5H－6

# リカレントニューラルネットワークを基本とした先読みモデルの提案

佐藤裕二　落合辰男*　波多野祥二
新情報処理開発機構　*日立マイコンシステム

## 1. はじめに

実世界における問題では，適応システムや機械学習で必要となる試行回数や学習時間が十分与えられるとは限らない．学習過程における部分的な情報を基に有効な行動を必要とされる場合が考えられる．この問題の解決策の一つとして，本稿では，リカレント型ニューラルネットワークを基本とした先読みモデルを提案し，その有効性を検討する．ニューラルネットワークの構造および結合荷重の学習には，遺伝的アルゴリズム(GA)を用いる．tic-tac-toeゲームを用いてこのモデルの有効性の評価を行う[1].

## 2. 先読みモデルの提案

本稿で扱うリカレントニューラルネットワーは，環境を介した帰還ループを持つことを特徴とする．例えばフィードフォワード型のニューラルネットワークに時間遅れのある自己帰還ループを加えたJordan型の時系列発生モデル[2]に，環境を介した帰還ループを付加した場合を考える．本稿で提案する先読みモデルを図１に示す．提案するモデルは，先読みを実現するために，行動を決定するネットワーク (Action Network) の他に，環境の内部モデル (Internal Model) を持つ．いずれも，前記リカレントニューラルネットワークを基本とする．環境との応答を通して，環境の内部モデルを自己学習しながら先読みを行い，投機的な行動を外界に対して行う．先読み結果は行動を決定するネットワークの学習にも利用する．すなわち，提案するモデルは，行動を決定するネットワークと内部モデルが相互作用を持ちながら成長するモデルとなっている．

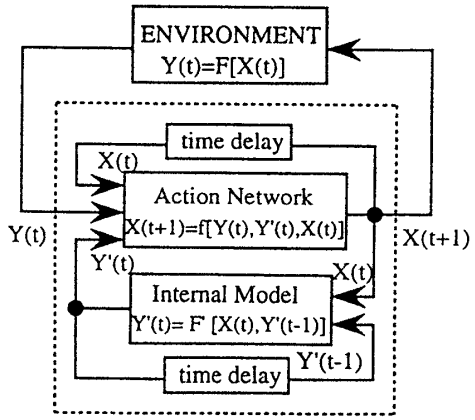


図１　提案する先読みモデル

Proposal of a Lookahead Planning Model based on Recurrent Neural Networks
Yuji Sato, Tatsuo Ochiai* and Shoji Hatano
Real World Computing Partnership
* Hitachi Microcomputer System LTD.

## 3. GAによるニューラルネットワークの構造および結合荷重の決定

### 3.1 行動を決定するネットワークの学習

時刻 t の出力と時刻 t の出力に対する環境からの入力の組み合わせに対して，可能性のある時刻 t+1 の出力を各々染色体とする．初期状態では，各染色体の個体数を同じに設定する．従って，最初はランダムにある染色体が選ばれる．学習は２種類の方法で行う．第１に，ある染色体が選ばれたことにより最終的にゲームに勝ったか負けたかという評価を行い個体数の増減を行う．第２に，内部モデルを用いた先読み結果の勝敗により個体数の増減を行う．

### 3.2 内部モデルの学習

ここでは，２種類の方式に関して検討を行った．

#### 3.2.1 学習方式１：アルゴリズム学習方式

学習方式１では，入力層，隠れ層群，出力層から構成される，層状結合型のニューラルネットワークを用いる．各ニューロンは，０または１の状態をとる．ある時点における入力信号の総和が０よりも大きい場合は自分自身の次の状態を１にし，それ以外の場合は０にする．ニューロン間の結合荷重は１，-１または０のいずれかをとる．０の時は，ニューロン間に結合がないことを表わす．次に，回路網の構造と結合荷重を表わす染色体を定義する．あるニューロンから別のニューロンへの結合荷重（１，-１または０）を要素とする遷移行列を染色体とする．

#### 3.2.2 学習方式２：入出力の対応表学習方式

学習方式２では，入力から出力への２次元遷移行列を染色体と考える．入力または出力として取りうる状態をそれぞれ異なるシンボルに割り当てる．

## 4. 実験

### 4.1 実験方法

tic-tac-toeゲームを用いてモデルの評価を行った．解析を簡単化するために，以下の仮定を行った．

（1）先手は，自分のシンボル○を，最初に中央に置く，後手は，Xを，最初に左上隅に置く．

（2）先手は，熟練者であり，負けることはない．

（3）後手は，先手のアルゴリズムを知らない．

（4）後手は初期状態として以下の知識を持つ．

（知識1）：自分の番の時に，自分のシンボルを3個続けて一直線上に置ければ置く．

（知識2）：（知識1）が適用できない時，もし次に相手が，相手のシンボルを3個続けて一直線上に置けるようなら阻止する．

（知識3）：（知識1），（知識2）共に適用できない場合は，空いている場所にランダムに打つ．

すなわち，後手は，行動を決定するネットワークの内，（知識1）と（知識2）に関しては予め学習済みであり，（知識3）の戦略を経験により更新する．上記仮定に基づくゲームの流れを図2に示す．

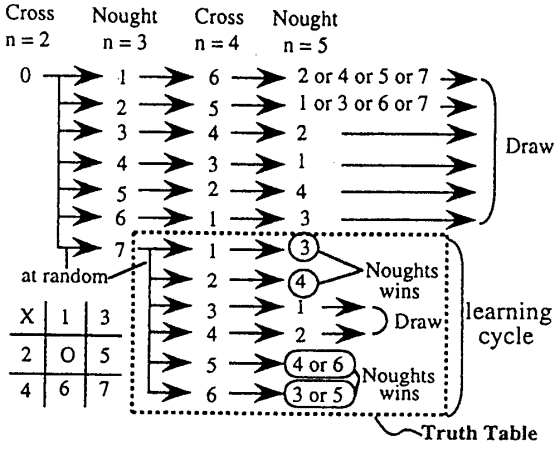


図2 （知識1）～（知識3）に基づくtic-tac-toeゲームの流れ

### 4.2 実験結果

以下の4つのケースに関して実験を行った．

（実験1）：（知識3）の更新を，先読み無しで行う場合．内部モデルは使用しない．

（実験2）：（知識3）の更新を，先読みも利用して行う場合．内部モデルは学習方式1を用いる．

（実験3）：（知識3）の更新を，先読みも利用して行う場合．内部モデルは学習方式2を用いる．

（実験4）：（知識3）の更新を，先読みも利用して行う場合．但し，内部モデルとして，先手の完全なアルゴリズムを最初から使用する．

試行回数と後手（シンボルX）が負ける確率の累積平均との関係の一例を，図3に示す．

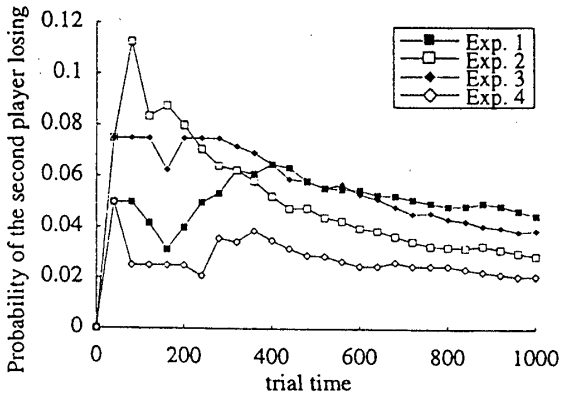


図3 後手が負ける確率（累積平均）と試合数の関係

## 5. 考察

（実験1）と他の3つの実験結果の比較から，内部モデルを用いた先読みが，勝率向上に有効であったことが分かる．先読み結果を学習に利用することにより，学習に必要な試行回数を削減できる可能性を持つ．（実験2）と（実験3）では，（実験2）の方が高い先読み効果を示す傾向にあった．この原因を調査するために，内部モデルの学習ターゲットの詳細な調査を行った．その結果，内部モデルの学習ターゲットはXOR論理を部分的に含んでおり，内部モデルは，ターゲットとなる真理値表のランダムな変化に伴い，XOR論理の生成と消滅を動的に繰り返す必要があったことが分かった．このことから，動的に変化する非線形論理の学習には，入出力の対応を学習する方式よりも，アルゴリズムを学習する方式の方が有効であると推測される．

## 6. おわりに

本稿では，リカレント型ニューラルネットワークを基本とした先読みモデルを提案し，tic-tac-toeゲームを用いて有効性の評価を行い，以下の結論を得た．（1）環境の内部モデルを用いた先読みにより，実際の環境を介した学習回数を削減できる．（2）内部モデルの学習方式として，アルゴリズムを学習する場合の方が，入出力の対応を記憶する場合よりも，先読み効果が高い可能性を持つ．今後，より深い先読みを必要とする問題での調査が必要である．

### 参考文献

[1] Sato, Y. et al. : Lookahead Planning and Co-Evolution in Recurrent Neural Networks, Proc. of ICEC'94-Orlando, pp. 764-769 (1994).

[2] Jordan, M. I. : Serial order : A parallel distributed processing approach, Advances in Connectionist Theory : Speech, Erlbaum (1989).