オセロゲームのための適応型意思決定システムの提案

白神　真一　　佐藤　裕二
法政大学情報科学部

Abstract

Strategy of Othello adopts mainly on the search, while the strategy has weak point that processing time greatly depends on environment. In this paper, we adopt strategy that processing time does not depend on environment by using some score tables. Then we propose that a score table creating method using genetic algorithm, and evaluate the method. As a result, we able to confirm a tendency to improvement in winning rate, and obtain a refinement that the evaluation of the game of go.

1. まえがき

二人零和有限確定完全情報ゲームでは，ルールと初期状態から最善手筋が決まるため，探索によって解を求めることができるといわれている[1]．しかし，現実には探索を進めるにつれ計算量が膨大になるため，多くのゲームで初期状態から最善手筋を明らかにすることは困難である．オセロもそのようなゲームの一つであり，戦略も探索を中心として構成されている．一方，探索中心の戦略はマシンの能力に処理時間が大きく依存してしまう弱点が存在する．

そこで本稿では，意思決定の際の判断基準としてマシンの能力に処理時間が大きく依存することの無い盤面に重みを付けた得点テーブル(以下，評価値リスト)を用いた戦略の強化を考える．明確な教師データが無くても強化学習が可能なこと，および並列的な探索が可能な遺伝的アルゴリズム(GA)によって状況に適応した評価値リストを自律的に生成する手段の提案と評価を行う．

2. オセロエンジンの概要

a	b	c	d	d	c	b	a
b	e	f	g	g	f	e	b
c	f	h	i	i	h	f	c
d	g	i	0	0	i	g	d
d	g	i	0	0	i	g	d
c	f	h	i	i	h	f	c
b	e	f	g	g	f	e	b
a	b	c	d	d	c	b	a

a..i: -100から100までの整数
図1. 評価値リスト

従来，評価値リストを用いた戦略は，その時点での盤面を評価値リストによって評価し，得点が大きくなるマスに石を打つという単純な仕組みになっている．従って，意思決定の際にかかる時間は短いが状況判断能力が低く，一般的に用いられることが少ない[2].

本研究では，評価値リストを用いた戦略に改良を加えることによって弱点を克服するアプローチを採る．中盤以降の盤面に多様性を持たせるため11手まではランダムに石を打ち，その後評価値リストを用いた戦略に移行する．評価値リストは図1を用いる．

5手先の盤面の状況を探索し，評価値リストによって盤面に付けられた重みを自分の石が置かれているときは加算し，相手の石が置かれているときは減算した合計点に着手可能手数に50を乗算したものを加算して評価点とし，もっとも評価点の高かった打ち手を選択する．

尚，本研究では評価値リストを2種類用い，従来の考え方に基づく評価値リスト{100, -40, -20, 5, -80, -1, -1, 5, 1}と，33手以降，かつ2つ以上の角に同色の石が置かれている状況，つまり対局の優劣がほぼ決まった際に用いる評価値リスト(以下, 条件評価値リスト)を使い分ける．

その後，残り15手以内となり，最善手筋を読み切るために必要な計算量が十分に少なくなった時点から最善手筋を計算する．

3. GAを用いた評価値リスト生成の提案

本研究では，GAの初期個体の生成の際に，隅をとると良い，Xは危険などといった，人間がオセロをする際の考え方を含めることによって，人間の考え方を含みながらも人間では考え得ない個体を生成することを図った.対局の評価では，勝敗だけでは判断できない面を補うべく独自の評価方法によって対局を評価する．

エリート個体の選択の際には，オセロエンジンの序盤戦略の影響により，能力の低い個体がエリートとして選択されてしまう可能性を，前世代のエリートと対局させて再評価することで予防する．

以上の考えから，状況に応じた評価値リストをGAによって自律的に生成する手法を提案する．

3.1. 文字列の設定と評価値リストの生成方法

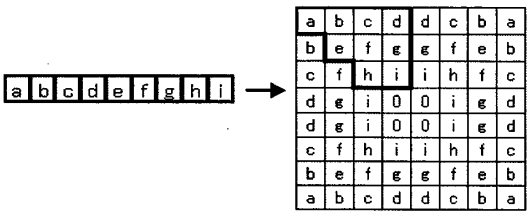

図2. 文字列から評価値リストの生成

Proposal of the Adaptive Decision Support System for Othello
Shinichi Shirakami, Yuji Sato
Faculty of Computer and Information Sciences,
Hosei University.

本研究では，オセロの盤面 64bit 分の文字列を GA によって生成することはせず，盤面の対称性を利用することによって文字列を 9bit まで減らし探索空間を限定することで処理速度の向上を考える．

文字列からの評価値リスト生成例を図 2 に示す．評価値リストを生成する際に中央 4 マスは初期配置であるので得点 0 として設定する．生成された評価値リストは，2 節のオセロエンジンの条件評価値リストとして扱われる．

3.2.本実験の流れ

実験の流れを図 3 に示す．

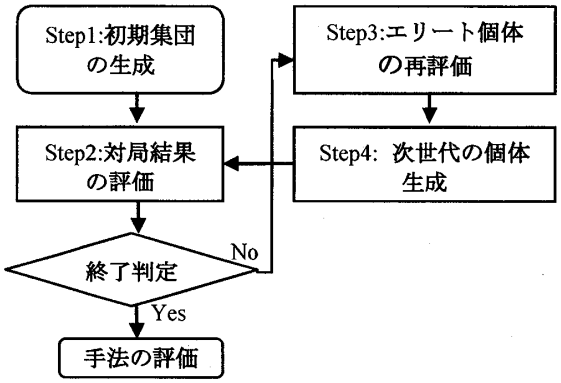

図 3. 本実験の流れ

3.2.1. 初期集団の生成：Step1

集団の大きさを設定し，人間の考え方に基づく個体{100,40,-20,5,-80,-1,-1,5,1}として個体 1 に含め，それ以外の個体をランダムに生成する．

個体の生成方法として，各個体の各要素に-100 から 100 の値をランダムに割り当てることによって生成する．

3.2.2. 対局結果の評価方法：Step2

本研究では対局の優劣がほぼ決まった際の評価値リストを生成するため，対局の評価は，勝利による評価にだけではなく，盤面の最終状態を評価に加えるものとし，次式で表す

$$p_i = 16w_i + \frac{1}{4}s_i \qquad s_i = \sum_{j=1}^{n} a_j - b_j + 64 \qquad (1)$$

p_i: 評価点 w_i: 勝利数 a_j: 自分の石数 b_j: 相手の石数 n: 対局数

(1)式で先攻 10 戦，後攻 10 戦の計 20 戦の対局を評価し，その合計点を個体の評価点とする．

個体 1 に関しては，全ての個体との対局結果の合計値から平均点を計算して評価点とする．

3.2.3. エリート個体の再評価：Step3

オセロエンジンの序盤戦略の影響により，能力の低い個体がエリートとして選択されてしまう可能性があるため，エリート個体を先代のエリート個体と先攻 50 戦，後攻 50 戦の計 100 戦対局させ，より優秀な個体を次の世代の個体 1 として残す．評価点の計算方法は Step2 と同一の方法を採る．

3.2.4. 次世代の個体の生成：Step4

選択にはルーレット選択を用いる．ルーレット選択によって 2 つの個体を選択し，交叉，突然変異させることによって次世代の個体とする．

4. 評価実験

4.1. 評価方法

本実験では，3 節の手法を世代数 50，個体数 20，交叉率 0.8，突然変異率 0.001 に設定し実験を行った．その結果生成された各世代の個体と未調整状態の基準となる個体{100,-40,20,5,-80,-1,-1,5,1}を先攻 250 戦，後攻 250 戦の計 500 戦対局させて評価点を計算し，各世代の勝率，評価点の推移を評価することで行い，評価点の計算方法は盤面の最終状態を評価するために(2)式によって計算する，その結果を勝率は 50%，評価点は全て引き分けであった際の得点 8000 を基準として学習効果を評価する．

$$p_i = \frac{1}{4}s_i \qquad s_i = \sum_{j=1}^{n} a_j - b_j + 64 \qquad (2)$$

p_i: 評価点 a_j: 自分の石数 b_j: 相手の石数 n: 対局数

4.2. 実験結果と考察

要素	学習無し	25世代	50世代	増加量
勝率(%)	50.061	50.131	50.201	0.14
評価点	8002.8	8017.375	8031.95	29.15

図 4. 世代毎の評価点・勝率の推移(試行 5 回平均)

今回の実験では，試行回数が十分に取れなかったため評価点，勝率ともに近似線を引くことによって傾向を観ることにする．近似線によると評価点，勝率ともにわずかではあるが向上傾向がみられ，提案手法による学習効果があったと考えられる．

大きな勝率の向上に繋がらなかったのは，対局の評価方法に改善の余地があったためと考えられる．本手法による評価値リストの生成の際には，生成する評価値リストの役割によって，適宜対局の評価方法を変更しなければならず，今後は対局の評価方法の設定方法を含め検討する必要がある．

5. まとめ

本稿では，マシン性能に依存しないオセロゲーム実現のため，GA を用いたオセロ戦略のための評価値リスト自律生成手段を提案し，実験，評価を試みた．その結果として，GA を用いた条件評価値リストの生成は効果がある可能性を示した，一方，大幅な向上を目指すためには，今後，対局の評価方法などの改善が必要である．

文　献

[1] 三浦琢磨，オセロ最善手筋の並列処理による解法の提案と縮小版オセロによるその評価，p1-5，2006.

[2] Seal Software，評価関数，リバーシのアルゴリズム，p102，工学社，東京，2003