義 2 勝負手探索を実行するプレイヤー X の先読みの深さを d_x 、セミランダムプレイを実行するプレイヤー Y の先読みの深さを d_y とおく. このとき、 $d_x \geq d_y + 2$ の条件下で、セミランダムプレイに対する

は、ゲーム木の各部分木のルートノードでの自分の勝ちの期待値を最大化する指し手を選択する.

各部分木のルートでのプレイヤー X が勝ちになる期待値は、プレイヤー X が勝つ確率を計算することで得られる. このようにして得られる期待値は、一般にはプレイヤー X とプレイヤー Y の観点からは必ずしも等しくはないことに注意すべきである. つまり、先読みの深さの差を利用して、プレイヤー Y のミスを最大限に期待するのである.

図 2 に上述した勝負手探索の例を示す. この例では、先手プレイヤーは先読みの深さが 4 であり、勝負手探索を実行する. 一方、後手は、先読みの深さが 1 のセミランダムプレイを行う. 図のルート局面は先手の手番である. 後手は、深さ 1 のノードにある局面 (#2, 3, 4) で、深さ 1 の先読みが可能であるが、先手が図のゲーム木のすべてを先読みできる.

図 2 のゲーム木では、左側の部分木には、6 個の全葉ノード (#12, 13, 26, 27, 28, 29) の中に 2 個の勝ちノード (#12, #13) がある. 真中の部分木には、4 個の全葉ノードの中に 1 個の勝ちノードがある. そして、右側の部分木には、4 個の全葉ノードの中に勝ちノードは存在しない. 左側の部分木での、全葉ノードに対する先手勝ちノードの割合は、他の部分木より大きいので、先行研究 GPW99 で提案した勝負手探索によれば、先手は 1 → 2 の指し手を選択するはずである. 一方、各部分木のルートノード (#2, 3, 4) における先手勝ちの期待値を考慮した本稿で提案する新たな勝負手探索では、先手は指し手 1 → 3 を選択する. なぜなら、ノード #2 における先手勝ちの期待値は $1/3$ であるのに対し、ノード #3 では $1/2$ 、ノード #4 では 0 だからである.

2.2 勝負手探索の実験・結果

本稿で新たに提案した勝負手探索を用いて、セミランダムプレイと対戦試合を行った。先手と後手、それぞれ10,000試合ずつ、合計20,000試合を行った。先手が勝負手探索を用いた結果を表2に、そして、後手が勝負手探索を用いた結果を表3に示す。ここで、 P_i が勝負手探索を行えるのは、条件 $i \geq j+2$ を満たすときだけであり、もし $j \geq i+2$ であれば、 P_j が勝負手探索を実行する。それ以外の場合には、各プレイヤーはセミランダムプレイを行う。

2.3 戦略の融合を考える

表4に、セミランダムプレイと勝負手探索(可能な場合)によるそれぞれの勝率を示す。

表4: 勝負手探索を最良のタイミングで用いた場合の結果(勝率の増加分)

	P0	P1	P2	P3	P4	P5
P2	1.99 1.39					
P3	5.50 15.86	0.85 1.92				
P4	4.91 13.46	1.72 7.63	7.50 1.85			
P5	1.97 10.86	2.39 4.21	16.39 13.39	13.02 7.84		
P6	2.17 8.93	2.01 1.02	15.75 10.76	11.74 3.61	19.26 3.78	
P7	2.92 11.11	2.01 1.46	15.44 29.72	12.07 23.56	18.77 27.32	19.72 3.24

各数字は勝率を表す。上段は先手番で勝負手探索を用いた場合、下段は後手番で勝負手探索を用いた場合である。

これらの結果(表2~4)からわれわれは次のような知見を得る。

- 本稿で提案する勝負手探索によって、セミランダムプレイよりも平均的に勝率が高くなる。
- 自分の手番時に、常に勝負手探索を実行する('111')のは得策ではない。主な理由として以下の2つがある。勝負手探索だけを用いた結果が、混合戦略の組合せの中で最良となるケースは、先手の場合(表2参照)で1/3、後手の場合(表3参照)で約1/2である。また、セミランダムプレイに比べて勝負手探索は計算コストが高い。

- 勝負手探索を用いるタイミング(タイプの選択、戦略の制御)を間違えると、セミランダムプレイよりも劣る結果となる。
- 混合戦略のタイプ'001'が相対的によくない成績である。相手よりも先読みの深さが大きいという利点を生かせない、と解釈する。
- 混合戦略のタイプ'1xx'が'0yy'と比べて相対的によい成績である。つまり、できるだけ早い段階(序盤)で、勝負手探索を用いることが有効である、と解釈する。

3 まとめ

セミランダムプレイに対する勝負手探索を提案し、その有効性をTICTACTOEを用いた実験で確認した。ただし、勝負手探索特有のリスク[4]により例外となった場合もある。同時に、セミランダムプレイと勝負手探索の混合戦略についても検討した。それぞれが、ミニマックス探索および相手モデル探索に相当すると仮定すると、混合戦略(戦略の制御)についていくつかの有益な示唆が得られた。

今後は、TICTACTOE以外のゲームでも本稿で行ったような実験を行う予定である。

参考文献

- [1] Gao, X., Iida, H., Uiterwijk, J.W.H.M., and Herik, H.J. van den (1999). A Speculative Strategy. In Proc. Internat. Conf. on Computers and Games, CG'98 (Eds. H.J. van den Herik and H. Iida), *Lecture Notes in Computer Science*, vol.1558, Springer, Heidelberg, pp.74-93.
- [2] Iida, H., Uiterwijk, J.W.H.M., Herik, H.J. van den, and Herschberg, I.S. (1993). Potential Applications of Opponent-Model Search; part 1: the domain of applicability. *ICCA Journal*, 16(4), pp.201-208.
- [3] Iida, H., Uiterwijk, J.W.H.M., and Herik, H.J. van den (1993). Opponent-Model Search. Technical Report #CS 93-03, Department of Computer Science, Universiteit Maastricht, Maastricht, The Netherlands.
- [4] Iida, H., Uiterwijk, J.W.H.M., Herik, H.J. van den, and Herschberg, I.S. (1994). Potential Applications of Opponent-Model Search; part 2: risks and strategies. *ICCA Journal*, 17(1), pp.10-14.

表 2: セミランダムプレイと勝負手探索の混合戦略の結果 (先手番).

W\B	Type	P0	P1	P2	P3	P4	P5	Average
P2	000	89.28						89.28
	100	90.85						90.85
	010	89.43						89.43
	001	88.55						88.55
	110	90.00						90.00
	101	89.83						89.83
	011	88.46						88.46
	111	91.27						91.27
P3	000	93.68	93.93					93.81
	100	94.61	94.66					94.64
	010	97.08	93.30					95.19
	001	93.88	93.69					93.79
	110	98.85	92.60					95.73
	101	95.49	94.78					95.14
	011	97.57	93.60					95.59
	111	99.18	93.42					96.30
P4	000	93.96	93.53	52.30				79.93
	100	95.04	95.25	59.80				83.36
	010	97.40	93.80	56.04				82.41
	001	94.18	93.48	53.01				80.22
	110	98.72	93.29	39.82				77.28
	101	95.71	94.77	59.26				83.25
	011	97.57	93.88	55.56				82.34
	111	98.87	93.19	38.83				76.96
P5	000	96.97	96.38	76.52	76.91			86.70
	100	98.80	98.77	89.16	89.27			94.00
	010	98.62	96.34	87.78	76.67			89.85
	001	96.99	94.68	76.68	76.08			86.11
	110	98.71	98.04	92.91	87.41			94.27
	101	98.91	98.52	89.54	89.93			94.23
	011	98.85	92.22	88.02	77.36			89.11
	111	98.94	96.97	92.69	87.42			94.01
P6	000	96.83	96.82	76.62	77.74	68.47		83.30
	100	98.93	98.99	89.80	89.46	87.40		92.92
	010	98.63	96.57	87.16	77.34	67.72		85.48
	001	97.57	94.95	77.95	77.29	68.13		83.18
	110	98.86	98.23	92.37	87.49	87.45		92.88
	101	99.00	98.32	89.23	89.48	87.48		92.70
	011	98.81	92.49	88.24	76.86	68.01		84.88
	111	98.93	96.76	92.37	87.74	87.73		92.71
P7	000	96.67	96.90	77.96	77.34	68.66	68.00	80.92
	100	98.85	98.91	89.94	89.41	87.37	87.40	91.98
	010	98.57	96.32	86.35	79.48	73.21	66.95	83.48
	001	97.50	94.84	77.87	77.11	69.06	68.78	80.86
	110	99.59	98.14	93.04	87.27	87.34	87.56	92.16
	101	98.99	98.46	89.63	89.13	87.33	87.72	91.88
	011	98.81	92.52	89.15	79.70	72.17	67.57	83.32
	111	99.51	96.74	92.59	87.32	87.43	87.29	91.81

各数字は勝率を表す。後手はセミランダムプレイである。タイプ'000'とは先手が勝負手探索をしない戦略を表す。同様に、'100'は最初の手番のときだけ、'110'は最初と次の手番の2回、勝負手探索を行う戦略を表す。イタリックで表された数字は混合戦略で最も良い結果を得た場合である。

表 3: セミランダムプレイと勝負手探索の混合戦略の結果 (後手番).

B\W	Type	P0	P1	P2	P3	P4	P5	Average
P2	000	69.03						69.03
	100	70.14						70.14
	010	68.12						68.12
	001	67.93						67.93
	110	70.35						70.35
	101	69.94						69.94
	011	69.32						69.32
	111	70.42						70.42
P3	000	70.76	68.07					69.42
	100	72.81	69.31					71.06
	010	79.76	67.03					73.40
	001	78.24	67.87					73.06
	110	83.81	68.60					76.21
	101	78.75	69.24					74.00
	011	83.22	67.73					75.48
	111	86.62	69.95					78.29
P4	000	73.67	73.03	23.73				56.81
	100	75.98	73.58	24.80				58.12
	010	78.76	72.22	23.99				58.32
	001	80.89	73.18	23.58				59.22
	110	83.54	79.02	22.73				61.76
	101	81.01	76.24	25.58				60.94
	011	83.51	72.91	23.45				59.96
	111	87.13	80.66	22.58				63.46
P5	000	75.88	73.28	26.89	22.95			49.75
	100	76.80	75.37	32.95	30.38			53.88
	010	80.91	64.23	34.41	23.73			50.82
	001	81.66	73.86	26.23	23.09			51.21
	110	83.79	70.26	39.73	30.03			55.95
	101	82.55	77.49	33.23	30.79			56.02
	011	85.06	65.09	34.39	23.31			51.96
	111	86.74	72.05	40.28	30.29			57.34
P6	000	77.69	77.63	16.79	16.48	10.73		39.86
	100	76.30	76.18	19.88	20.09	14.43		41.38
	010	84.16	72.31	23.42	16.49	11.17		41.51
	001	82.99	78.65	16.05	16.58	10.84		41.02
	110	84.25	71.78	27.55	19.73	13.67		43.40
	101	81.43	75.83	20.29	19.83	14.51		42.38
	011	86.62	72.87	23.17	16.57	11.44		42.13
	111	86.30	73.49	27.54	19.50	13.97		44.16
P7	000	78.22	78.56	15.85	16.68	11.07	11.11	35.25
	100	81.13	80.02	42.53	38.77	38.99	13.72	49.09
	010	83.62	71.41	22.66	16.44	11.23	11.08	36.07
	001	83.57	78.25	16.08	16.80	11.07	11.83	36.27
	110	86.88	77.53	44.97	39.30	37.99	14.20	50.15
	101	84.25	79.78	41.05	39.60	38.29	14.35	49.55
	011	86.68	72.72	23.88	16.64	11.15	10.71	36.96
	111	89.33	77.33	45.57	40.24	38.36	14.02	50.81

表 2 の説明を参照.

- [5] Iida, H., Yoshimura, J., Morita, K., and Uiterwijk, J.W.H.M. (1999). Retrograde Analysis of the KGK Endgame in Shogi: Its Implications for Ancient Heian Shogi. In: Proc. Internat. Conf. on Computers and Games, CG'98 (eds. H.J. van den Herik and H. Iida), *Lecture Notes in Computer Science*, Vol. 1558, pp. 318–340. Springer-Verlag, Heidelberg.
- [6] Kajihara Y., Sakuta M., and Iida H. (1999). Semi-Random Self-Play in Game Playing. *Proceedings of Joint International Conference on Advanced Science and Technology (JICAST'99)*, Hangzhou, China, pp.205–208.
- [7] Kajihara Y., Sakuta M., Iida H, Uiterwijk, J.W.H.M., and Herik, H.J. van den (1999). A Speculative Play against Semi-Random Self-Play. *Proceedings of Game Programming Workshop in Japan '99*, Hakone, Japan. pp.109–112.
- [8] Newborn, M. (1985). A Hypothesis Concerning the Strength of Chess Programs. *ICCA Journal*, 8(4), pp.209–215.
- [9] 佐々木宣介, 橋本剛, 梶原羊一郎, 飯田弘之 (1999). チェスライクゲームにおける普遍的指標, 情報処理学会ゲーム情報学研究会 99-GI-1, pp.91–98.