# Policy Networkと既存オセロAIの融合による人間の棋譜との着手一致率を維持した棋力調整手法

山名　琢翔[1,a)]　星野　准一[2,b)]

**概要**：現代のオセロ AI には，棋力調整時に人間の着手との一致率が低下するという課題がある．本稿では，着手一致率を高く保ちながら棋力を調整する手法を提案する．既存オセロ AI“Egaroucid” の評価値を方策に変換し，人間の棋譜で学習した Policy Network の方策と重み付き幾何平均で融合した．さらに，人間のミスした着手でファインチューニングしたモデルも用意し，融合する方策を変更することで幅広い棋力への調整を実現した．評価は人間の棋譜を収録した複数のデータセットによる着手一致率と，AI 同士の対戦から推定した Elo レーティングで行った．その結果，オセロ AI と Policy Network の融合によって着手一致率が向上し，提案手法は幅広い棋力帯で既存オセロ AI を上回る着手一致率を示した．また，模倣対象の人間の棋力帯によって，着手一致率を向上させる方策の組み合わせが異なることを確認した．

**キーワード**：オセロ，棋力調整，Policy Network，着手一致率

# Strength Adjustment by Blending a Policy Network and an Existing Othello AI While Maintaining Move-Matching Accuracy with Human Game Records

Takuto Yamana[1,a)]　Junichi Hoshino[2,b)]

***Abstract:*** Adjusting the playing strength of Othello AI often reduces its move-matching accuracy with human moves. We propose a method that maintains high accuracy across a wide range of playing strengths. We convert move evaluations from the existing Othello AI “Egaroucid” into a policy and blend it with a Policy Network trained on human game records using a weighted geometric mean. We also prepare a model fine-tuned on human mistakes and vary the policy combinations to adjust playing strength. We evaluate move-matching accuracy using multiple human game datasets and estimate Elo ratings through AI-versus-AI matches. The results show that blending Othello AI with a Policy Network improves move-matching accuracy and that the proposed method achieves higher accuracy than existing Othello AI across a wide range of playing strengths. We also find that effective policy combinations differ depending on the skill levels of the human players being modeled.

***Keywords:*** Othello, playing strength adjustment, policy network, move-matching accuracy

## 1. はじめに

オセロ AI は 1997 年に人間の世界チャンピオンに勝利し [1]，2023 年には弱解決が報告され [2]，現在では超人的な強さとなっている．そのため，人間の対戦相手としてオセロ AI を活用するには意図的にオセロ AI を弱くする必要がある．既存オセロ AI の多くは，探索の深さを浅くすることで棋力調整が可能であるが，この方法では人間から見て不自然な着手となってしまう．チェスでは強い AI を意図的に弱くしたときに人間らしくない不自然な着手を行うことが報告されている [3]．

人間の着手を模倣するオセロ AI では，人間の棋譜を使っ

[1] 筑波大学人間総合科学学術院人間総合科学研究群（博士前期課程）デザイン学学位プログラム
Master’s Program in Design, Graduate School of Comprehensive Human Sciences, University of Tsukuba
[2] 筑波大学システム情報系
Institute of Systems and Information Engineering, University of Tsukuba
[a)] yamana.takuto.rg@alumni.tsukuba.ac.jp
[b)] jhoshino@esys.tsukuba.ac.jp

© 2026 Information Processing Society of Japan

て学習したニューラルネットワーク（Policy Network）によって高い着手一致率を実現できたと報告されている [4]．また，複数の方法で学習した Policy Network を融合させることで様々なゲームにおいて着手一致率を高める取り組みがある [5]．一方，これらの研究では棋力の調整は行われていない．棋力を調整する方法としては，AlphaZero 方式のゲーム AI においてニューラルネットワークが出力した Value を使用して意図的に探索時のパラメータを調整し，互角に戦えるようにした事例が存在する [6]．しかし，この研究では人間との着手一致率に焦点を当てていない．

本稿では，棋力を自由に調整でき，どの強さでも人間との着手一致率が高いオセロ AI を提案する．Policy Network を持たない既存のオセロ AI の出力を加工して方策とし，人間の着手で学習した Policy Network と重み付き幾何平均で融合した．この重みを調整することで棋力の調整を行う．また，人間がミスした手でファインチューニングした Policy Network を融合することで，人間の弱いプレイヤーの着手との一致率が高く，弱いオセロ AI も実現した．

## 2. 関連研究

本節では，人間との着手一致率の高いゲーム AI，動的に棋力を調整できるゲーム AI，人間らしさを維持した棋力調整の先行事例を概説する．

### 2.1 人間との着手一致率の高いゲーム AI

様々なゲームにおいて人間の着手を推定する手法が Ogawa らによって提案された [5]．この手法では，人間の棋譜で学習した Policy Network を棋力帯ごとに複数用意し，AlphaZero 型の強いゲーム AI の Policy Network の出力と幾何平均で融合させ，将棋，囲碁，チェスにおいて人間との高い着手一致率を中級者・上級者それぞれで実現した．本稿では，Policy Network を持たないオセロ AI の出力を加工し，方策として利用した点，人間のミスした手でファインチューニングした方策を融合することで弱い人間の着手を模倣できるようにした点，既存オセロ AI も交えて強さと人間との着手一致率の関係を評価した点が異なる．

人間の棋譜を学習した Policy Network 単体で人間との着手一致率が 62.7%，複数の Policy Network の bagging による併用で 64.0%に達したとする研究が存在する [4]．この先行研究に対し，本稿では既存のオセロ AI の出力と Policy Network を融合することで人間との着手一致率を向上させたほか，対戦で強さの評価も行った．なお，本稿では Policy Network の設計においてこの先行研究で人間との着手一致率が一番高いと報告されたモデルを採用した．

### 2.2 動的に棋力を調整するゲーム AI

様々なゲームにおいて，AlphaZero 型 AI の棋力を動的に調整する研究がある [6]．この先行研究では，AlphaZero 型 AI 内のニューラルネットワークが出力した Value によって，PV-MCTS の探索回数の変更や，ニューラルネットワーク自体の精度の変更等を行い，棋力調整を行った．一方，この先行研究では着手の人間らしさについては評価していない．本研究では，既存オセロ AI と Policy Network の融合による棋力調整方法を提案し，人間の棋譜との着手一致率を高く保ったまま棋力を調整した点が異なる．なお，本稿の手法はパラメータ調整によって棋力変更が可能であるため，[6] の棋力調整手法を使うことで，実際の対局での棋力調整が可能になると考えている．

著者らによってオセロ指導対局ロボット “Minoth” が開発されている．展示の観察により，オセロ初心者に対してロボットが強すぎて人間の学習機会を十分に提供できないという問題が発覚した [7]．本研究はこの問題を解決する，手加減を行う独自オセロ AI の開発と位置づけられる．

### 2.3 人間らしさを維持した棋力調整

Jacob らの取り組み [8] では，人間を模倣したプレイヤーが棋力に劣るため，探索によって人間との着手一致率を維持したまま棋力を改善した．特に，人間を模倣した Policy Network を活用した PV-MCTS による探索を行った．本稿では，人間を模倣したプレイヤーに対して，強いオセロ AI と弱いオセロ AI の両方を作成した点が異なる．また，人間らしさと棋力調整を両立させる手法が複数 AI の出力の単純な融合である点が異なる．

Hsueh らの取り組み [9] では，囲碁において中級者程度の棋力を持つ AI と勝率 50%程度で対戦できる，人間らしい手を打つ AI を提案している．この先行研究では，強豪囲碁 AI の形勢評価と人間の着手を学習した Policy Network を本稿とは異なる方法で併用している．それぞれの AI が強いまたは人間らしいと評価した手を MCTS によって探索し，現在の局面の形勢を用いて，人間らしさをなるべく損なわずに打つ手の強さを決定している．また，[9] では人間との着手一致率を評価していないが，本稿では棋力と人間との着手一致率の双方を評価した．

## 3. 提案手法

本稿では，既存オセロ AI の探索結果を活用した方策と人間の棋譜を学習した Policy Network をそれぞれ複数用意した．これらのモデルから 2 つを選んで融合することで棋力調整と人間との着手一致率の向上を図った．

### 3.1 既存オセロ AI を活用した方策出力

既存のオセロ AI の多くは $\alpha\beta$ 法で実装されており，Policy Network を持たない．そこで，Egaroucid for Console 7.8.1 [10]（以後，単に Egaroucid と表記する）レベル 1 および 21 によって各合法手の評価値（予想最終石差）を計算し，式 (1) によって方策に変換した．$v_{s,a}$ を Egaroucid が

© 2026 Information Processing Society of Japan

局面 $s$ で手 $a$ に与えた評価値，$\tau > 0$ を定数，$\pi_{AI,s,a}$ を方策値とする．なお，レベルとは Egaroucid の強さを決める整数であり，探索深さや Multi-ProbCut [11] の確率を決定する．レベル 21 は Egaroucid のデフォルトで使われており，人間に対して十分な強さであることから，強い棋力を実現するために使用した．一方，レベル 1 は設定可能な最弱のレベルであり，弱い棋力を実現するために使用した．

$$\pi_{AI,s,a} = \frac{\exp(\tau v_{s,a})}{\sum_{b \in A(s)} \exp(\tau v_{s,b})} \tag{1}$$

式 (1) は，オセロ AI による評価値を強い手を選ぶ際の有望度合いに変換する意味を持つ．最善手とそれ以外の手で評価値の差が激しければ，$\pi_{AI,s,\text{最善手}}$ が 1 に近づき，$\pi_{AI,s,\text{次善手以降}}$ は 0 に近くなる．また，$\tau$ が大きくなるほど，評価値の差に対して $\pi_{AI,s,a}$ の差が大きくなる．

### 3.2 人間の棋譜を学習した Policy Network

人間によるオセロの対局データを用いて，Policy Network を学習した．入力は，手番側・相手側各 64 マスの石の有無を表すため，8×8 ピクセルの画像 2 枚とした．出力層はオセロ盤上の中央 4 マスを除いた 60 ノード，活性化関数は softmax で，各マスに人間が打つ確率を推測する．中間層は Liskowski らによる先行研究 [4] で人間との着手一致率が一番高いと報告された，8 層の畳み込みニューラルネットワーク（**図 1**）を使用した．畳み込み層のフィルタは $3 \times 3$ で，ストライドは 1 とし，ゼロパディングを行う．畳み込み層および 128 ノードの全結合層には Batch Normalization を使用し，活性化関数は ReLU とした．

なお，式 (2)，(3) に示すように，出力された方策のうち，非合法手の方策は 0 とし，すべての合法手の方策値の和が 1 となるように補正した．$s$ は局面，$a$ は $s$ における 1 つの手，$A(s)$ は $s$ における合法手集合，$\hat{\pi}_{s,a}$ は Policy Network が出力した方策値，$\pi_{PN,s,a}$ は補正後の方策値とする．

$$P_{s,a} = \begin{cases} \hat{\pi}_{s,a} & (a \in A(s)) \\ 0 & (a \notin A(s)) \end{cases} \tag{2}$$

$$\pi_{PN,s,a} = \frac{P_{s,a}}{\sum_{b \in A(s)} P_{s,b}} \tag{3}$$

### 3.3 方策の融合

本稿では，複数の方策モデルを用意した上で，2 つの方策モデルを選び，式 (4) で融合した．この融合手法は，Ogawa らの手法 [5] と同様であるが，[5] では着手一致率の向上のために融合を行ったのに対し，本稿では棋力の調整のために融合を行った点で用途が異なる．$0 \leq \alpha \leq 1$ を融合度合いを調整する実数パラメータとした．$\alpha = 0$ で $\pi_0$，$\alpha = 1$ で $\pi_1$ の出力となる．本稿では，$Q_{s,a}$ が最大となる手を選択することとした．

$$Q_{s,a} = \pi_0^{1-\alpha} \times \pi_1^{\alpha} \tag{4}$$

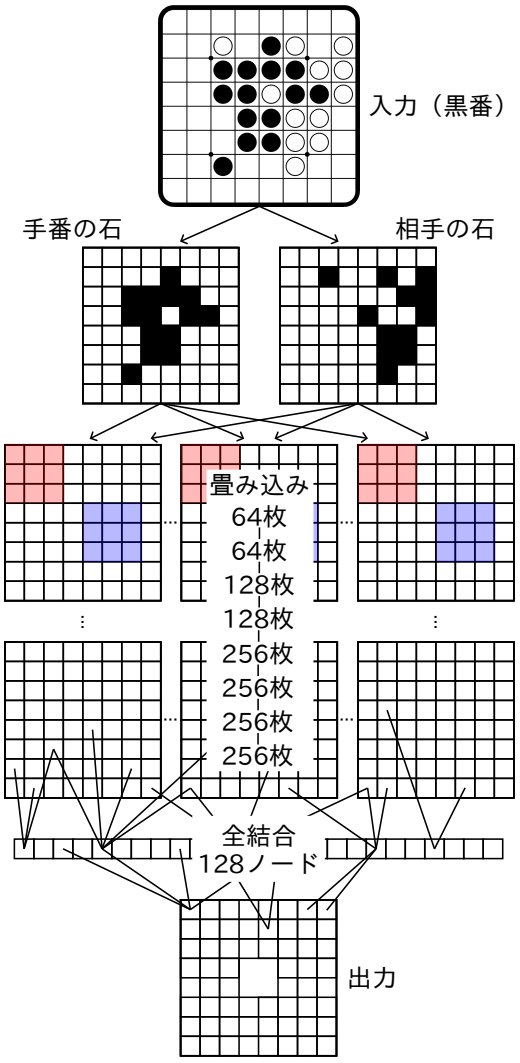



**図 1** Policy Network の構造

ここで $\pi_1 = \pi_{AI}$ のときを考える．$\sum_{b \in A(s)} \exp(\tau v_{s,b})$ は同一局面内で共通であるから，式 (1)，(4) より，選ぶ手は

$$\underset{a \in A(s)}{\operatorname{argmax}} (Q_{s,a}) = \underset{a \in A(s)}{\operatorname{argmax}} \left( \pi_0^{1-\alpha} \times \exp(\alpha \tau v_{s,a}) \right)$$

と表せる．ここで，$\alpha' = 1 - (1-\alpha)/(1-\alpha+\alpha\tau)$ と置くと，右辺で自然対数を取り，$1 - \alpha + \alpha\tau$ で割って，

$$\underset{a \in A(s)}{\operatorname{argmax}} (Q_{s,a}) = \underset{a \in A(s)}{\operatorname{argmax}} \left( (1-\alpha') \ln \pi_0 + \alpha' v_{s,a} \right)$$

である．定義から $0 \leq \alpha' \leq 1$ であり，$\alpha'$ のみで $\alpha$ および $\tau$ で操作できる範囲を操作できるため，$\tau$ は冗長である．そのため，本稿では $\tau = 1$ で固定した．

## 4. 学習

本稿では，フランスオセロ連盟が公開している，オセロの大会等の棋譜を集めたデータセット（WTHOR データ）[12] を使い，実際に人間が打った手を学習した．さらに，WTHOR データのうち，一定以上の石損がある着手のみでファインチューニングを行うことで，人間のミスを模倣した Policy Network も作成した．

© 2026 Information Processing Society of Japan

### 4.1 事前学習

まず，WTHOR データ [12] 全てで 3.2 節の Policy Network を学習した．学習に使用したデータは，1977 年から 2024 年までの終局している棋譜 134,181 局である．同一棋譜由来の局面が学習・検証・テストの各データに入らないよう，まずこれらの棋譜を 3 つに分割し，120,768 局を学習，6,707 局を検証，6,706 局をテストに使用した．これらの棋譜に含まれる局面数はそれぞれ 7,232,087，401,650，401,545 であり，各局面で人間が選んだ手を正解ラベルとした．学習時には，学習データのみ回転・反転による変換を適用し，8 倍に水増しした．この水増し手法は Liskowski らの取り組み [4] で着手一致率が向上したと報告されている．

損失関数には，人間の着手を正解ラベルとするクロスエントロピーを用い，非合法手の確率も 0 に補正せずに計算した．最適化手法には Adam を用い，学習率を $10^{-3}$ に固定し，バッチサイズを 512 として，NVIDIA GeForce RTX 3090 で 10 エポック学習し，検証 Loss が最小のモデルを採用した．学習時間は 4 時間 57 分であった．このモデルを「通常モデル」とする．

### 4.2 ファインチューニング

4.1 節で学習した通常モデルに対して，WTHOR データのうち一定以上の石損がある着手のみでファインチューニングを行った．石損とは，オセロにおいて実際に打った手が最善手から何石損だったかを表す数値である．本稿では，WTHOR データの全ての着手を Egaroucid [10] レベル 10 で解析して石損を求め，2 以上の石損を持つデータをファインチューニングに使用した．なお，Egaroucid のレベルは探索深さを設定する数値で，レベル 10 では中盤 10 手読み，20 マス空きから完全読みを行う．ファインチューニングに使用したデータを**表 1** に示す．

**表 1** 石損数ごとの WTHOR データ数

| | データ（盤面）数 | | |
|---|---|---|---|
| 石損数 | 学習（水増し前） | 検証 | テスト |
| 全て | 7,232,087 | 401,650 | 401,545 |
| 2 以上 | 1,668,844 | 93,447 | 94,189 |

表 1 に示した 2 石損以上の学習データを 4.1 節と同様に回転・反転で 8 倍に水増しして学習を行った．事前学習モデルの全ての層を学習するようにした上で，最適化手法には新たに初期化した Adam を用い，学習率を $10^{-4}$ に固定し，バッチサイズを 512 として 10 エポック学習した．他の学習条件は 4.1 節と同じとし，検証 Loss が最小のエポックのモデルを採用した．このモデルを「石損モデル」とする．

## 5. 評価手法

本稿で行う着手一致率と強さの評価手法を解説する．

### 5.1 強さの評価手法

各プレイヤーの強さは総当たり戦で計算した推定 Elo レーティングによって評価した．対戦では，XOT [13] で開始局面にばらつきを与えた．各開始局面で手番を変えて 2 回対戦を行い，合計石差の多い方を勝者とした．各モデルが他の全モデルに対して各 50 組（先手・後手各 50 戦）の対戦を行い，勝ちを 1 点，引き分けを 0.5 点とした勝ち点から Elo レーティングを推定した．なお，全てのプレイヤーの組み合わせで独立して XOT から 50 個の開始局面を重複なく無作為に抽出した．各対局が終わるたびにオセロ AI の置換表のデータを消去し，置換表に登録されたデータが複数の対局で流用されないようにした．

#### 5.1.1 Elo レーティングの推定方法

プレイヤー A のプレイヤー B に対する予想勝率を $p_{A\to B}$，それぞれの Elo レーティングを $R_A, R_B$ とすると，

$$p_{A\to B} = \frac{1}{1+10^{-(R_A-R_B)/400}}$$

である．総当たり戦の各プレイヤーの組み合わせにおいて，予想勝率 $p_{A\to B}$ が実際の勝率に近づくようにレーティングを最適化した．実際には，式 (5) を最小化するようにニュートン法によってレーティングを決定した後，平均値が 1500 となるように補正した．

$$L = -\sum_{i<j}\Big((S_{i\to j}+0.5)\log_{10} p_{i\to j} + (N_{ij}-S_{i\to j}+0.5)\log_{10}(1-p_{i\to j})\Big) \quad (5)$$

$S_{i\to j}$ はプレイヤー $i$ が $j$ に対して得た合計勝ち点，$N_{ij}$ はプレイヤー $i$ と $j$ の対戦回数（常に 50）である．また，式 (5) 内の 0.5 は，全勝や全敗でレーティングが発散しないよう，あらかじめ 0.5 勝と 0.5 敗を加える処理である．

信頼区間はブートストラップ法で求めた．2 プレイヤーの対戦における 50 組の対戦結果から重複を許してランダムに 50 組の対戦を抽出する操作を，全てのモデルの組み合わせにおいて行い，レーティングの推定を行う．これを 20,000 回繰り返し，各モデルのレーティング推定値の 2.5 から 97.5 パーセンタイルの区間を 95%信頼区間とした．

### 5.2 着手一致率の評価手法

人間との着手一致率は，各プレイヤーの選ぶ手が正解ラベル（人間が実際に選んだ手）と一致した割合で評価した．評価に使用したデータセットを**表 2** に示す．本稿の着手一致率の計算では，盤面に対称性があり，人間が実際に着手した手と実質的に同じ別の手が存在する場合，対称的な手も正解とした．また，石損は 4.2 節の手法と同様に計算した．WTHOR 検証データは，着手一致率が最大となる融合時の定数を探すために使用した．

WTHOR データに加え，様々な棋力帯の着手を収録したデータとして，オンライン対戦システムであるオセロク

© 2026 Information Processing Society of Japan

エスト [14] の 5 分対局における棋譜（オセロクエストデータ）も用いた．このデータでは，オセロクエスト内で計算されたレートによってプレイヤーの棋力を数値的に判定できるため，レート帯によって分割し，様々な棋力で人間の着手との一致率を求めた．各レート帯において，条件を満たす局面全ての中からランダムに 10 万局面を選んで着手一致率の計算に使用した．なお，本稿では，黒と白の両方のプレイヤーが同一のレート帯に属していた場合の棋譜のみを用いた．これは，オセロクエストでは対局中に相手のレートを見る設定が可能で，自分のレートと比べて相手のレートが極端に違う場合に，十分に思考せずに着手してしまう場合が考えられるためである．

**表 2**　着手一致率の評価に使用したデータセット

| データセット | 対象局面 | 使用局面数 |
|---|---|---|
| WTHOR | 検証データ全体 | 100,000 |
| WTHOR | テストデータ全体 | 100,000 |
| WTHOR | テストデータ（石損 2 以上） | 94,189 |
| オセロクエスト | レート $[800, 1000)$ | 100,000 |
| オセロクエスト | レート $[1000, 1200)$ | 100,000 |
| オセロクエスト | レート $[1200, 1400)$ | 100,000 |
| オセロクエスト | レート $[1400, 1600)$ | 100,000 |
| オセロクエスト | レート $[1600, 1800)$ | 100,000 |
| オセロクエスト | レート $[1800, 2000)$ | 100,000 |
| オセロクエスト | レート $[2000, 2200)$ | 100,000 |

### 5.3　信頼区間の計算方法

本稿では，Elo レーティングと着手一致率をあわせた信頼区間をブートストラップ法により推定した．$i$ 番目のプレイヤーの推定 Elo レーティングを $R_i$，ブートストラップ法による $B = 20,000$ 回の試行の中で $b$ 回目に得られた推定値を $R'_{i,b}$，$B$ 回の試行によって得られた $R'_i$ の平均値を $\bar{R}'_i$ とする．すると，$R'_i$ の標準偏差 $\sigma_{R',i}$ は

$$\sigma_{R',i} = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}\left(R'_{i,b} - \bar{R}'_i\right)^2}$$

である．同様にして着手一致率 $p'_i$ の標準偏差 $\sigma_{p',i}$ も得る．

ブートストラップ法の各試行において，プレイヤー $i$ の Elo レーティングと着手一致率が推定値から最大で標準偏差の何倍ずれたかを表す $T_b$ を式 (6) で計算する．

$$T_b = \max\left(\max_i \frac{\left|R'_{i,b} - R_i\right|}{\sigma_{R',i}}, \max_i \frac{\left|p'_{i,b} - p_i\right|}{\sigma_{p',i}}\right) \qquad (6)$$

ブートストラップ法によって $T_b$ を $B$ 回計算し，その 95 パーセンタイルを $q$ として，各プレイヤーの Elo レーティングの信頼区間を $R_i \pm q\sigma_{R',i}$，着手一致率の信頼区間を $p_i \pm q\sigma_{p',i}$ とした．グラフ上でこの 95%信頼区間を長方形で図示し，同じプレイヤー群内で隣り合う点の信頼区間領域を含む最小の凸多角形で繋いだ．

## 6.　着手一致率と強さの評価

提案手法で実際に強さを調整できるか，人間との着手一致率を高められるかの 2 点を検証する．

### 6.1　Policy Network 単体での人間との着手一致率

4.1 節および 4.2 節で学習した各モデルについて，人間の着手との一致率を計測した．

#### 6.1.1　WTHOR データでの着手一致率

表 2 の WTHOR のテストデータ（全体・石損 2 以上）によって人間との着手一致率を評価した結果を**表 3** に示す．同一テストデータ内で着手一致率が高い方を強調した．

**表 3**　各モデルの WTHOR テストデータにおける着手一致率（%）

| テストデータ | 局面数 | モデル名 | 着手一致率 | 95%信頼区間 |
|---|---|---|---|---|
| 全体 | 100,000 | 通常モデル | **63.73** | 63.43–64.03 |
| | | 石損モデル | 50.82 | 50.51–51.13 |
| 石損 2 以上 | 94,189 | 通常モデル | 36.42 | 36.12–36.73 |
| | | 石損モデル | **45.06** | 44.75–45.38 |

石損モデルはファインチューニングの成果により，石損 2 以上のテストデータにおける着手一致率が通常モデルよりも高かった．しかし，通常モデルがテストデータ全体で 63.73%の着手一致率だったことに対し，石損モデルは石損 2 以上のテストデータで着手一致率が 45.06%に留まった．これは，人間によるミスを模倣することが，ミスのない手を模倣することよりも困難である可能性を示唆している．なお，[4] では本稿と同等のニューラルネットワークを WTHOR データで学習し，62.7%の着手一致率を実現している．学習データの分割手法に違いはあるが，本稿の通常モデルの着手一致率と近い数値となった．

表 3 での 95%信頼区間は，Wilson 法によって求めた．$n$ 件のテストデータのうち $h$ 件で着手が一致したとして，観測した着手一致率を $\hat{p} = h/n$ とする．真の着手一致率として $p$ を仮定したとき，着手一致率の標準偏差は二項分布より $\sqrt{p(1-p)/n}$ である．95%信頼区間を求めるために正規分布表より $z \approx 1.96$ を導入して，

$$|p - \hat{p}| \le z\sqrt{\frac{p(1-p)}{n}}$$

を満たす $0 \le p \le 1$ の範囲を 95%信頼区間とした．

#### 6.1.2　オセロクエストデータでの着手一致率

表 2 のオセロクエストデータでの着手一致率を**表 4** に示す．同一テストデータ内で着手一致率が高い方を強調した．通常モデルと石損モデルの両方で，低レート帯での着手一致率は低くなる傾向が見られた．また，通常モデルは高レート帯での着手一致率が石損モデルよりも高い傾向，逆に石損モデルは低レート帯での着手一致率が高くなる傾向が見られた．

© 2026 Information Processing Society of Japan

**表 6** 評価に使用した提案手法のプレイヤー

| 系列名 | $\pi_0$ | $\pi_1$ | $\alpha$ |
|---|---|---|---|
| レベル 21-通常モデル | Egaroucid レベル 21 | 通常モデル | 0.0 , 0.2, 0.4, 0.6, 0.7, 0.8, 0.9, 1.0 |
| 通常モデル-レベル 1 | 通常モデル | Egaroucid レベル 1 | (0.0), 0.2, 0.4, 0.6, 0.8, 1.0 |
| 通常モデル-石損モデル | 通常モデル | 石損モデル | (0.0), 0.2, 0.4, 0.6, 0.8, 1.0 |

**表 4** 各モデルのオセロクエストデータにおける着手一致率（%）

| レート帯 | モデル名 | 着手一致率 | 95%信頼区間 |
|---|---|---|---|
| $[800, 1000)$ | 通常モデル | 34.48 | 34.19–34.78 |
| | 石損モデル | **36.42** | 36.12–36.72 |
| $[1000, 1200)$ | 通常モデル | 39.26 | 38.96–39.56 |
| | 石損モデル | **41.04** | 40.74–41.35 |
| $[1200, 1400)$ | 通常モデル | 45.68 | 45.38–45.99 |
| | 石損モデル | **46.33** | 46.02–46.64 |
| $[1400, 1600)$ | 通常モデル | **51.84** | 51.53–52.15 |
| | 石損モデル | 49.34 | 49.03–49.65 |
| $[1600, 1800)$ | 通常モデル | **58.06** | 57.75–58.36 |
| | 石損モデル | 50.92 | 50.61–51.23 |
| $[1800, 2000)$ | 通常モデル | **63.07** | 62.77–63.36 |
| | 石損モデル | 51.61 | 51.30–51.92 |
| $[2000, 2200)$ | 通常モデル | **65.99** | 65.69–66.28 |
| | 石損モデル | 51.66 | 51.35–51.97 |

### 6.2 各モデルの強さの測定・使用するプレイヤーの決定

3.3 節の手法では，複数のモデルから 2 つのモデルを選び，融合する．本稿では，既存オセロ AI (Egaroucid) のレベル 21 および 1 を活用した方策，WTHOR データによって学習した通常モデル，石損モデルの合計 4 つのモデルを融合対象とするため，融合時に 2 つのモデルを選ぶ必要がある．本稿では棋力を調整するため，棋力の離れた 2 モデルを融合する必要がある．本節では，融合に使用するモデルを選ぶため，4 つのモデルの総当たり戦を行った．

総当たり戦の結果および 4 プレイヤー内での推定 Elo レーティングを**表 5** に示す．Egaroucid レベル 21 が一番強く，その次に通常モデルが強いことがわかった．Egaroucid レベル 1 と石損モデルは Elo レーティングの 95%信頼区間が大きく重なっており，明確な棋力差は確認されなかった．

**表 5** 各モデルの総当たり対戦における勝率と Elo レーティング

| モデル名 | 合計勝率（%） | 推定 Elo | 95%信頼区間 |
|---|---|---|---|
| Egaroucid Lv.21 | 100.00 | 2164.5 | 2139.3–2204.4 |
| Egaroucid Lv.1 | 20.33 | 1162.9 | 1087.9–1222.0 |
| 通常モデル | 60.00 | 1515.3 | 1469.0–1572.5 |
| 石損モデル | 19.67 | 1157.3 | 1095.9–1206.0 |

本稿では，各モデルの強さを踏まえ，提案手法として，**表 6** に示すプレイヤーを使用した．なお，通常モデル-レベル 1 の $\alpha = 0.0$，通常モデル-石損モデルの $\alpha = 0.0$ はレベル 21-通常モデルの $\alpha = 1.0$ と一致するため，1 つのプレイヤーとしてまとめた．

**表 7** 評価に使用した既存手法のプレイヤー

| 系列名 | 使用したレベル |
|---|---|
| Egaroucid | (1), 2, 3, 5, 9, 13, 17, (21) |

また，提案手法の比較対象として，強力なオープンソースのオセロ AI として知られる Egaroucid [10] を使用した．使用したプレイヤーを**表 7** に示す．レベルは探索深さを設定する整数であり，大きいほど深い先読みや早い段階からの終盤完全読みを行うため，強い傾向がある．なお，レベル 1 およびレベル 21 は，それぞれ提案手法の通常モデル-レベル 1 における $\alpha = 1.0$，レベル 21-通常モデルの $\alpha = 0.0$ と一致するため，1 つのプレイヤーとしてまとめた．

### 6.3 WTHOR で着手一致率を最大化するプレイヤー

WTHOR 検証データを使い，表 6 のプレイヤーすべての着手一致率を 6.1.1 節と同様の方法で評価した．その結果を**表 8** に示す．レベル 21-通常モデルの $\alpha = 0.7$ で着手一致率が最大となったが，95%信頼区間は $\alpha = 0.8$ と広く重なっており，これらの着手一致率の違いはほとんど見られなかった．表 8 ではレベル 21-通常モデルでの $\alpha = 0.7$ の着手一致率と信頼区間，$\alpha = 0.8$ の信頼区間を強調した．

### 6.4 強さと人間との着手一致率の関係

横軸を対戦結果から推定した，プレイヤー間での Elo レーティング，縦軸を人間との着手一致率としたグラフを**図 2**，**図 3**，**図 4**，**図 5**，**図 6** に示す．着手一致率の評価に使用したデータは表 2 に示した WTHOR テストデータ（全体・石損 2 以上），オセロクエストデータ（レート $[800, 1000)$，$[1400, 1600)$，$[2000, 2200)$）である．図中の半透明の領域は Elo レーティングおよび着手一致率の両者を加味した 95%信頼区間を表す．なお，縦軸は図ごとに違う．

図 2 では，提案手法（レベル 21-通常モデルおよび通常モデル-レベル 1）が既存手法（Egaroucid）と比べ，レーティング 1000 から 2400 程度までで 95%信頼区間全体が上回っており，人間との着手一致率が高いことが確認できる．また，6.3 節の結果と同様に，通常モデルと Egaroucid レベル 21 の融合により，融合していない通常モデル単体（63.73%）や Egaroucid レベル 21 単体（63.46%）よりも高い着手一致率を実現できた．検証データで着手一致率が高かった $\alpha = 0.7, 0.8$ では，テストデータでも高い着手一致率を示し，それぞれ 66.44%，66.60%であった．

© 2026 Information Processing Society of Japan

表 8　WTHOR 検証データ 100,000 局面における着手一致率と 95%信頼区間（%）

| 系列名 | レベル 21-通常モデル | | | | | | | | 通常モデル-レベル 1 | | | | | 通常モデル-石損モデル | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ | 0.0 | 0.2 | 0.4 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 | 0.2 | 0.4 | 0.6 | 0.8 | 1.0 | 0.2 | 0.4 | 0.6 | 0.8 | 1.0 |
| 着手一致率 | 63.59 | 65.05 | 65.57 | 66.36 | **66.66** | 66.58 | 65.55 | 63.41 | 62.79 | 60.49 | 58.02 | 55.79 | 53.10 | 63.47 | 62.06 | 57.91 | 54.37 | 50.65 |
| 下限 | 63.29 | 64.75 | 65.27 | 66.06 | **66.37** | **66.28** | 65.25 | 63.11 | 62.49 | 60.19 | 57.71 | 55.49 | 52.79 | 63.17 | 61.76 | 57.60 | 54.06 | 50.34 |
| 上限 | 63.88 | 65.35 | 65.86 | 66.65 | **66.95** | **66.87** | 65.84 | 63.71 | 63.09 | 60.79 | 58.32 | 56.10 | 53.41 | 63.77 | 62.36 | 58.22 | 54.67 | 50.96 |

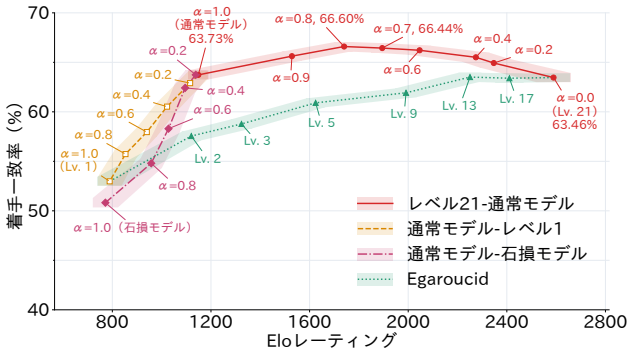



**図 2**　強さと人間との着手一致率の関係
（WTHOR テストデータ全体）

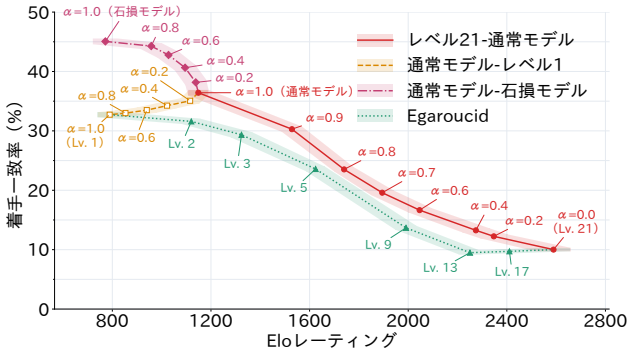



**図 3**　強さと人間との着手一致率の関係
（WTHOR テストデータ・石損 2 以上）

レベル 21-通常モデルの $\alpha = 0.8$ で WTHOR テストデータ全体において人間と着手が一致しなかった局面を詳しく調査した．本モデルが正解できなかった局面のうち，テストデータ内に対称性を加味して同一とみなせる局面が 3 つ以上あり，さらに人間の着手もデータごとに複数に分かれていた局面は 17.6%であった．これらは人間の中でも着手がばらついた局面であり，その中で本モデルが，人間の選ぶ確率が同率 1 位を含めて一番高かった手を選んだ割合は 65.5%であった．つまり，本モデルと人間の実際の着手が食い違った局面のうち，11.5%（$= 17.6\% \times 65.5\%$）は，本モデルが多くの人間の選ぶ手を選べていた．

図 3 では，強い人間プレイヤーのミスした手として，WTHOR データにおける石損が 2 以上の局面での着手一致率を評価した．提案手法（レベル 21-通常モデルおよび通常モデル-石損モデル）がレーティング 800 から 2400 程度にかけて 95%信頼区間全体で既存手法よりも高い着手一致率となった．また，本条件では，提案手法・既存手法ともに弱い AI ほど着手一致率が向上する傾向が見られた．

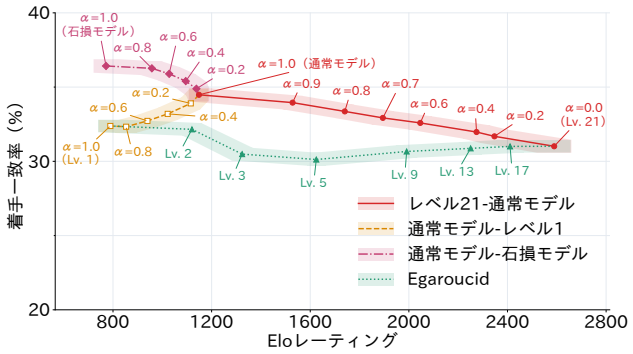



**図 4**　強さと人間との着手一致率の関係
（オセロクエスト・レート [800, 1000)）

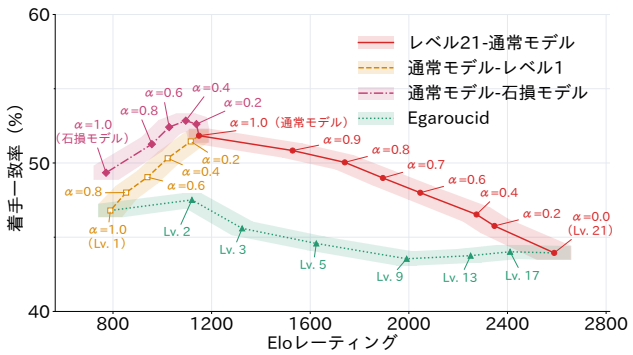



**図 5**　強さと人間との着手一致率の関係
（オセロクエスト・レート [1400, 1600)）

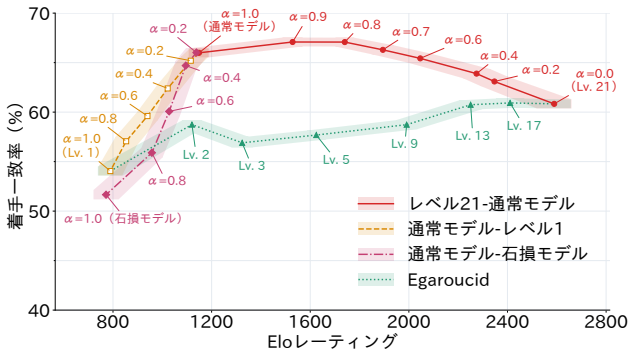



**図 6**　強さと人間との着手一致率の関係
（オセロクエスト・レート [2000, 2200)）

図 4 では，オセロクエストにおける比較的レートの低いプレイヤー（800 以上 1000 未満）との着手一致率を評価した．提案手法（レベル 21-通常モデルおよび通常モデル-石損モデル）の着手一致率がレーティング 800 から 2200 程度にかけて 95%信頼区間全体で既存手法を上回った．また，図 2 の WTHOR（強い人間プレイヤーの着手）における着手一致率とは異なり，提案手法を同程度の棋力帯で

© 2026 Information Processing Society of Japan

比べたとき，通常モデル-石損モデルの着手一致率が通常モデル-レベル 1 よりも高い傾向が見られた．これは，オセロ AI としては同程度の強さであっても，Egaroucid レベル 1 よりも強い人間プレイヤーのミスの方が弱い人間プレイヤーの着手に近い可能性を示唆している．

図 5 はオセロクエストにおける中程度のレート（1400 以上 1600 未満）を持つプレイヤーの着手一致率によって比較を行った結果である．このテストデータでも，提案手法のうち，レベル 21-通常モデルおよび通常モデル-石損モデルの着手一致率がレーティング 800 から 2400 程度にかけて 95%信頼区間全体で既存手法を上回った．

図 6 はオセロクエストにおける非常に強いプレイヤー（レートが 2000 以上 2200 未満）の着手との一致率を測定したものである．レベル 21-通常モデルおよび通常モデル-レベル 1 がレーティング 1000 から 2400 にかけて 95%信頼区間全体で既存手法の着手一致率を上回った．なお，図 6 は WTHOR での着手一致率（図 2）と非常に似た傾向・着手一致率の値を示した．この結果から，WTHOR に記録された着手データとオセロクエストにおける非常に強いプレイヤーの着手が似ていることが推測される．

## 7. 結論

本稿では，オセロにおいて複数の方策を重み付き幾何平均で融合することで，幅広い強さで既存オセロ AI よりも人間との着手一致率が高いオセロ AI を実現した．本稿で使用した方策は以下に示す 4 つのモデルである．

- 通常モデル：WTHOR で強い人間の着手を学習した Policy Network
- 石損モデル：WTHOR のうち石損が 2 以上である手でファインチューニングを行った Policy Network
- Egaroucid レベル 1
- Egaroucid レベル 21

WTHOR [12] およびオセロクエスト [14] におけるプレイヤーのすべての条件において，既存手法と同程度の強さを実現したとき，人間との着手一致率が同等以上であるオセロ AI を実現できた．また，WTHOR データ全体では，Egaroucid レベル 21 と通常モデルを融合すると，それぞれの単体での着手一致率よりも着手一致率が向上する傾向が確認できた．

石損モデルと Egaroucid レベル 1 はほぼ同じ強さであったが，強い人間プレイヤーのミスした手や弱い人間の着手との一致率が高いのは石損モデル，強い人間プレイヤーの着手全体との一致率が高いのは Egaroucid レベル 1 であった．そのため，弱い棋力を実現する場合や強いプレイヤーを模倣するなら Egaroucid レベル 1 と通常モデルを融合する手法，弱いプレイヤーを模倣するなら石損モデルと通常モデルを融合する手法を使うことが望ましいとわかった．

本稿の手法では，融合に使用するモデルと融合度合いを調整するパラメータ $\alpha$ を調整することで，様々な強さのオセロ AI を作成可能である．局面の形勢によってパラメータを調整すれば，Fujita の取り組み [6] と同様の手法で動的な棋力調整に応用可能である．

**謝辞**　本研究では，Mindwalk 株式会社よりオセロクエストの棋譜データをご提供いただきました．深く感謝申し上げます．

### 参考文献

[1] Buro, M.: The Othello Match of the Year: Takeshi Murakami Vs. Logistello, ICCA Journal, Vol.20, No.3, pp.189-193 (1997).

[2] Takizawa, H.: Othello is solved, arXiv preprint arXiv:2310.19387 (2023).

[3] McIlroy-Young, R., Sen, S., Kleinberg, J., et al.: Aligning Superhuman AI with Human Behavior: Chess as a Model System, Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp.1677-1687, doi: 10.1145/3394486.3403219 (2020).

[4] Liskowski, P., Jaśkowski, W. and Krawiec, K.: Learning to play Othello with deep neural networks, IEEE Transactions on Games, Vol.10, No.4, pp.354-364, doi: 10.1109/tg.2018.2799997 (2018).

[5] Ogawa, T., Hsueh, C.-H. and Ikeda, K.: More Human-Like Gameplay by Blending Policies From Supervised and Reinforcement Learning, IEEE Transactions on Games, Vol.16, No.4, pp.831-843, doi: 10.1109/tg.2024.3424668 (2024).

[6] Fujita, K.: AlphaDDA: strategies for adjusting the playing strength of a fully trained AlphaZero system to a suitable human training partner, PeerJ Computer Science, Vol.8, p.e1123, doi:10.7717/peerj-cs.1123 (2022).

[7] 山名 琢翔，星野 准一: オセロ学習支援ロボットのインタラクションと展示からの設計示唆, 日本デザイン学会研究発表大会概要集, Vol. 2026, No. 73, セッション ID C-05-01 (2026).

[8] Jacob, A.P., Wu, D.J., Farina, G., et al.: Modeling Strong and Human-Like Gameplay with KL-Regularized Search, International Conference on Machine Learning, pp.9695-9728 (2022).

[9] Hsueh, C.-H. and Ikeda, K.: Improvement of move naturalness for playing good-quality games with middle-level players, Applied Intelligence, Vol.54, No.2, pp.1637-1655, doi: 10.1007/s10489-023-05210-2 (2024).

[10] 山名 琢翔: Egaroucid for Console（オンライン），入手先〈https://www.egaroucid.nyanyan.dev/ja/console/〉(2026.10.22).

[11] Buro, M.: Experiments with Multi-ProbCut and a new high-quality evaluation function for Othello, NEC Research Institute, Technical Report No. 96 (1997).

[12] Fédération Française d'Othello: La base WTHOR (online), available from 〈https://www.ffothello.org/informatique/la-base-wthor/〉(2026.10.22).

[13] Berg, M.: Welcome to xot.xmav.eu (online), available from 〈https://berg.earthlingz.de/xot/index.php?lang=en〉(2026.10.22).

[14] Mindwalk 株式会社: Mindwalk 株式会社 オンラインで対戦できる頭脳ゲーム・学習アプリをつくっています．将棋クエスト，囲碁クエスト，オセロクエスト，五目クエスト，激戦！英単語．（オンライン），入手先〈https://www.mindwalk.co.jp/〉(2026.10.22).

© 2026 Information Processing Society of Japan