ニューラルネットワークによる将棋評価関数の学習

今 龍　鈴木 康広　大堀 隆文　渡辺 一央
北海道工業大学

概要

コンピュータ将棋において、評価関数部分に階層型ニューラルネットワーク（ＮＮ）を用い、既存の棋譜に基づいた誤差逆伝搬法（ＢＰ）による自動学習法を提案した。プロの棋譜から、中盤戦の局面とその局面における着手を取材し、それらをデータセットとして学習および学習結果を検証する評価実験に用いた。また、評価実験は次の一手問題によっても行った。その結果、設定した評価項目と探索深さの内容に対して充分な性能を獲得することができた。

Learning of Shogi evaluation function by a neural network

Kon Ryu　Suzuki Yasuhiro　Ohori Takafumi　Watanabe Kazuhisa
Hokkaido Institute of Technology

Abstract

We propose a technique for learning an evaluation function in Shogi using the BP neural network(NN). Teacher signals are a set of the pair of a position and a move in the position based on the professional games. We prepared one hundred data sets from professional player's games used by learning experiment and one hundred data sets from next-one-move problems from shogi magazine. And we computed a correct answer rate for these sample patterns. By this experiment, it is concluded that the proposal method is effective for learning the shogi evaluation function.

１．はじめに

将棋系のゲームを人間がプレイする時、着手すべき局面、あるいは先読みの中に出現した局面において、何らかの基準による静的評価を行うことにより、下位の手を切り捨て、最善の一手を決定する。コンピュータ将棋においても、静的評価関数の役割は局面の正しい順位付けにある。

しかし、序・中・終盤といった局面の進行度や、居飛車・振り飛車といった戦型によって、評価関数で考慮すべき項目の重み付けは異なっていると考えられる。また、探索の中に相手戦略を盛り込む場合においても、相手戦略の獲得という行程が不可欠である。

このような、時と場合に応じた戦略を評価関数に組み込もうとする際、人間が手作業によってそれを行うのは非常に手間がかかり、非現実的である。そこで、対象とするプレイヤの棋譜という、相手の特徴をよく表したデータに基づいて自動学習する方法が考えられる。

今回我々は、３層ＮＮを評価関数に用い、その評価関数を利用した探索による着手とデータセットの登録着手との一致情報に基づいたＢＰ学習法を提案する。

2．手法の特徴

この手法は、ＮＮを静的評価関数として利用し、既存の棋譜に基づいてその評価関数をチューニングする。目的は、利用する棋譜に記録された着手の性質をＮＮの結合係数に反映し、実験システムを同じ性質の手を指せるようなプレイヤにすることである。

棋譜がある人物によるものばかりならば、実験システムはその人物のモデルを目指す。棋譜がある戦型に限定されているならば、実験システムはその戦型に特化した性質を目指す。このように、データセットの作り方によって学習の方向性を制御できる。

設定した項目によっては、棋譜の着手の性質を正しく評価できず、学習を失敗するパターンが出てくることが予想される。また、探索が着手を導いた思考に対して浅い場合は、やはり学習の失敗が予想される。

しかし、学習はＢＰによって行われるため、失敗の学習も有効に機能することが期待できる。そうして学習した結果は、棋譜に出現していない局面に対してもある程度の対応ができる汎化性能を持つと考えられる。

3．実験手順

学習システムはデータセットから１つのサンプルを読み込み、探索によってその局面における自分の着手を決定する。

得られた着手をサンプルの着手と比較し、正解・不正解を判定する。完全一致のみを正解とし、それ以外はすべて不正解とする。

学習のための教師信号は、次のように決定する。

$$\begin{cases} 1 : \text{着手が一致} \\ 0 : \text{着手が不一致} \end{cases}$$

教師信号に基づき一度だけ逆伝搬した後、学習回数に達していなければ次のサンプルを読み込み、同じ操作を繰り返す。

逆伝搬のためには、学習すべき局面における項目値を再度ＮＮに入力し出力を算出する必要がある。今回学習局面として、データセットの登録局面から実験システムが探索によって決定した手順の、葉の局面を利用する。

4．実験システムの設計

4．1　データセット

使用するデータセットのサンプルは、棋譜の中に出現した局面とそこで実際に指された着手の対で表現される。局面は、将棋盤の上にあるすべての駒の状態（生・成）と所有者（先手・後手）の情報を持ち、各手番が所有するすべての持ち駒を種類ごとの枚数として持つ。着手はＣＳＡの標準形式にしたがう。

データセットの棋譜は、日本将棋連盟発行の月刊誌「将棋世界」に連載されている『タカミチの将棋実戦コーナー』[1]より取材したプロ公式戦のものである。

序盤は定跡に頼るところが大きく、終盤は詰みの計算が重要であるため、棋譜の中盤よりサンプルを取った。中盤の判断は著者らの直感により行った。

今回は 2296 パターンを収集し、その中の 1000 パターンを学習用、残った中の 100 パターンを評価用とした。学習手法が成立することを検証するため、戦型や個人を特定せずランダムにパターンを選択した。

さらに、学習した結果の性能を確認する評価実験のため、「将棋世界」に過去連載されていた『やさしい次の一手』[2]よりサンプルを 100 パターン用意した。これは、棋譜の中の手という絶対の一手とは必ずしもいえないサンプルによってのみ評価したのでは、一致しないが間違いではない手を正解に数えられないことから、絶対の一手が存在する局面での評価が必要と考えたためである。

次の一手問題は、中盤戦と判断でき、７手まで読めれば決定できそうなものを選択した。

4．2　評価関数

静的評価関数として、３層ＮＮを用いた（Fig.1）。評価項目の数に応じた入力層素子を持ち、１つの出力層素子を持つ。このＮＮの出力に微小乱数を加えたものを評

価値とした。

図の黒丸は線形素子、白丸はシグモイド素子である。しきい素子は用いていない。

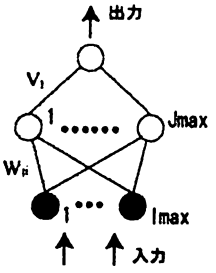

Fig. 1　ＮＮ評価関数模式図

４．３　評価項目

今回は、以下に示す５つの項目を用意した。中盤戦で特に注目されそうなものを意識している。

１．総駒価値
２．総利き数
３．優利き数
４．ただ取り
５．成り

項目値は、先後それぞれで得点を計算した後に、手番側の得点から非手番側の得点を引いた差分値としている。

総駒価値は、インターネットから入手した、ある市販ソフトの技術解説に掲載されている駒価値の表に基づき計算する。

総利き数は、盤上の駒が利いているマスの数である。マスの数であるから、複数の駒が利いていてもカウントは重複しない。

優利き数とは、あるマスについて先後両方の利きがあり利いている駒の数に差がある時、数で勝っている方に１ポイントを与えるものである。

ただ取りはただで取れる駒の駒価値をポイントとする。

成りは次に一手あれば成れる駒の、（成りの駒価値－生の駒価値）をポイントとする。

ただし、ただ取りと成りは手番によって得点の計算が異なる。仮に、評価する局面が先手番であったとすると、先手はすぐにでも駒を取るか、成るかを実現できるが、後手は先手にその手を阻まれる可能性がある。そこで、この場合は後手の得点を２分の１で計算することにする。

５．シミュレーション実験

５．１　実験条件

実験用評価関数として、上の項目のうち総駒価値・総利き数・優利き数の３項目を使用する「利き数関数」、総駒価値・優利き数・ただ取りの「ただ取り関数」、ただ取り関数＋成りの「成り＋関数」の３種類を用意した。

学習実験に際しては、次のような条件を設けた。

Table 1　学習実験条件

学習係数	0.1
サンプル数	1000
結合係数初期値	-0.5～0
探索	全幅３手

学習は繰り返し無しで、1000 サンプルを学習したら終了とした。

評価実験については、1000 回学習終了時の結合係数を用いて 100 サンプルの正解率を観察する。この時の探索は以下の３種類とした。

Ａ　全幅３手
Ｂ　全幅５手
Ｃ　７手（80-60-40-20-10-10）

Ｃの７手読みでは、カッコ中に示したように探索幅を制限する。左の数字が根ノードの幅であり、以下深さ順に数値を示してある。

比較のため、２つの市販ソフトによっても評価用サンプルの正解率を計った。

５．２　学習実験結果

1000 回の学習を 50 回ごとに区切り、区切った範囲の正解率の推移を Fig.2 に示す。横軸に学習回数、縦軸に正解率を取った。

利き数関数のグラフでは、上昇傾向は見られるがほぼ 0.1 付近で収束している。ただ取り関数では、一部極端に下がっているものの、始めから一定の値を保っている。

最後に成り＋関数の結果だが、下降傾向が見られる。しかし最終的には、ただ取り関数と同程度の正解率に落ち着いている。

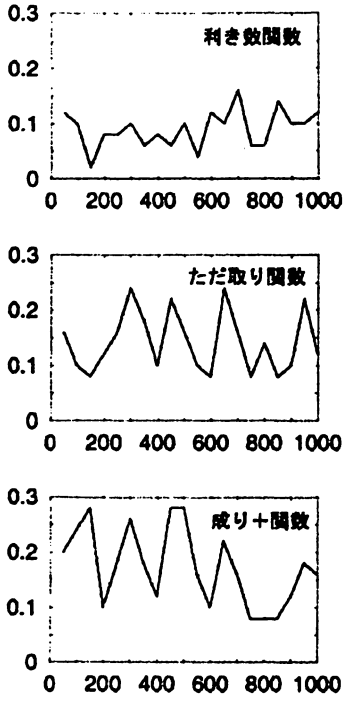

Fig. 2　学習時正解率推移

5．3　評価実験結果

Fig.3 に示すように、棋譜データの正解率は市販ソフトの最強レベルにだいぶ劣っている。しかし、市販ソフトは評価関数が我々のシステムよりも複雑であり、探索も深いことを考えれば、その半分ほどを正解する性能は高く評価できると思われる。

Fig.4 の次の一手では、市販ソフトの最高値と実験システムの最高値がさらに開いている。これは、７手を完全に読んでぎりぎり正解できるレベルの問題が多数あるために、探索の深い市販ソフトでは正解できても、実験システムが正解できていないということが考えられる。

6　おわりに

既存の棋譜を利用し、ＢＰによってＮＮ評価関数のチューニングを行う手法を提案し、実験により有効性を検証した。

その結果、市販ソフトには及ばないものの、基本的な手については学習できていることを確認できた。水平線効果や項目の不備により明らかな疑問手を指す局面もあり、項目の追加・探索の工夫などによって改良してゆきたい。

また、学習係数の設定や次善手データの利用による学習の精度向上も今後の課題である。

市販システム		正解率	実験システム(7手)	正解率
A将棋		39	利き数	15
K将棋	Lv.7	40	ただ取り	21
	Lv.4	43	成り＋	25
	Lv.1	43		

Fig. 3　棋譜データ評価実験

市販システム		正解率	実験システム(7手)	正解率
A将棋	標準	19	利き数	25
	最強	26	ただ取り	20
K将棋	Lv.7	57	成り＋	20
	Lv.1	10		

Fig. 4 次の一手評価実験

参考文献

[1]『タカミチの実戦コーナー』(「将棋世界」1994 No.2～1997 No.3、日本将棋連盟)

[2]『やさしい次の一手』(「将棋世界」1995 No.1～1996 No.12、日本将棋連盟)