ニューラルネットワークを用いた囲碁の評価関数の設計

永吉宏之　等々力賢
東京大学大学院工学系研究科システム量子工学専攻
naga@lyman.q.t.u-tokyo.ac.jp

概要
本研究では、囲碁の評価関数のモデルとしてユニット間の接続を近傍のみに限定しパラメータの共通化を行うことにより対称性を考慮したニューラルネットワークを提案した。プロの棋譜を用いて学習を行い、結果について検証を行った。対局終了時における石の死活や途中局面に対する地の予測確率で良好な結果が得られ、本手法の妥当性が示された。

Design of Evaluation Function using Neural Networks in the game of go

Hiroyuki Nagayoshi, Masaru Todoriki
Department of Quantum Engineering and Systems Science,
School of Engineering, The University of Tokyo

Abstract
In this paper, we propose a neural network evaluation function in the game of go. The features of our neural network are local connection of its neural units and parameter share for considering symmetry of go board states. We train our neural network with professional game records, and then we obtain good results in the territory estimation and life and death distinction at the endgame.

1．はじめに

情報処理の分野において、オセロ、チェス、囲碁などの二人完全情報ゲームは、研究の対象としてよく用いられてきた。その中で囲碁は、コンピュータにとって最も難しいゲームとされている。その理由として、一局面あたりの可能な指し手が多いため探索空間が広い、明確な評価法が存在しないなどが挙げられる。そのうち探索空間が広いという特徴は、囲碁の持つ性質であり、探索手法を工夫することにより読みの省略は可能だが、本質的には探索空間を小さくすることはできない。一方で精密な評価ができれば、浅い探索でも強いプログラムを作ることが可能になるため、評価関数の改良は探索空間の広大さを補う方法として有力である。そのため、評価関数の研究というのは強い囲碁プログラムを作るために重要となっている。

実際に囲碁の評価関数を作るとき、２通りのアプローチを考えることができる。ひとつは囲碁の強い人の評価方法をプログラム上で実現することであり、もうひとつは対局や問題集を用いて機械学習などにより評価関数を作成するという方法である。前者の場合、強い人の大局観をアルゴリズム化しなければならないが、本人でさえもどのように評価しているか明確にできない要素もあり、完全に再現することは難しい。後者の場合、囲碁の評価関数として適当な関数モデルを与え、そのパラメータをチューニングすることにより学習させることになる。この場合、最初にどのような関数を与えるかによって、評価関数の近似精度の限界が決定され、関数をどのような手法で学習させるかにより、うまくパラメータをチューニングできるかが決定される。そのため、関数のモデルと学習手法により学習が成功するか失敗するかが決まってくる。

そこで本研究では、囲碁の評価関数のモデルとしてユニット間の結合を近傍のみに限定し、パラメータの共通化を行うことによって対称性を考慮した多階層ニューラルネットワークを提案し、それに対しプロの棋譜を用いた学習を行うことにより評価関数を作成し、その有効性を示すことを目的とする。

2. ニューラルネットワークの構造

囲碁の評価関数を設計する場合、関数の出力として、単に評価値のみを出力する関数と、盤上の各々について地になる確率を求め、その合計を評価値として出力する関数が考えられる。本研究では、棋譜から得られる情報を有効に活用するために、後者を採用した。

本研究で用いるニューラルネットワークとして、多階層フィードフォワード型ニューラルネットワークを用いた。ユニット間の接続に工夫を凝らすことにより、局面の対称性などを考慮したニューラルネットワークを設計した。

ニューラルネットワークの入力として局面情報を数値化したデータを用いる。数値化は黒白別々に行い、それぞれ局面上の各点に対して石の有無により1と0を割り当てる。盤のサイズをＮとすると、入力層のユニット数は$2N^2$となる。

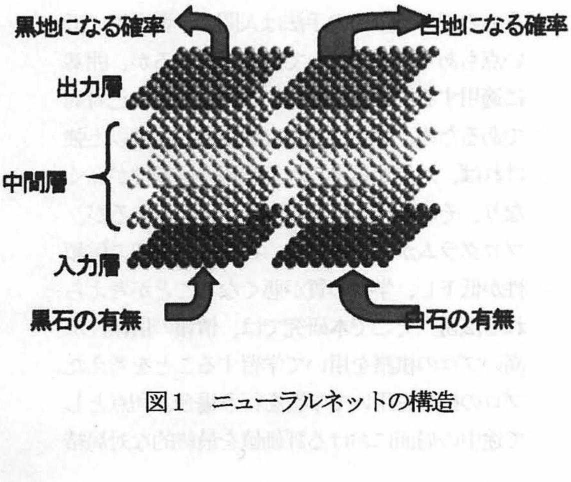

図1　ニューラルネットの構造

ニューラルネットワークの出力は、盤上の各々の点における地になる確率である。そのため出力層のユニット数は$2N^2$となる。

中間層のユニット数は自由に取ることができるが、本手法では入力層や出力層と同じ構造をもつように$2N^2$個とした。このような構造により、局面の対称性の考慮が容易になる。

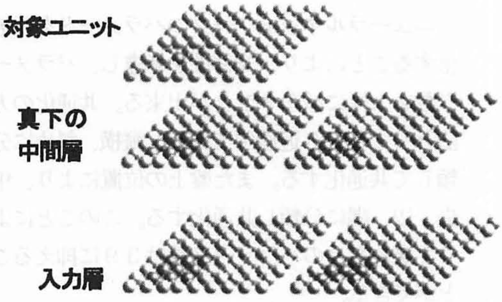

図2　ユニット間の接続

ユニット間の接続は、上方のユニットから見て、一段下の中間層と入力層のユニットとのみ接続し、各々の黒白に属する３×３近傍に位置する３６ユニットと接続している（図２）。このような構造により、ユニットの持つ情報の直接的影響を受ける領域は近傍のみになり、それ以上遠距離へは、情報は逐次的に伝達する。中間層の数が１つ増えるにつれ、距離が１つ遠い場所へ影響が及ぶようになり、例えば中間層が３層の場合には距離４までを考慮することができる。ユニット間の接続を近傍のみにした場合のメリットとデメリットを下表に示す。

	近傍接続型	全接続型
関数近似	一般的に悪い	非常に良い
パラメータ	少ない	多い
計算速度	速い	遅い

表1　近傍接続型の特徴

ニューラルネットワークのパラメータを共通化することにより、対称性を考慮し、パラメータ数を大幅に減らすことが出来る。共通化の方法として３×３近傍中で真下、縦横、斜めに分類して共通化する。また盤上の位置により、中央、辺、隅に分類し共通化する。このことにより１層あたりのパラメータ数は３９に抑えることができる。

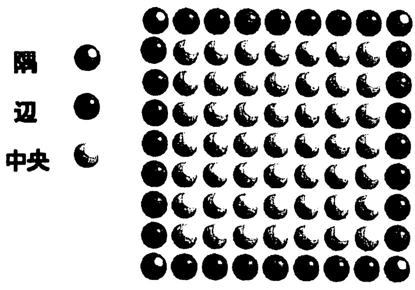

図3　パラメータの共通化

その他の処理として局面上において同一の連に属するユニットの出力値を、そのまま出力とせず、連内で平均化することを行う。この処理により同一の連に属する石は死活を共にすることを表現できる。

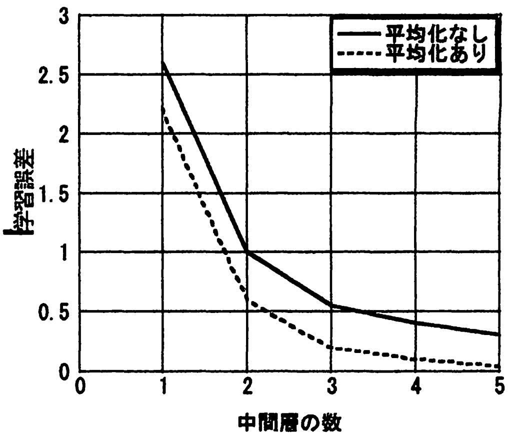

図4　連内の出力平均化の効果

3．棋譜を用いた学習

評価関数を学習させる方法として、最も理想的なのは教師データとして真の評価値を用いることである。しかしながら、囲碁などのゲームにおいて、真の評価値が得られるのは、終局直前の局面に対してだけであり、事実上不可能である。そのため、実際の学習には、真の評価値に変わる指標が必要となる。そのひとつの手段としてＴＤ学習[1]などの強化学習がある。強化学習は直接的な教師データを必要とせず、経験（対局など）を通して適応的に学習を行うというものである。この手法は人間の学習過程に近い点もあり学習法としては優秀であるが、囲碁に適用する場合、経験に相当するのが自己対局であるため、プログラム自身がある程度以上強ければ、対局から得られる情報の信頼性が高くなり、その結果、良好な学習が可能であるが、プログラムが弱い場合は、得られた情報の信頼性が低下し、学習の質が悪くなることが考えられる[2,3]。そこで本研究では、情報の信頼性が高いプロの棋譜を用いて学習することを考えた。プロの棋譜を用いて学習を行う場合、利点として途中の局面における評価値を最終的な対局結

果から推定しても大きな誤差は生じないと仮定できる点がある。欠点として学習できる局面が棋譜上に現れる局面に限定され任意の局面に対して学習を実行できない点がある。本研究においては棋譜中の局面に対して、事後確率に相当する対局終了時の地の状態を教師データとする学習を行った。これにより、単に評価値だけでなく、局面上の模様や勢力を学習することが可能になる。

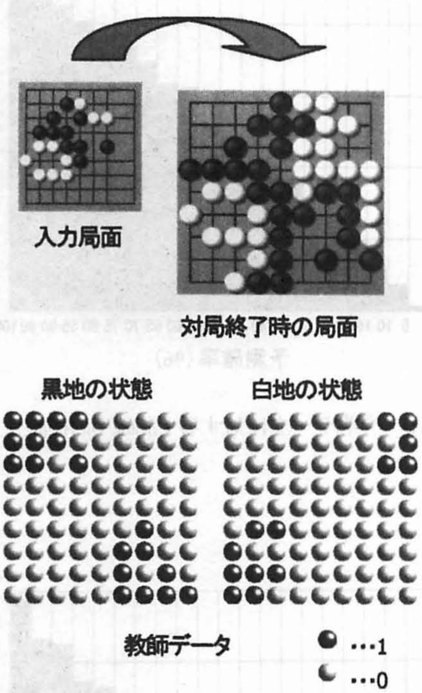

図5　教師データの与え方

実際の学習は、評価関数の良さを判断する基準となる目的関数を作成し、それが最小になるようにパラメータをチューニングすることにより行う。ニューラルネットワークの学習の場合、目的関数のパラメータによる勾配情報はバックプロパゲーションを用いることにより容易に求めることができる。得られた勾配情報をもとに、非線形最適化手法を用いることにより目的関数を最小化するパラメータを決定する。通常、ニューラルネットの学習には最急降下法が用いられることが多いが、多階層ニューラルネットワークに適用した場合、学習速度が遅く有効ではない[4]。そこで本研究では準ニュートン法[5]を用いる。準ニュートン法は、ニュートン法で必要とされるヘッセ行列（目的関数の２階微分行列）を勾配情報から適応的に求めることにより最適化を行う手法である。一般に準ニュートン法は最急降下法と比べ学習速度が速い(図６)。

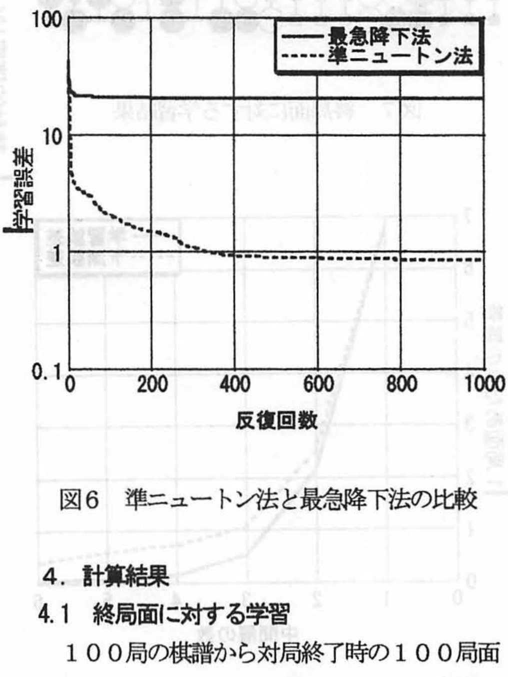

図６　準ニュートン法と最急降下法の比較

４．計算結果

4.1　終局面に対する学習

１００局の棋譜から対局終了時の１００局面を抽出し、そのうちの８０局面に対して学習を行い、残り２０局面を検証用データとした。学習の反復回数は１００００回、ニューラルネットワークの中間層の数を１～６とした。その結果、中間層の数が増えるにつれ学習誤差が少なくなり、中間層が５層以上の場合において、石の死活、欠け目をほぼ正しく判断することが出来た。また予測誤差についても、中間層の数が増えるに従って減少しており、過学習の傾向は見られなかった。結果を図に示す。

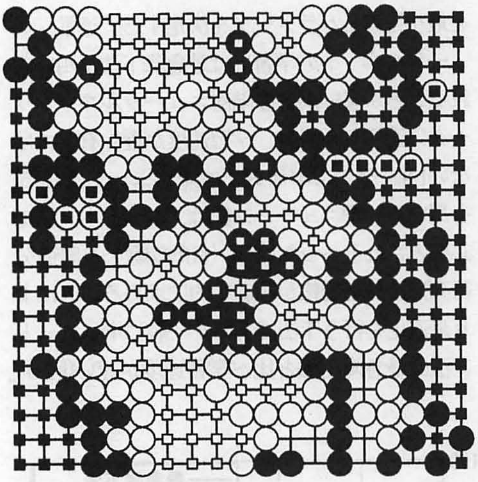

図７　終局面に対する学習結果

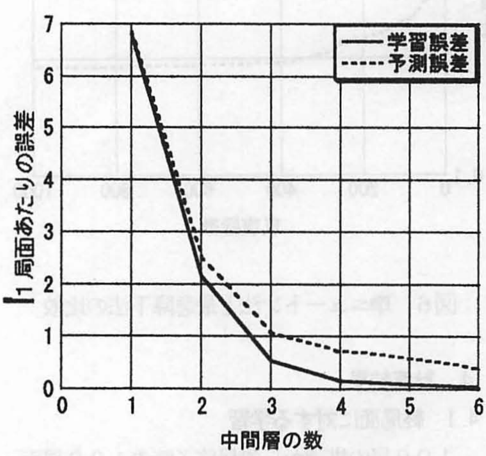

図８　中間層数と誤差の関係

4.2　対局途中の局面に対する学習

５０局の棋譜を用意し、そのうち３０局の棋譜を用い、棋譜中の局面すべてに対して学習を行った。その結果に対し、学習で用いていない残りの２０局を用いて統計的な検証を行った。統計的検証法として、予測した確率を５％ごとに区切りヒストグラムをとり、それぞれに対し求められる統計的な確率と比較した。例えば黒地になる確率が５０％の点を１００点集め、それらの点に対応する終局面上の点が５０程度黒地になっていれば予測確率は妥当であると考えることができる。検証の結果、本手法により得られた確率と棋譜により得られた統計的確率を比較するとよく一致し、本手法により妥当な評価が得られることがわかった。

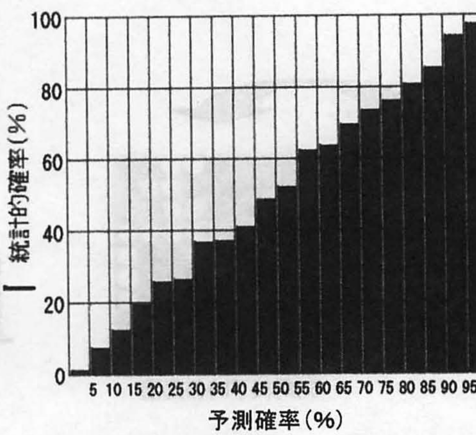

図９　学習データに対する予測確率の検証

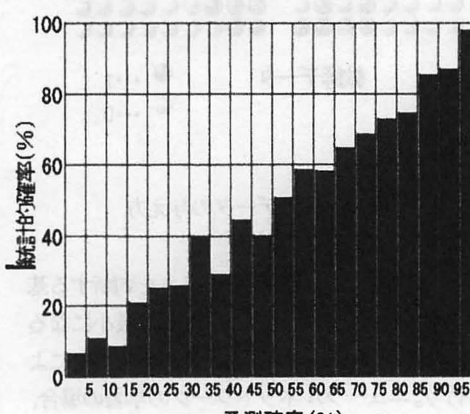

図１０　非学習データに対する予測確率の検証

5．考察

本研究で用いたニューラルネットワークはユニット間の接続を近傍のみに限定し、さらにパラメータの共通化を行っている。これにより通常であれば、ニューラルネットの持つ高い関数近似能力が得られないと考えられる。しかしながら、本ニューラルネットワークは、囲碁局面上の地になる確率をよく近似できることがわかった。このことは、石の影響力が近傍から遠方へ逐次的に拡散するという仮定が妥当なものであることを示しており、対称性を考慮してパラメータを共通化することの妥当性も示している。

6．今後の課題

本研究で用いたニューラルネットワークには、ユニット間接続の近傍のみの限定、対称性を考慮したパラメータの共通化、同一の連に属するユニットの出力平均化などの特徴を持ち、それらは有効に働くことがわかった。しかしながら、本来取り入れるべきものが残っている。連と連との関係、各連の持っているダメの数、各連がいくつ眼を持っているか、などの情報である。これらは石の死活に直接関わる重要な要素であり、この情報抜きに死活の判定を学習させるのは無謀とも言える。本研究で取り入れなかった理由として、これらの要素が場合によっては探索も伴うため時間が掛かる要素である点を挙げておく。しかし、これらの要素は非常に重要であるため、何らかの形で取り入れる必要があると思われる。

6．まとめ

本研究において、ユニット間の接続を近傍のみに限定しパラメータの共通化を行うことにより対称性を考慮したニューラルネットワークを囲碁の評価関数のモデルとして提案し、プロの棋譜を用いて学習を行った。その結果、終局時における石の死活判定と対局途中の局面における予測確率において良い結果が得られ、本手法の妥当性を示した。

参考文献

[1] R. Sutton, "Learning to Predict by the Method of Temporal Difference Learning", *Machine Learning*, 3, 9-44, 1988.

[2] M. Enzenberger, "The Integration of A Priori Knowledge into a Go Playing Neural Network", available from Internet.

[3] N. N. Schraudolph, P. Dayan, T. J. Sejnowski, "Temporal Difference Learning of Position Evaluation in the Game of Go", *Advances in Neural Information Processing 6*, Morgan Kaufmann, San Francisco 1994.

[4] D. P. Bertseks, J, N, Tsitsiklis, "Neuro Dynamic Programming", Athena Scientific, Belmont, M.A. 1996.

[5] 矢部博、八巻直一、"応用数値計算ライブラリ　非線形計画法", 朝倉書店　1999.