

4K-7

対戦ゲームにおける評価関数の学習 —ニューラルネットワークを用いた方法—

宮坂大也 中西正和

慶応義塾大学 理工学研究科 計算機科学専攻

1. はじめに

チェスや将棋などの対戦ゲームにおいて強い手をコンピュータに選ばせるという課題は、多くの人によって研究されている。

MiniMax 法 [1] などを用いコンピュータに手の先読みをさせる場合には、局面の良さを評価する評価関数が必要になる。しかし、ゲームによっては、評価関数を人間が適当に設定することが難しい場合がある。

本研究では、より良い評価関数を作り上げていく方法を考え、その手法を○×ゲームと五目並べに適用した。

2. 評価関数の学習

本手法では、評価関数を階層型ニューラルネットワークを用い表現し、そのニューラルネットワークをある方針に基づき学習させることにより、評価関数をより適したものに変わっていく。

2.1 評価関数の表現

評価関数は、ある局面が“先攻プレイヤーにとってどのくらい有利か”を表すものとし、以下のものから構成する（図1参照）。

- n 個の階層型ニューラルネットワーク $N_1, N_2, \dots, N_n$
- n 個の写像 $H_1, H_2, \dots, H_n$
- 関数 G

N_i ($i = 1, 2, \dots, n$) を「評価ネットワーク」と呼ぶことにする。評価ネットワークの出力層のユニットは1個である。 H_i は、局面の情報を N_i への入力ベクトル

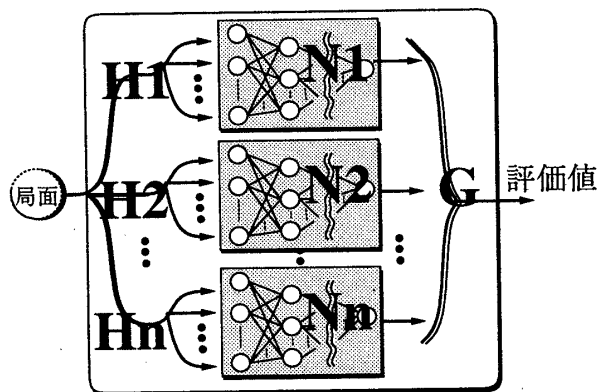


図1: 評価関数の表現

ルに変換する写像である。 G は $N_1, N_2, \dots, N_n$ の出力を一つの実数値に変換する関数であり、その値が局面の評価値となる。なお、 n, H_i ($i = 1, 2, \dots, n$), G は、対象とするゲームによって固定して考える。

2.2 評価ネットワークの学習

上で述べた評価関数を基に MiniMax 法を用い手を決めるプレイヤーを「ニューラルプレイヤー」と呼ぶことにする。ニューラルプレイヤーに与える初期の評価ネットワーク内の結合荷重は乱数で決める。

そして、ニューラルプレイヤーの評価ネットワークをより適したものに変わっていくために、次のことを行なう。

- 1) ニューラルプレイヤーを色々なプレイヤーと一定回数対戦させる。
- 2) 対戦中に現れた局面を基にゲーム木を作る。
- 3) ゲーム木の中から適当な学習データを抽出する。
- 4) 学習データをニューラルプレイヤーの評価ネットワークに学習させる。

4) は、Back Propagation 法 [2] を用いて行なう。評価ネットワークを学習させる方針は次の二つである。

- 評価関数が、先攻プレイヤーにとって有利な局面に対しては、大きな値を出力するようにする。

- 評価関数が、後攻プレイヤーにとって有利な局面に対しては、小さな値を出力するようにする。

この方針に合うような学習データを3)において抽出するのである。

以上のことを繰り返し行なうことにより、ニューラルプレイヤーの評価ネットワークは、学習した局面に対しては適した判断をするようになる。

3. 実験

2.で述べた手法を○×ゲームと五目並べに適用し、評価関数をより良いものに変えていく実験をした。実験において、評価ネットワークは3層に限定し、中間層のユニット数が30,80,130である評価ネットワークを持つニューラルプレイヤーを各々10人ずつ用意し、そのグループ毎に本手法を適用した。

○×ゲームに対する実験における、中間層のユニット数が130個である、10人のニューラルプレイヤーの「強さ」の変移を図2に示す。「強さ」は以下のプレイヤーとの対戦成績の、(勝率(%))+(引き分け率(%))である。

- 人間が作った評価関数を基に MiniMax 法を用いた手を決めるプレイヤー
- 選択できる手の中からランダムに手を決めるプレイヤー

全ニューラルプレイヤーのうちの最も弱いプレイヤー、平均的なプレイヤー、最も強いプレイヤーに対する、(1)何も学習していない時の強さと(2)本手法を適用した後の強さについて、表1(○×ゲーム)と表2(五目並べ)に示す。

	(1)	(2)
最も弱いプレイヤー	24	45
平均的なプレイヤー	32	78
最も強いプレイヤー	40	99

表1: 強さの推移 (○×ゲーム)

	(1)	(2)
最も弱いプレイヤー	49	50
平均的なプレイヤー	50	64
最も強いプレイヤー	44	97

表2: 強さの遷移 (五目並べ)

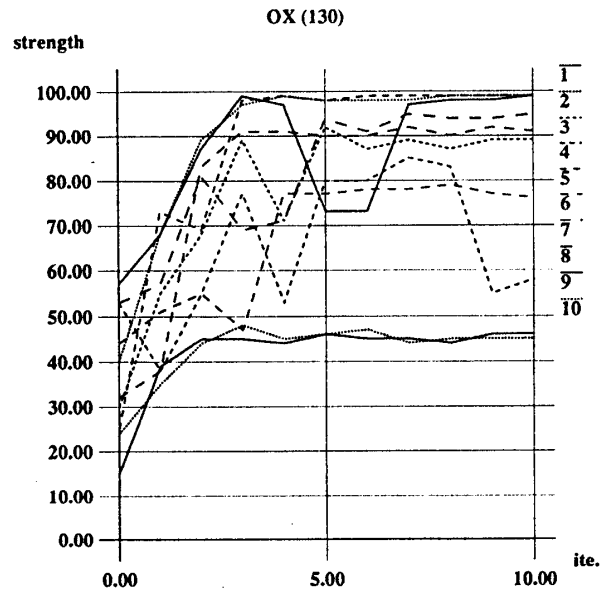


図2: あるニューラルプレイヤー10人の強さの変移 (横軸:手法の適用回数,縦軸:強さ)

4. 考察

評価ネットワークが新しい局面について学習したのにも関わらず、弱くなる場合があった。これは、Back Propagation法を用いた結果、学習していない局面に対する出力値が学習する前の時よりも不適切になることがあるためだと思われる。

5. 結論

対戦ゲームにおける評価関数を学習させる方法を考え、その手法を○×ゲームと五目並べに適用した。その結果○×ゲームにおいては平均で、78%の割合で勝つ又は引き分けるような評価関数を作り上げることが出来た。五目並べにおいては平均で、64%の割合で勝つ又は引き分けるような評価関数を作り上げることが出来た。

参考文献

- [1] P. H. Winston: ARTIFICIAL INTELLIGENCE, Addison-Wesley Publishing Company, 1977.(長尾, 白井訳: 人工知能, 培風館)
- [2] R. Beale, T. Jackson: Neural Computing:An introduction, IOP Publishing Ltd, 1990.(八名ほか訳: ニューラルコンピューティング入門, 海文堂)