

ハーモニーサーチアルゴリズムと ニューラルネットワークによるオセロ戦術の学習

(Othello Strategies Learned By Neural Network and Harmony Search)

1

A1178647 岡田 良健

Index

- 研究概要
- 機械学習
- ニューラルネットワーク
- ハーモニーサーチアルゴリズム
- ゲーム木
- オセロ
- 実験

研究概要

➤ 目的

大きな複雑度を持つ問題に対し, 専門知識をベースとせずに機械学習によって状況評価を学習すること

➤ 方法

- 十分複雑な問題の題材 : オセロ
- オセロAIの状況評価の機械学習 : ニューラルネットワーク
- ニューラルネットワークの重みの最適化 : ハーモニーサーチアルゴリズム

機械学習

- ▶ 機械学習とは
人間の知能的能力をコンピュータで実現することを目指す分野または技術
- ▶ 人間の知能の三つの能力
 - ▶ 知識の獲得をする**学習**能力
 - ▶ 物事の特徴やパターンを**認識**して**理解**する能力
 - ▶ 蓄積した経験や知識をもとに新たな問題の解決方法を**推論**する能力
- ▶ コンピュータに入力される**データ** = 人間の**経験**
数多くのデータ → コンピュータは上記の能力を習得, 発揮



ニューラルネットワーク

- 生物神経システムにおける**神経ネットワークの構造と機能**を模した数学モデル

各ノード：ニューロン(神経細胞)

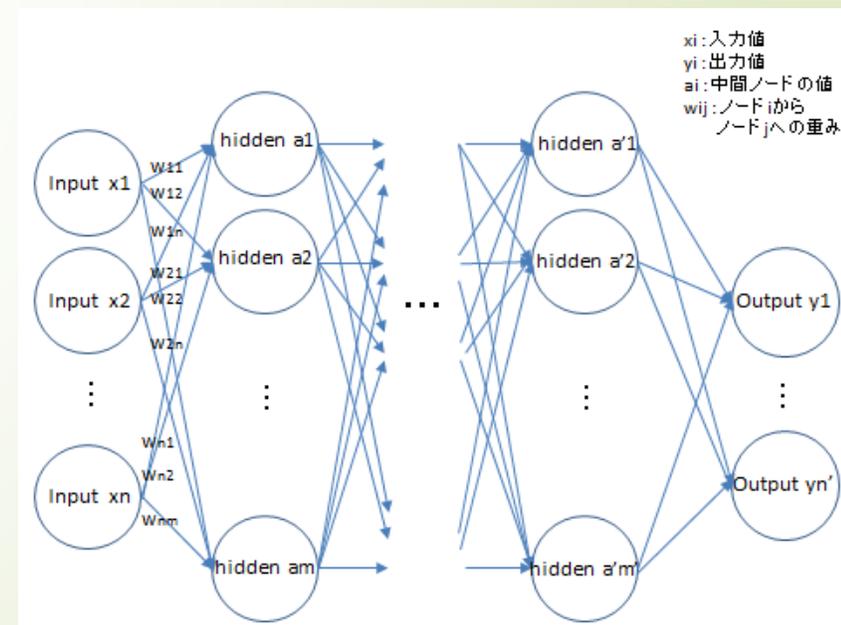
ノード間の枝：シナプス

ノード間の枝の重み：結合の強さ

- n次元のデータのベクトルの入力に対して、何らかの変換を行い、m次元のベクトルを出力

学習の結果、

求めたい**未知の関数を近似する関数**となる



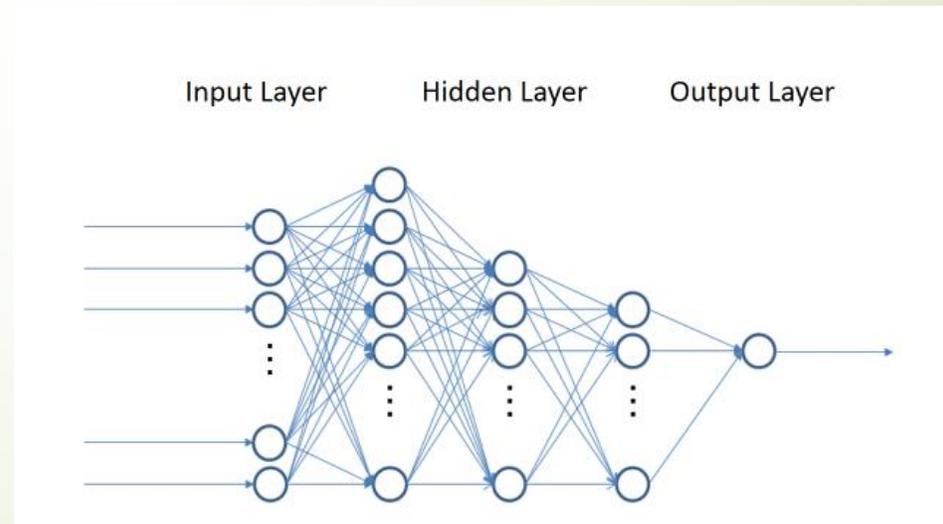
ニューラルネットワーク

- 階層型ネットワーク (∈ニューラルネットワーク)

階層構造を持ち, 入力層から出力層へと
一方向のみに信号が伝搬するニューラルネットワーク

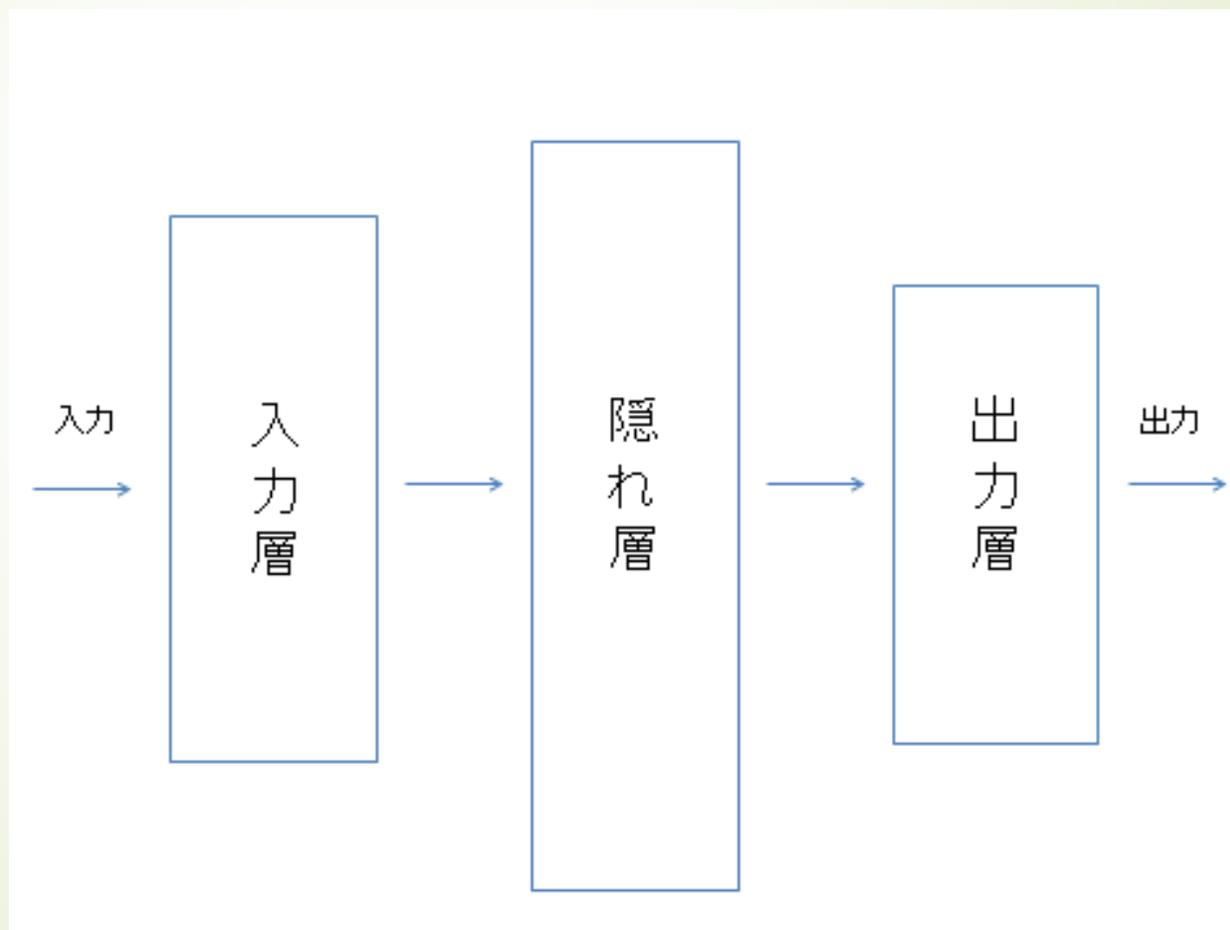
- 多層パーセプトロン (∈階層型ネットワーク)

隠れ層 (入力層と出力層の間の層) がある階層型ネットワーク



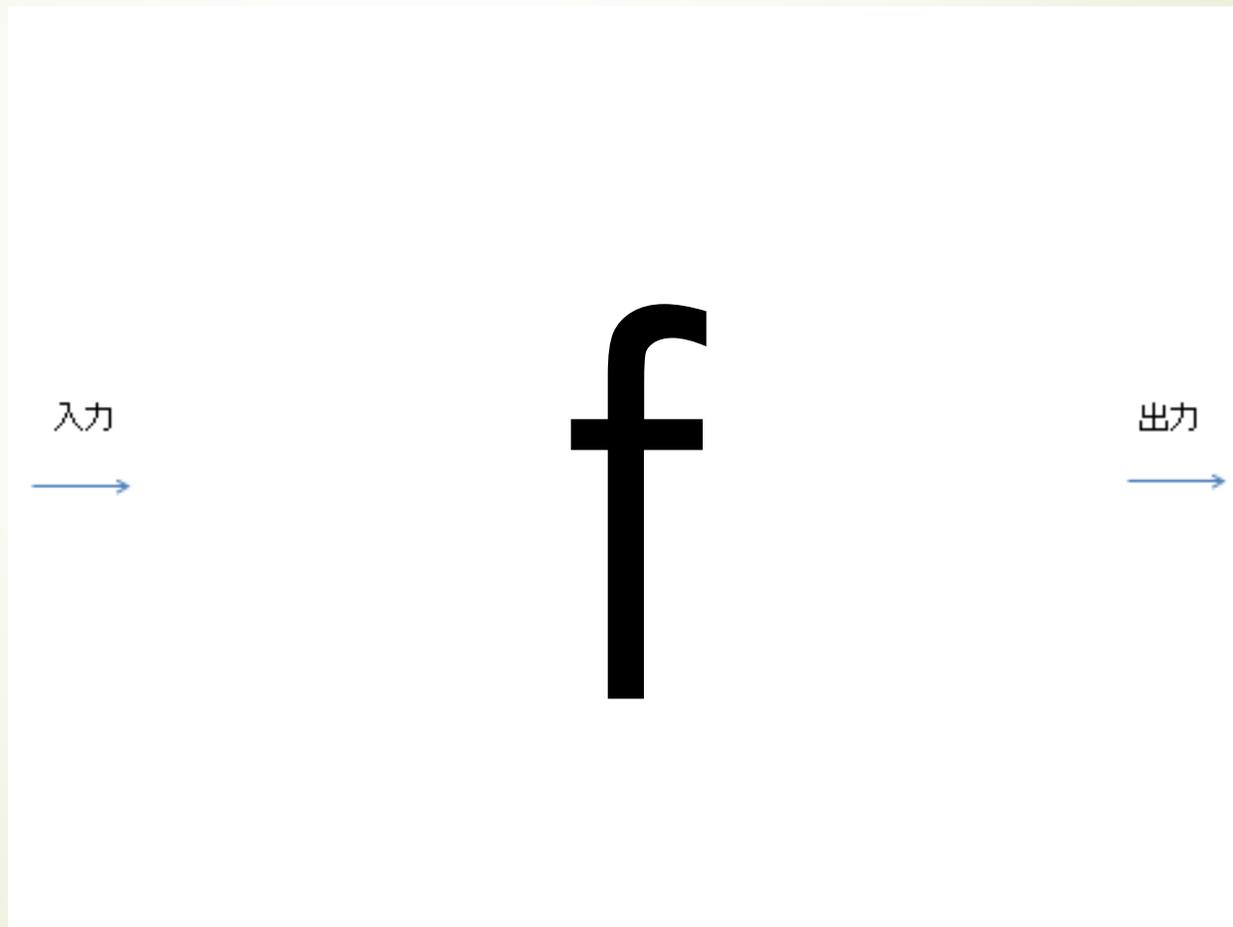
ニューラルネットワーク

つまり



ニューラルネットワーク

つまり



ハーモニーサーチアルゴリズム

- ▶ 個体群ベースのメタヒューリスティックアルゴリズムの一種
任意の問題に対して近似的に解を求めるアルゴリズム
- ▶ 音楽家の調和の探し方をアイデアのベースにして作られた
音符：決定変数
ハーモニー：解ベクトル
最高のハーモニー：準最適解ベクトル, 大域最適解
→ ハーモニーを改善し、最高のハーモニーを求める
- ▶ 構造がシンプル
- ▶ 組み合わせ最適化やパラメータ最適化に適している



ハーモニーサーチアルゴリズム

Optimize (minimize or maximize) $f(A)$: (1)

Subject to:

$h_i(A) = 0; i = 1, \dots, p;$: (2)

$g_i(A) \geq 0; i = 1, \dots, q;$: (3)

$a_i \in X_i = \{ x_i(1), x_i(2), \dots, x_i(K_i) \}$ or $x_i^L \leq a_i \leq x_i^U$: (4)

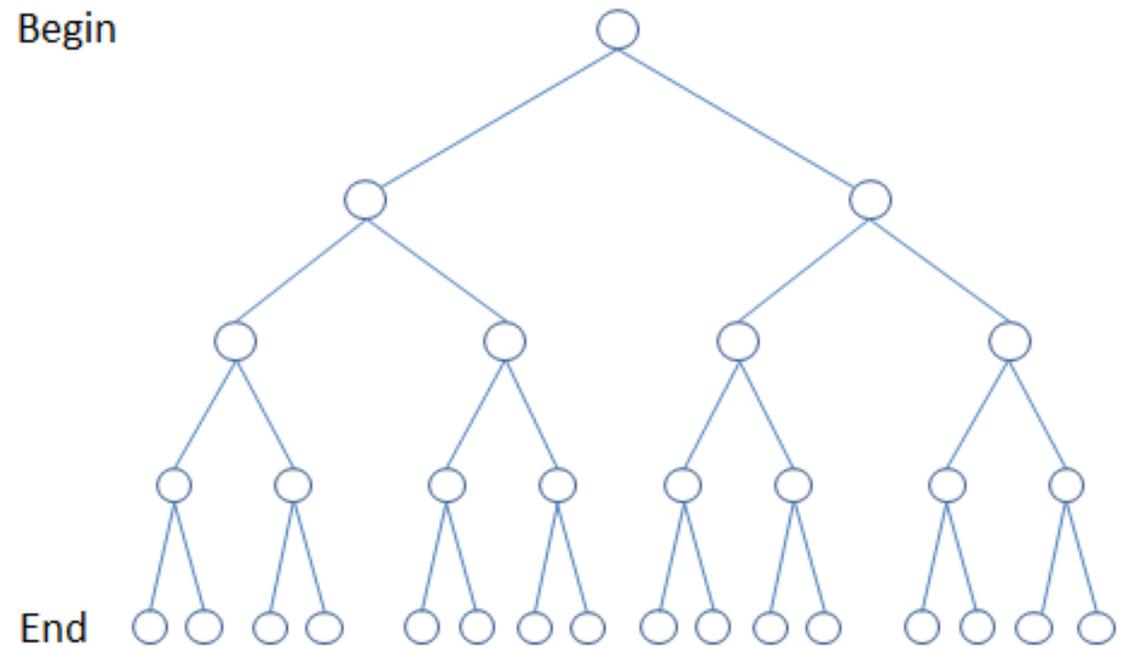
(1):目的関数, (2)~(4):制約条件 { (2):等式, (3):不等式, (4):値の範囲 }

与えられた制約条件の下で目的関数に対して
準最適解ベクトル $A = \{ a_1, a_2, \dots, a_n \}$ を求める

ゲーム木

- ▶ ゲームの展開を木構造で表したものの
- ▶ ゲームの**先読み**に利用
深さ：何手先読みか

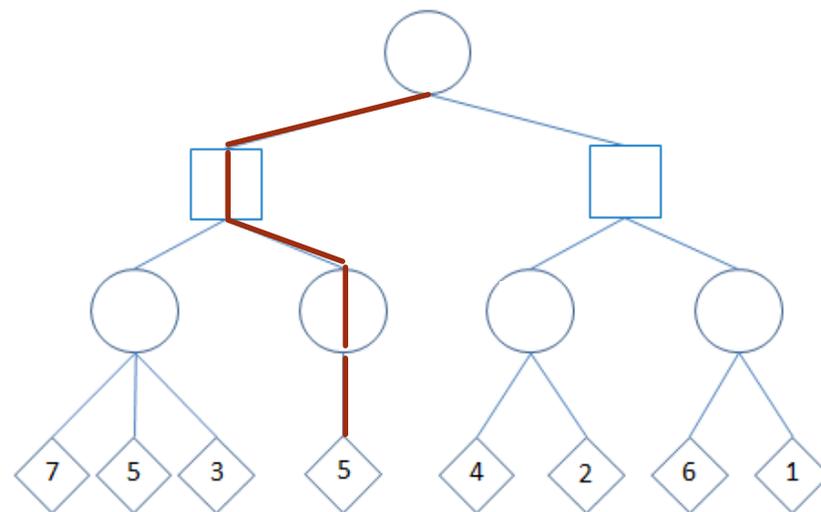
Begin



End

ゲーム木

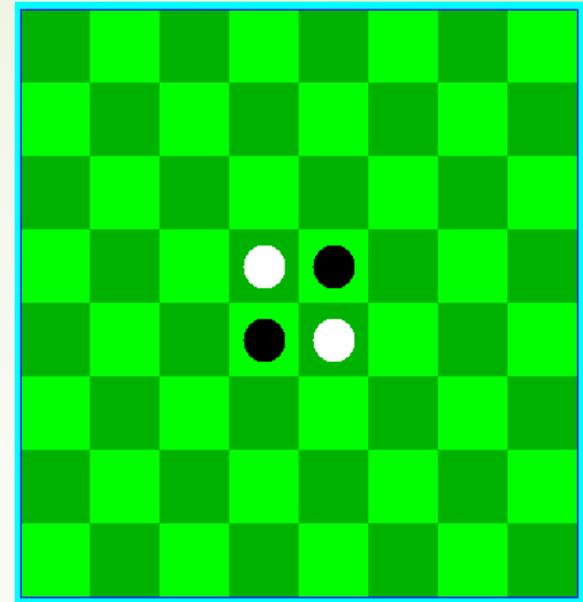
- ▶ ゲーム木探索アルゴリズム
 - ▶ **最良の手を探す**ための方法
 - ▶ 先読み先の状況の価値を見て判断
 - ▶ Minimax法, $\alpha\beta$ 法



数値：先読み先の状況の価値

オセロ

- ▶ 縦横8×8のマス目のある盤と白黒の石を使い
2人のプレイヤーが対戦するボードゲーム
- ▶ プレイヤーの**目的は勝利**
プレイヤーは各局面において優勢になるように行動
- ▶ ゲームの複雑度は人間がゲーム木の全展開を把握可能な程度を越えている
ゲーム探索空間は 10^{60} 程度
(三目並べ: 10^3 未満, チェッカー: 10^{30} , チェス: 10^{120} , 将棋: 10^{220} 囲碁: 10^{360})



実験

▶ 目的

オセロに対して専門知識をベースとせずに
機械学習によってその状況評価を学習すること

▶ 方法

- ▶ オセロAIの状況評価の機械学習：ニューラルネットワーク
- ▶ ニューラルネットワークの重みの最適化：ハーモニーサーチアルゴリズム
- ▶ 対戦時の先読み能力： $\alpha\beta$ 法
- ▶ エージェント間の対戦を繰り返して優秀なものを残す

エージェント：それぞれのハーモニーのこと。

また、その重みベクトルデータを元に形成されたニューラルネットワークを
頭脳として持つ、学習済みオセロAIのこと

実験

▶ ニューラルネットワークの詳細設定

▶ [パラメータ設定]

- ・ 入力層のノード数 input layer size : 68
- ・ 中間層1のノード数 hidden layer size : 208
- ・ 中間層2のノード数 hidden layer size2 : 60
- ・ 中間層3のノード数 hidden layer size3 : 20
- ・ 出力層のノード数 output layer size : 1
- ・ バイアスの数 bias size : 1 (bias = -1)

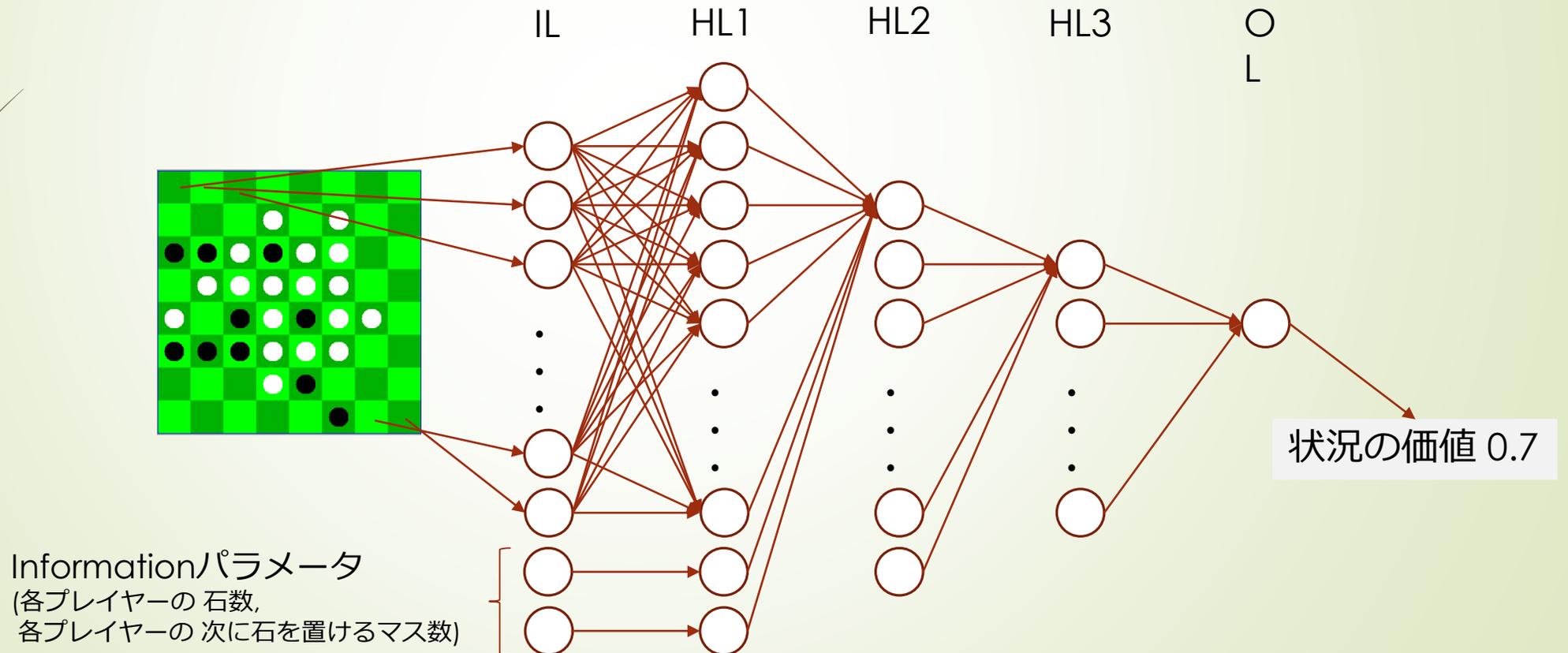
▶ [活性化関数]

- ・ シグモイド関数

$$f(x) = 1/(1+e^{(-ax)}) , (a = 1)$$

実験

ニューラルネットワークの設定



実験

- ▶ ハーモニーサーチアルゴリズムの詳細設定
 - ▶ [パラメータ設定]
 - MI (maximum improvisation) : 1000
 - hms (harmony memory size) : 128
 - hmcr (harmony memory considering rate) : 0.95
 - par (pitch adjusting rate) : 0.3
 - fw (fret width, formerly bandwidth) : 0.005
 - | Allow Range | ≤ 1

実験

- ▶ ハーモニーサーチアルゴリズムの詳細設定

- ▶ 目的関数 F

- ▶ F : エージェント集合内での相対的強さ

- ▶ Fの計算

- ▶ 以下を指定した試合数繰り返す

- ▶ 第一試合の前にFを0で初期化する

- ▶ 個体間の試合の勝敗から以下の点数を得る

- ▶ 勝利 : +10

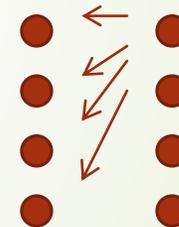
- ▶ 負け : + 0

- ▶ 引き分け : + 1

- ▶ 試合数で割る

- ▶ Fの最大化を目指す

- ▶ 勝利を目指す



実験

▶ ゲーム木探索アルゴリズムの詳細設定

▶ $\alpha\beta$ 法

- ▶ 深さ2 (次の次の自分の手まで)
- ▶ 先読み先の状況の価値
 - ▶ ニューラルネットワークにより算出した値
- ▶ 先読み先の状況の価値の評価に, オセロの対称性を利用し元々の局面を 90° 回転, 180° 回転, 270° 回転させたものも活用
 - ▶ オセロの局面評価の安定性の向上とともに, 学習の促進が狙い

実験

▶ オセロ対戦の設定

▶ [パラメータ設定]

実験における学習のためのオセロの対戦のパラメータは以下の通りで、

- ・ 対戦相手の数 MaxOppNum : 11
- ・ 対戦相手との試合数 MaxGametimes : 1

である。

実験

▶ 実験結果

▶ 対観測者

▶ 観測者：

学習課程で、改善がきちんと行われているかを確認するために
対戦相手用に用意したAI

▶ AI (No.1)：自分の石が一番多くなる手を最良の手とするもの

▶ AI (No.2)：盤の各マスの価値を知識として持ち、
石の位置による局面評価の値が一番大きくなる手を最良の手とするもの

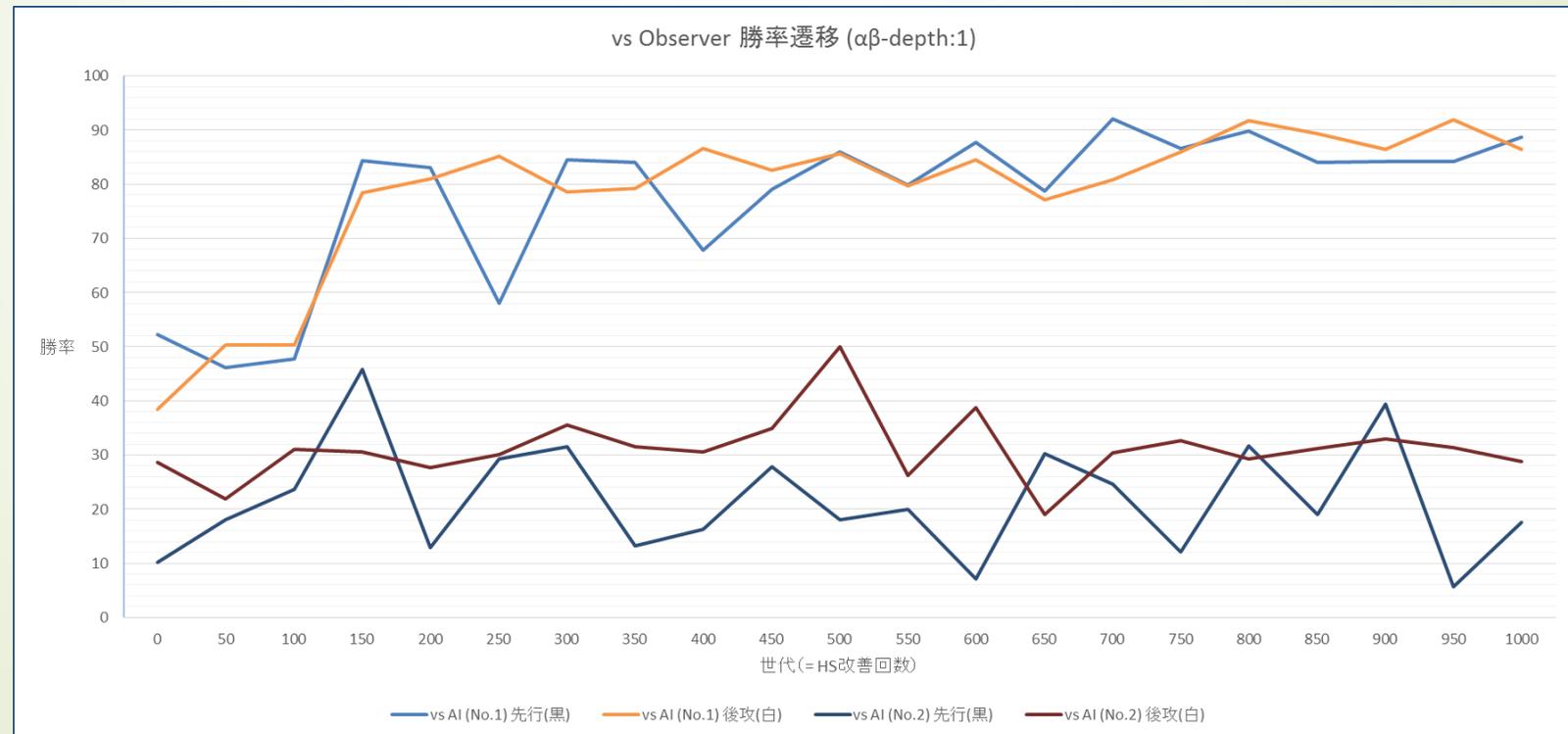
▶ **優良な上位3エージェントの勝率の平均**を結果として記録

▶ 学習の過程の50世代ごとに各観測者と50戦

▶ 組み込む $\alpha\beta$ 法の深さ1(= $\alpha\beta$ なし)の場合と深さ2の場合の両方で観察

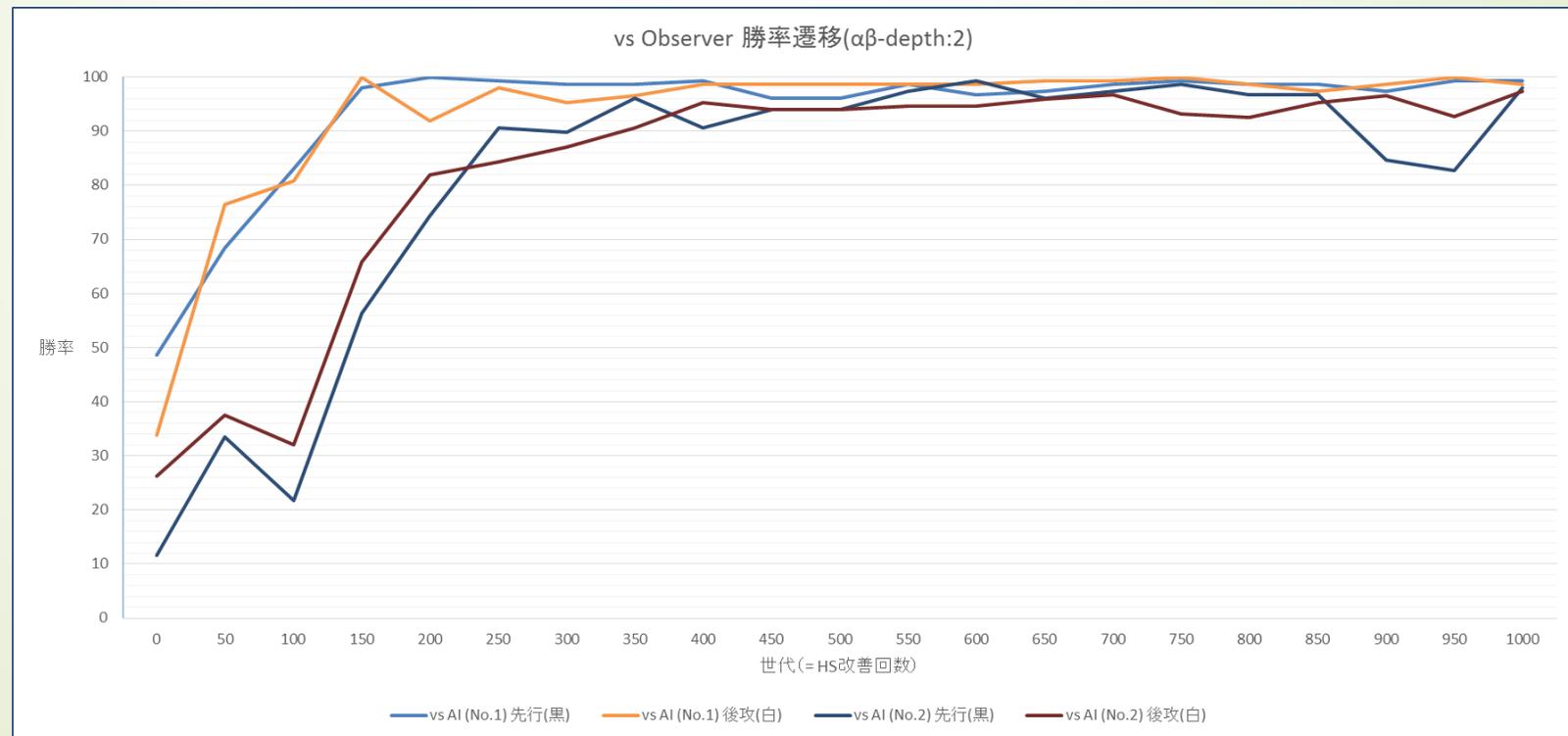
実験結果

対観測者



実験結果

対観測者



実験

- ▶ 実験結果

- ▶ 対人

- ▶ **最優良な個体の勝率の結果を記録**

- ▶ 36人に対して各約3戦ずつ, 計126戦

- ▶ 対戦相手の人は無作為に選ばれた大学生

実験結果

対人

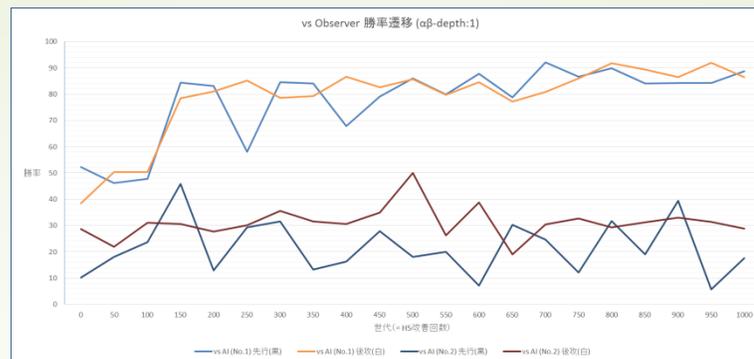
表1 対人試合結果の集計

総対戦者数	総試合数	勝利数	敗北数	勝率(%)
36	126	101	25	80.15873016

表2 対人（経験なし, 経験少ないを除く）試合結果の集計

総対戦者数	総試合数	勝利数	敗北数	勝率(%)
24	79	57	22	72.15189873

実験



考察

対観測者

- 対観測者
 - ある相手に対して勝率が上昇しても、他の相手に対して勝率が下降している過程が存在
 - ある相手の戦術に強い戦術は、他の相手の戦術に通用するとは限らない
 - $\alpha\beta$ 探索ありの場合の対戦成績の改善度 > $\alpha\beta$ 探索なしの場合の対戦成績の改善度
 - $\alpha\beta$ 探索がAIの強さを上げるために優秀
 - 学習する戦術が $\alpha\beta$ 探索ありの、2手先読みをした場合に効率的な戦術に進化した
 - 対戦時の先読みの程度によって学習した戦術の価値が異なる
 - 先攻と後攻の戦術の学習の改善度の差は特にない
 - 450世代あたりから勝率の変化が停滞
 - 観測者が改善度を観測できなくなった
 - 本実験の実験方法で学習できる戦術の限界に到達した可能性
 - 勝率(ゲームの強さ)の観点では同等であるが、戦術としては新しいものを学習し続けている可能性

実験

- 考察

- 対人

- 全試合の勝率は80.16%, 経験が少ない人との対戦を除いても72.15%
 - オセロの戦術を十分に学習したといえる

実験

■ 考察

■ 戦術面

- 後の展開を考慮しない悪手が稀に存在

■ 原因

■ 先読み

- 先読みが足りていない

- 学習時と学習後の対戦時, 共に $\alpha\beta$ 探索法のゲーム木の探索の深さ2(次の次の自分の手まで)

- “ゲームの強さ”を求めるためには, 状況評価の学習だけではなく, ゲーム木の探索が重要

■ 学習過程における個体間の対戦において出現頻度の低かった状況

- 後の展開が悪くなる典型パターンは避ける

- オセロの戦術知識におけるウイングの形など

- 後の展開が悪くなる局面を敗北により学習している

- ある世代において出現頻度が高い状況を一旦学習したとしても後の世代でその出現頻度が低くなった場合には, それを忘れてしまう可能性

- 複雑な問題に対して個体間の対戦による問題の本質の学習では, 問題の中で稀少な場合にも対応した学習は容易ではない

ご清聴ありがとうございました