

初期局面の有利不利に適応した レーティング手法の検証

高尾 海州^{1,a)} 松崎 公紀^{1,b)}

概要: ゲームやスポーツにおいて、プレイヤーの強さはレーティングによって数値化される。オセロやチェス、将棋などを代表とした運に左右されないゲームでは、イロレーティングという手法が広く用いられている。将棋の手合割のように、プレイヤーの強さに差がある場合にハンディキャップを与えることはよくあるが、その場合には既存のイロレーティングでは、適切なレートを出算できない可能性がある。そこで、本研究では2014年に森田らが提案した初期局面の有利不利に適応したレーティング手法をオセロを用いて検証する。オセロの途中盤面をいくつか用意し、シミュレーションによって盤面ごとの勝率を求め、その数値を森田のレーティング手法に用いた。実験では、既存のイロレーティングと森田のレーティングにより、初期局面の有利不利がある場合のレートを求め、本来のレートと比較する。本論文では、検証のための実験の設計と結果について報告する。

キーワード: オセロ, レーティング

1. はじめに

ゲームやスポーツなどの勝敗を競う分野において、プレイヤーの強さはレーティングによって数値化される。チェスや将棋などのゲームでは、イロレーティング [1] というレーティングアルゴリズムが広く使われている。イロレーティングは、2つのプレイヤーの対戦結果とレートから対戦後のレートを求める。勝ったプレイヤーのレートが相手のレートより小さいほどレートの増減量は大きくなる特徴がある。

トランプを使ったゲームやコンピュータを使うゲームでは、ランダム性があることにより初期局面に有利不利がある。イロレーティングは、初期局

面が対等であることを前提としているため、そのようなゲームにおいて単純にイロレーティングを用いると適切なレートを算出できない可能性がある。

そこで、本研究では2014年に森田ら [2] が提案したレート計算式を用いて、初期局面の有利不利に適応したレーティング手法をオセロを使って検証する。

森田が提案したレート式は、イロレーティングのレーティングアルゴリズムに初期局面での勝率を加えることでレートの変化量を補正する。

実験では、オセロの試合を途中で止めた途中盤面を初期局面とすることで初期局面の有利不利を再現する。いくつかの手法で、途中盤面ごとの勝率を求める。どの手法が有効なのかを検証するために、途中盤面の勝率から得られたレートと既存のイロレーティングから得られたレートを、本来のプレイヤーのレートと比較する。本論文では、検証のため

¹ 高知工科大学

^{a)} 190334v@ugs.kochi-tech.ac.jp

^{b)} matsuzaki.kiminori@kochi-tech.ac.jp

の実験の設計と結果について報告する。

本論文の貢献は大きく以下の3つである。

- 途中盤面からプレイするオセロにおいて、初期盤面勝率の存在について確認した(第4.2節)。
- 初期局面にランダムに有利不利を与えたゲームにおいて、森田らの手法によってより適切なレートが計算できる場合があることを確認した(第5.1節)。
- より強いプレイヤーに対して不利な初期局面を与えるゲームにおいて、森田らの手法によって十分適切なレートが計算できる場合があることを確認した(第5.2節)。

本論文の構成は以下のとおりである。第2節では本論文で使用するイロレーティングとそれを拡張したレーティングの計算方法について解説する。第3節では実験で使用するプレイヤーについて解説しそのプレイヤーの基準となるレートを求める。第4節では初期盤面勝率の存在をテスト盤面により確認し、実験で使用する途中盤面について初期盤面勝率の分布を示す。第5節では途中盤面の勝率から得られたレートとイロレーティングから得られたレートを、本来のレートと比較する。さらに、強いプレイヤーに対して不利な盤面を与えた場合についても検証し、考察する。第6節で関連論文を紹介し第7節で本論文をまとめる。

2. イロレーティングとその拡張

2.1 イロレーティング

チェスなどの2人ゲームにおいて、プレイヤーの実力を算出する計算手法のひとつにイロレーティング [1] がある。

イロレーティングでは、プレイヤー間の対戦結果によりレートを更新する。プレイヤーAのレートが R_a 、プレイヤーBのレートが R_b であるとする。このとき、プレイヤーAのプレイヤーBに対する期待勝率 e_{ab} を

$$e_{ab} = \frac{1}{1 + 10^{-(R_a - R_b)/400}}$$

とする(図1)。プレイヤーAとプレイヤーBの対戦の結果 s_{ab} (ただし、プレイヤーAが勝ったときは $s_{ab} = 1$ 、プレイヤーAが負けたときは $s_{ab} = 0$) に

より、プレイヤーAの新たなレート R'_a は

$$R'_a = R_a + K(s_{ab} - e_{ab})$$

と計算される。ここで、 K はレートが増減する幅を決定する定数である。本論文を通して $K = 16$ とする。各プレイヤーについて、レートの初期値は1500とする。

例えば、プレイヤーAとBがいずれもレート1500であるとし、対戦の結果プレイヤーAが勝つとすると、プレイヤーAのレートは $1500 \rightarrow 1508$ と増加し、プレイヤーBのレートは $1500 \rightarrow 1492$ と減少する。プレイヤーAのレートが1600、プレイヤーBのレートが1400であるとき、プレイヤーAが勝つとプレイヤーAのレートは $1600 \rightarrow 1601.9$ 、プレイヤーBのレートは $1400 \rightarrow 1398.1$ と変化する。このように、イロレーティングではレートの高いプレイヤーがレートの低いプレイヤーに勝つとレートの増加値が小さくなる。逆に、レートが低いプレイヤーがレートの高いプレイヤーに勝つとレートの増加値が大きくなる。

2.2 初期盤面勝率を考慮に入れた拡張

イロレーティングでは、期待勝率が二人のプレイヤーのレート差のみから決まるという仮定をおいている。プレイヤー間に強さの差がある場合にハンディキャップを与えることはよくあるが、イロレーティングではそのような場合に適切なレートを計算できない。

この問題を解決する方法として、イロレーティングの期待勝率を求める式に初期盤面の有利不利を加味する拡張が森田ら [2] によって提案されている。期待勝率を拡張するにあたって、以下を仮定する。

- 初期盤面に有利不利がない場合には、期待勝率は二人のプレイヤーのレート差によって決まる。
 - 二人のプレイヤーのレート差がない場合には、期待勝率は初期盤面勝率 p_a によって決まる。
- これらの2つの仮定を満たす期待勝率を与える

式のひとつは、

$$e_{ab} = \frac{1}{1 + 10^{-((R_a - R_b)/400 + \log(p_a/(1-p_a)))}}$$

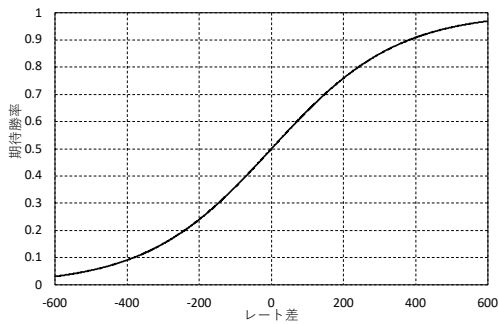


図 1 レート差による期待勝率の推移

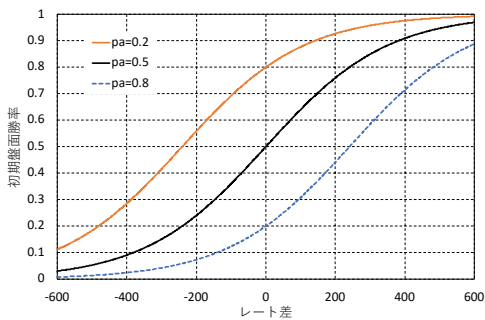


図 2 初期盤面勝率 p_a とレート差による期待勝率

である。この式は、イロレーティングにおける期待勝率の式において、プレイヤー間のレート差に初期盤面勝率 p_a から計算されるレート差 $\log \frac{p_a}{1-p_a}$ を足して計算することを意味している。

図 2 は、初期盤面勝率 p_a がそれぞれ 0.2, 0.5, 0.8 の場合に、プレイヤー間のレート差と期待勝率との関係をグラフにしたものである。

例えば、プレイヤー A のレートが 1600、プレイヤー B のレートが 1400 であり、初期盤面勝率が 0.2 である (B に有利な初期盤面) 状況において、プレイヤー A が勝ったとするとプレイヤー A のレートは $1600 \rightarrow 1608.9$ 、プレイヤー B のレートは $1400 \rightarrow 1391.1$ と変化する。これは、B に有利な初期盤面を与えることで、プレイヤー A の期待勝率が 0.44 とほぼ対等になったことにより、プレイヤー A が勝った場合のレートの変動が大きくなったことによる。

3. プレイヤ

本研究では 9 種類のオセロプレイヤーを用いて実

験を行った。本節では、まず各プレイヤーのアルゴリズムの概略を説明し、イロレーティングによって測ったそれらの強さを示す。

3.1 ランダムプレイヤー

プレイヤー **Random** は、石を置ける場所からランダムに選択する。

3.2 評価関数プレイヤー

プレイヤー **EV** は、塩田 [3] の評価関数を用いて、各盤面において評価値が最大となる手を選択する。

塩田の評価関数では、盤面の評価値は以下の 3 つのパラメータの重み和として計算される。

盤面位置 全てのマスに価値を持たせ、白石が置かれていれば加算、相手石が置かれていれば減算し、その合計値を算出する。

確定石 その後相手に絶対に取られない石のことを確定石と呼ぶ。塩田の評価関数では、外周の 4 辺における確定石のみを数え上げている。

候補数 ある局面において次に自分、もしくは相手が着手可能なマスの数を候補数と呼ぶ。

3.3 モンテカルロプレイヤー

モンテカルロプレイヤーは、互いにランダムに着手して終局まで進めるプレイアウトを用いたプレイヤーである。プレイヤー **MC** は、各局面において合法手のそれぞれについてプレイアウトを 50 回行い、勝率の最も高かった手を選択する。

3.4 モンテカルロ木探索プレイヤー

モンテカルロ木探索 [4] は、モンテカルロ法とゲーム木探索を組み合わせたアルゴリズムである。モンテカルロ木探索のアルゴリズムでは、以下の手順を繰り返し行い、最も評価回数が多かった手を選択する。

選択 ゲーム木の節点において UCB1 値の最も高い子節点を選ぶことを繰り返し、ゲーム木の葉を選択する。

展開 選択された葉に対してプレイアウト回数が閾値を超えている場合、その対応する局面の次の合法手をゲーム木に追加する。

プレイアウト 選択された葉から始めて、互いにランダムに着手して終局まで進めるプレイアウトを行う。

更新 根から葉までの経路上の節点に対して、プレイアウト結果（勝敗）を追加する。

プレイヤー **MCT** は、プレイアウト回数を 100 回、節点の展開の閾値を 10 と設定したモンテカルロ木探索プレイヤーである。

3.5 モンテカルロ木探索+評価関数プレイヤー

限られたプレイアウト回数ではモンテカルロ木探索プレイヤーはそれほど強くない。そこで、優れた評価関数を組み合わせるプレイヤーを強化する。

本研究では、モンテカルロ木探索への評価関数の組合せ方として、Early Payout Termination (EPT) [5] と呼ばれる手法を用いた。これは、プレイアウトをある深さまで打ち切り、その途中局面で評価関数を実行し、その評価値の正負をプレイアウトの勝敗とするものである。

本研究では、オープンソースの強いオセロプレイヤー Zebra [6] の評価関数を用いた。Zebra の評価関数は、11 種のパターンによる局所的局面評価値の和を評価値とする。

以下のとおり、2 通りのパラメータの組合せにより、強さの異なるプレイヤーを作成した。

MCT+EPT2k プレイアウト数 1730, ノード展開の閾値 10 とし、プレイアウトを深さ 1 で打ち切るモンテカルロ木探索+評価関数プレイヤーである。

MCT+EPT2k プレイアウト数 8500, ノード展開の閾値 10 とし、プレイアウトを深さ 1 で打ち切るモンテカルロ木探索+評価関数プレイヤーである。

3.6 アルファベータプレイヤー

指定した探索深さにおいて Zebra の評価関数により評価値を求め、アルファベータ枝刈りを行う minimax 探索を行う。異なる強さのプレイヤーを作るため、探索の深さを変えた以下の 3 通りのプレイヤーを作成した。

AlphaBeta1 深さ 1 まで探索を行うアルファ

表 1 本来のレート (通常の初期盤面から計算したレート)

プレイヤー	レート
Random	716.40
EV	1141.49
MC	1158.96
MCT	1301.01
MCT+EPT2k	1764.97
MCT+EPT8k	2111.30
AlphaBeta1	1415.82
AlphaBeta2	1837.92
AlphaBeta4	2052.13

ベータプレイヤーである。

AlphaBeta2 深さ 2 まで探索を行うアルファベータプレイヤーである。

AlphaBeta4 深さ 4 まで探索を行うアルファベータプレイヤーである。

3.7 各プレイヤーのレート

作成した 9 種類のプレイヤーの強さを、オセロの通常の初期盤面から対戦した結果をもとに、イロレーティングにより評価した。

対戦順はランダムに決定し、全ての異なるプレイヤー間で 100 戦行った。対戦順および勝敗によるレートのゆれが生じるため、対戦順を 40 通り生成した上で平均を求めた。このようにして求めた各プレイヤーのレート平均値を表 1 に示す。本研究では、表 1 のレートを本来のレートと呼ぶこととする。

表 1 に示すとおり、プレイヤー EV と MC, MCT と AlphaBeta1, MCT+EPT2k と AlphaBeta2, および MCT+EPT8k と AlphaBeta4 がそれぞれほぼ同じ強さとなっている。

4. テスト局面と初期盤面勝率の計算

本研究では、初期局面に有利不利がある状況において、本来のレート (表 1) に近いレートを算出する手法を見つけ検証することが目的である。レートの算出方法は、森田のレート式を用いる。森田のレート式では、初期盤面勝率を何らかの手段によって求めるものとしており、この値をどのようにして求めるのかを本実験で検証する。

4.1 テスト盤面の生成

初期盤面勝率について検証するためにテスト盤面を生成する。

テスト盤面は、第3節のプレイヤー9種類をランダムな組み合わせで対戦させ、10~40手のランダムな手数を進ませた状態で止めることで生成した。本研究では1000盤面生成し、そのうち2つの途中盤面を図3(盤面A)と図6(盤面B)に示す。

盤面Aは、先攻がShiotaEvで後攻がAlphaBeta4であり、15ターン経過した途中盤面である。盤面Bは、先攻がAlphaBeta4で後攻がShiotaEVであり、27ターン経過した途中盤面である。

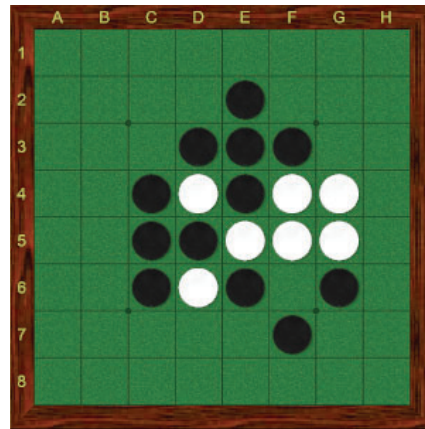


図3 盤面A

4.2 初期盤面勝率の存在の確認

得られた2つの盤面A, Bに対して、9種類のプレイヤーを100回ずつ総当たりで対戦させた。

その結果を元に、初期盤面勝率に応じた森田のレート式によるレートと本欄のレートを比較したグラフが図4と図7である。グラフではRandom, MCT, AlphaBeta4のデータのみを出している。実線が森田のレート式により求めたレートで、点線が本来のレートである。図4では、森田のレート式によって求めたレートが初期盤面勝率0.5に近いほど本来のレートに近くなっている。図7では、森田のレート式によって求めたレートの線が初期盤面勝率0.2もしくは0.8あたりのところで本来のレートの線と交わっている。

また、森田のレート式により得られたレート結果の絶対値による平均を求め、本来のレートとの差を表したグラフが図5, 図8である。盤面Aの図5では、初期盤面勝率が0.5で最も誤差が少なく、盤面Bの図8では、0.82で最も誤差が少ない。よって盤面Aでは、適切な初期盤面勝率は0.5で、盤面Bでは0.82であるといえる。

これらのことから、ランダムに生成した盤面については適切な初期盤面勝率が存在する場合があることが分かる。実験では、この初期盤面勝率を求める手法を検証する。

4.3 初期盤面勝率の算出

本論文の検証では、複数の初期盤面勝率により

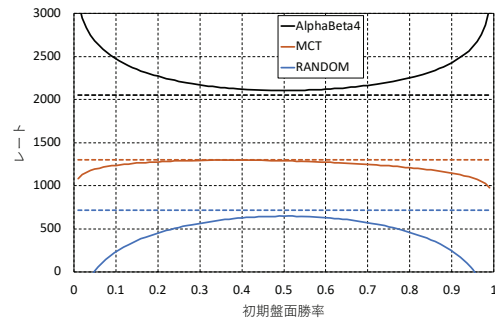


図4 森田のレートと本来のレート(盤面A)

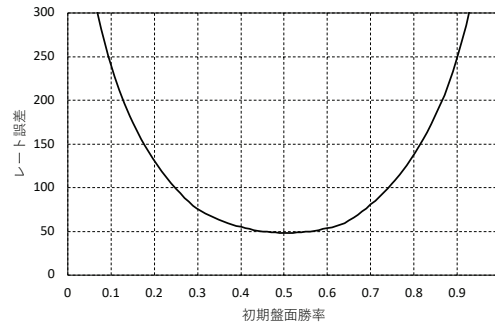


図5 レート誤差の絶対値の平均(盤面A)

求められたレートを比較する。生成した途中盤面1000盤面にランダム性のあるプレイヤー同士を各盤面ごとに対戦させた。この試行をランダム性のあるプレイヤー5種類で行うことにより、各盤面ごとに5種類の初期盤面勝率を得た。

図9~11は、初期盤面勝率の分布を表したグラ

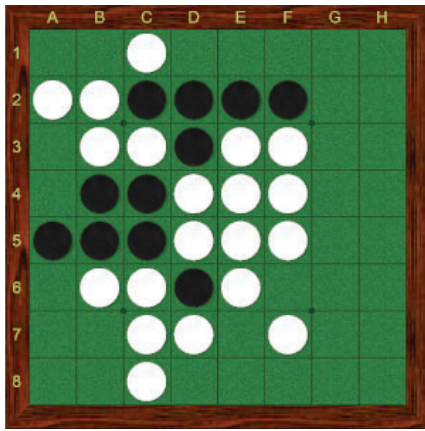


図 6 盤面 B

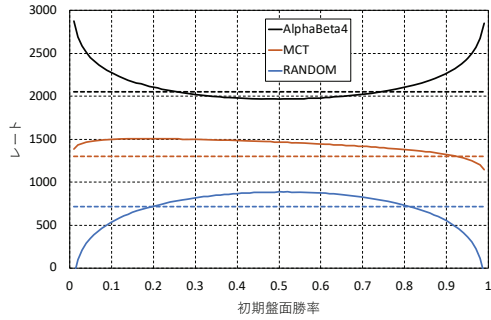


図 7 森田のレートと本来のレート (盤面 B)

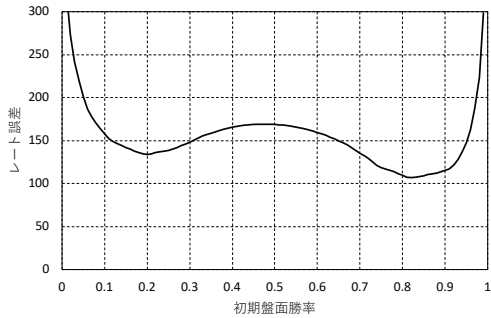


図 8 レート誤差の絶対値の平均 (盤面 B)

フである。レートの低いプレイヤーにより求めた初期盤面勝率の盤面の分布は、初期盤面勝率が0.5に近い所に多く分布している。レートの高いプレイヤーにより求めた分布は、初期盤面勝率が0か1に近い所に多く分布している。これは、強いプレイヤーは局面の有利不利が初期盤面勝率に強く影響する

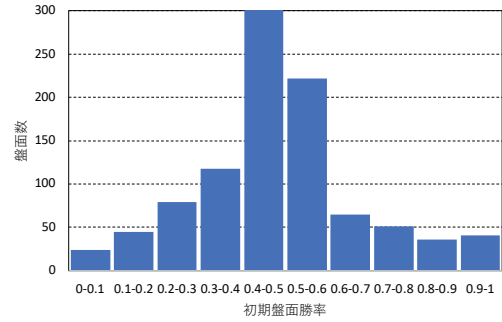


図 9 Random プレイヤの初期盤面勝率の分布

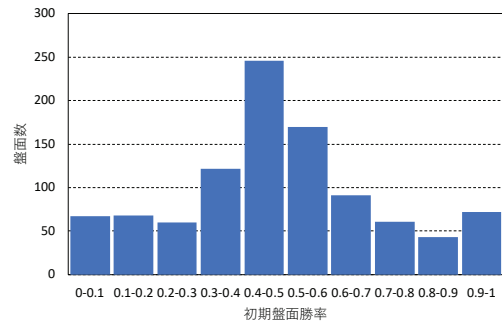


図 10 MCT プレイヤの初期盤面勝率の分布

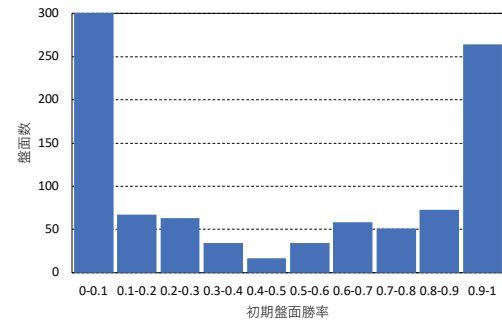


図 11 MCT+EPT8k プレイヤの初期盤面勝率の分布

ためだと考えられる。

5. 初期局面に有利不利のあるゲームでのレート算出

5.1 局面をランダムに与えた場合のレート算出

第 4.3 節で生成した 1000 盤面に対して、9 種類のプレイヤーを総当たりで 1 試合ずつ対戦させ (ただし対戦順についてはランダムとする), その対戦結果と各初期盤面勝率による森田のレート式を用

表 2 初期局面に有利不利のあるゲームでのレート差の平均

基準にしたプレイヤー	レート差
Random	95.32
MC	78.36
MCT	78.53
MCT+EPT2k	93.87
MCT+EPT8k	216.93
イロレーティング	118.18

いたレートとイロレーティングによるレートを算出した。得られたレート結果が表 4 である。

表 4 から、基準にしたプレイヤーの平均値と本来のレートとの差を表 2 に示した。表 2 から、MCT+EPT8k 以外の初期盤面勝率を基準にしたプレイヤーはイロレーティングでレートを求めた結果より、適切なレートを得ることができている。よって、この 1000 盤面においては森田 [2] のレート式による補正が期待通りに働いているといえる。MCT+EPT8k を基準にした初期盤面勝率がイロレーティングよりも悪い結果となった理由としては、図 11 より、初期盤面勝率が 0 か 1 に近い数値となっているため、初期盤面勝率が両極端になりやすいためだと考えられる。

図 12 は、初期盤面勝率の基準にしたプレイヤーごとのグラフである。グラフの縦軸は、通常の盤面でのレートとの差で系列は基準にしたプレイヤーである。Random, MCT, AlphaBeta4 プレイヤーのレート差のみを表示している。図 12 より、弱いプレイヤーの初期盤面勝率を基準にした場合は、弱いプレイヤーのレート差が正の方向に出ており、強いプレイヤーのレート差が負の方向に出ている。逆に強いプレイヤーの初期盤面勝率を基準にした場合は、弱いプレイヤーのレート差が負の方向に出ており、強いプレイヤーのレート差が正の方向に出ているという傾向がある。

例えば、MCT+EPT8k を基準にした系列では、初期盤面勝率が両極端な数値になっているため補正が強いが Random プレイヤーなどの弱いプレイヤーは初期盤面勝率とは異なった勝敗となることが多く、結果的にレート差が大きくなっている。

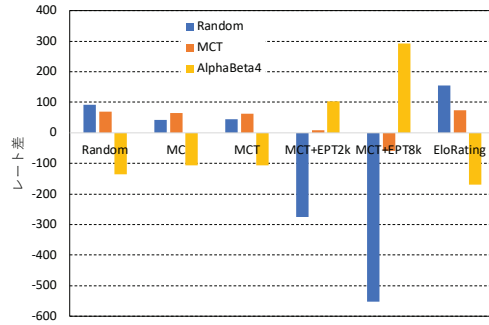


図 12 初期盤面勝率ごとのレート

表 3 強いプレイヤーに対して不利な盤面を与えた場合のレート差の平均

基準にしたプレイヤー	レート差
Random	98.56
MC	85.72
MCT	93.82
MCT+EPT2k	127.54
MCT+EPT8k	265.43

5.2 強いプレイヤーに対して不利な盤面を与えた場合のレート算出

レートを求める際に、レートの高いプレイヤーがレートの低いプレイヤーと試合する場合に、レートの高いプレイヤー側が不利となる状況のみでレートを算出する。得られたレートの差が表 5 である。

表 5 から、基準にしたプレイヤーの平均値と本来のレートとの差を表 3 に示した数値としては不利な盤面を与えていない実験 4.2 よりも差が大きい結果となった。

図 13 は、初期盤面勝率の基準にしたプレイヤーごとのグラフである。レート差の傾向としては、図 12 と同じだが、レート差の開きがより顕著に出ている。

6. 関連研究

Arpad Elo によって提案されたイロレーティング [1] は、チェスの他、サッカー、野球などにおけるプレイヤー（チーム）のレーティングに広く用いられている。また、イロレーティングを簡略化したレーティングアルゴリズムもよく用いられている。イロレーティングの問題として知られているも

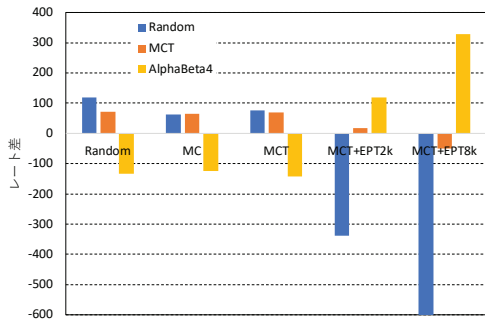


図 13 強いプレイヤーに対して不利な盤面を与えた場合のレート

のに、レート値のインフレーションの問題がある。この問題を解決するレーティングアルゴリズムとして、Mark Glickman によるグリコレーティング (Glicko および Glicko-2) [7] がある。グリコレーティングでは、各プレイヤーのレート値の他にレート値の標準偏差を管理してレーティングを行う。

イロレーティングやグリコレーティングは基本的に 2 人ゲームを対象としたレーティングアルゴリズムである。これに対して、複数人が参加する (特に複数人でチームを組むような) ゲームへ拡張したレーティングアルゴリズムとして TrueSkill [8] がある。TrueSkill では、複数人がチームを組む場合に、因子グラフ (factor graph) を構築してチーム単位のレート値を計算してレーティングを行う。

これらのレーティング手法では、初期局面に有利不利がないゲームを主に対象としている。本研究では、初期局面に有利不利があるようなゲームへのレーティングの適用を目指したものであることが大きく異なる。

7. まとめ

ランダムに生成した途中盤面にて、森田のレート式による補正が有効であることを確認することができた。

プレイヤーによって得られるレートは、組み合わせ次第で大きく変わるため初期局面の有利不利に適応した手法は見つけられなかったが、様々なプレイヤーを初期盤面勝率の基準にすることで、通常盤面とのレート差の傾向を示した。

今後の課題としては、初期盤面勝率の求め方を単一のプレイヤーだけでなく、複数のプレイヤーを使用し各局面における適切な初期盤面勝率を求める方法を検証したい。

謝辞 本稿に掲載した研究成果の一部は、高知工科大学の IACP クラスタ計算機を使用して得られたものである。

参考文献

- [1] Arpad Elo: *The Rating of Chessplayers, Past and Present*. Arco, 1978.
- [2] 森田 茂彦, 松崎 公紀: 大貧民における初期手札の不均等性を考慮したレーティングアルゴリズムの提案. 情報処理学会第 31 回ゲーム情報学研究会, 2014-GI-31(14), 2014.
- [3] 塩田 好: リバーシの評価関数について, 近畿大学理工学部情報学科卒業研究報告書, 2016.
- [4] Rémi Coulom: Efficient Selectivity and Backup Operators in Monte-Carlo Tree Search, In *Proceedings of 5th International Conference on Computers and Games (CG 2006)*, pp. 72–83, 2006.
- [5] Richard Lorentz: Using evaluation functions in Monte-Carlo Tree Search, *Theoretical Computer Science*, Vol. 644, pp. 106–113, 2016.
- [6] Gunnar Andersson: Zebra, <http://radagast.se/othello/index.html>.
- [7] Mark Glickman: *Mark Glickman's World*, <http://www.glicko.net/glicko.html>. (Reference on Dec 13, 2018)
- [8] Ralf Herbrich, Tom Minka and Thore Graepel: TrueSkill(TM): A Bayesian Skill Rating System. *Advances in Neural Information Processing Systems*, Vol. 20, pp. 569–576, 2007.

表 4 途中盤面 1000 盤面におけるレート算出結果

プレイヤー	本来のレート	初期盤面勝率を計算したプレイヤー					イロレーティング
		Random	MC	MCT	MCT+EPT2k	MCT+EPT8k	
Random	716.4	809.2	757.9	761.9	440.1	163.8	871.9
ShiotaEV	1141.5	1208.9	1184.8	1184.5	1025.0	891.0	1233.9
MC	1159.0	1276.4	1264.1	1263.0	1153.4	1044.1	1288.7
MCT	1301.0	1371.2	1365.6	1364.5	1308.5	1242.8	1375.0
MCT+EPT2k	1765.0	1763.5	1782.2	1782.0	1904.7	2020.5	1743.3
MCT+EPT8k	2111.3	1880.7	1904.7	1904.0	2087.2	2255.7	1853.7
AlphaBeta1	1415.8	1496.9	1496.8	1496.1	1498.3	1488.2	1495.9
AlphaBeta2	1837.9	1777.4	1798.2	1798.0	1927.3	2050.2	1754.9
AlphaBeta4	2052.1	1915.8	1945.8	1946.0	2155.5	2343.8	1882.7

表 5 強いプレイヤーに対して不利な盤面を与えた場合のレート算出結果

プレイヤー	本来のレート	初期盤面勝率を計算したプレイヤー				
		Random	MC	MCT	MCT+EPT2k	MCT+EPT8k
Random	716.4	834.6	778.3	791.6	378.5	76.7
ShiotaEV	1141.5	1204.2	1186.2	1190.6	957.0	803.3
MC	1159.0	1272.8	1267.0	1274.5	1124.3	991.7
MCT	1301.0	1371.6	1364.8	1370.3	1317.9	1251.7
MCT+EPT2k	1765.0	1749.8	1781.6	1779.7	1935.2	2052.4
MCT+EPT8k	2111.3	1892.1	1897.4	1875.9	2094.5	2279.9
AlphaBeta1	1415.8	1494.0	1506.5	1514.1	1558.4	1568.7
AlphaBeta2	1837.9	1761.7	1790.0	1793.9	1963.5	2095.1
AlphaBeta4	2052.1	1919.2	1928.3	1909.4	2170.7	2380.4