3G6-1

自己組織化回路素子へのフリップフロップ素子導入による時系列学習

Temporal Sequence Learning using Self-Organizing Network Elements with Flip-Flop Node

金　天海[*1]　出澤　純一[*1]　尾形哲也[*2]　菅野　重樹[*1]
Chyon Hae Kim　Jyun-ichi Idesawa　Tetsuya Ogata　Shigeki Sugano

[*1]早稲田大学 理工学術院 機械工学専攻
The Department of Mechanical Engneering Waseda University

[*2]京都大学 情報学研究科 知能情報学専攻
The Department of Intelligence Science and Technology Kyoto University

For the purpose of exploration in the universe or deep sea, the use of autonomous robot is expected. Such an autonomous robot needs learning method to learn unexpected situations, becuase expectation of such situations is difficult for robot creator. However, strict condition is required when we intend to use learning method in such a robot. Against these conditions, we porposed a learning system, Self-Organizing Network Elements (SONE).

In this paper, application of SONE for learning of long term sequence including hidden state is discussed. We introduced Flip-Flop node in SONE, and SONE could successfully learn long term temporal sequence.

1. はじめに

近年ロボット開発のなかで，遠隔操作が困難な宇宙や海底における探査・対人コミュニケーション等への応用を目的としたstand alone な自律型ロボットの開発が期待されている．このようなロボットの使用環境は，設計者にとって想定困難な状況を含むため，効果的な制御則を用意することが難しい．そこで，タスク・環境の中でロボット自身に制御則を獲得させる学習法に期待が寄せられている．

この場合，想定困難なタスク・環境に対応するために，学習系に幅広いタスク・環境へ応用できるロバスト性が求められる．さらには，ロボットが常にタスク・環境の変化にさらされるため，有意な時間において適応できる計算性能も必要となる．よってこのようなロボットには，タスク・環境毎の設定を省力化し，単純な外部パラメータにより実装された学習器によって，効果的な出力を有意な時間に探索・学習できる学習系が必要である（効果的な出力の自律的探索，単純な外部パラメータ，オンライン・リアルタイム性）．

従来効果的な出力の探索法として強化学習 (RL) に基づいた手法や，遺伝的アルゴリズム (GA) に基づいた手法が提案されており，外部パラメータの単純化も議論されている．

RL による手法としては Direct-Vision-Based Reinforcement Learning (DVB-RL) 等 [1] がある．これらの手法では，従来 RL の分野で一般に必要であったタスク・環境毎のセンサ状態空間分割を必要としない．しかし，タスク・環境毎に入力信号変換やニューラルネットワーク (NN) のトポロジー決定を必要とし，外部パラメータの単純化は十分とは言えない．

一方 GA による手法として NeuroEvolution of Augmenting Topologies (NEAT) [2] 等が提案されている．これらの手法では NN の構造と重みを GA によって決定し，状態空間分割，入力信号変換，トポロジー決定を必要としない．しかしながら，GA による手法では各表現型であるネットワークのテスト・評価に多くの評価時間を必要とするため，オンライン・リアルタイムな学習が困難である．

このような問題に対し，我々は自己組織化回路素子 Self-Organizing Network Elements (SONE) を提案し，ロボットシミュレーションによってその有効性を示した [6]．しかし，従来の SONE では長期の隠れ状態を含む時系列問題を扱うことが困難であった．SONE の応用範囲を広げ，高度な学習を実現するためには長時間の隠れ状態を含んだ時系列問題を学習することが必要であり，本稿では SONE によるネットワーク内にフリップフロップの機能を持つ素子を自己組織的に獲得させることでその解決を行った．

連絡先: 金天海，早稲田大学理工学部　菅野重樹研究室，東京都新宿区大久保 3-4-1 59 号館 319 室，03-5286-3264，tennkai@sugano.mech.waseda.ac.jp

2. 自己組織化回路素子

従来手法である MLRL，DVB-RL に代表される強化学習法では状態空間の分割，入力信号の変換，ネットワークトポロジーの決定といったパラメータをタスク・環境毎に設計者が設定する必要があった．また，NEAT に代表される GA による手法では評価時間に関する問題があった．以上の問題を解決する方法として著者らは自己組織化回路素子 Self-Organizing Network Elements (SONE) を提案し，オンライン・リアルタイムなネットワーク構造獲得による強化学習を実現している．

DVB-RL に見られるように，状態空間の分割，入力信号の変換法に関しては，NN のようなネットワーク型の学習器とロボットの間でセンサ入力，モータ出力を直結する手法が有効である [1]．そこで本手法も同様に，ネットワーク型の学習器を用い，それをロボットの入出力と直結する手法をとる．次に，学習を用いたネットワークトポロジーの決定について考える．NEAT のような GA に従った学習ではネットワーク全体を一個体として評価するため，その評価にはタスク・環境に応じた評価時間が必要となる．一方で，ネットワーク内の個々の素子を評価する場合，それと同等な評価時間は必要ない．なぜならば，DVB-RL や NEAT のようにネットワークがロボットの入出力と直結される場合，ネットワーク全体が学習するべき対象は，ロボットが学習するべきタスク全体と対応する．よって，ネットワーク全体の評価を行うためにはタスク全体に対する評価を行うだけの評価時間を必要とする．一方でネットワーク内の個々の素子を評価する場合，タスク全体に対する評価は必ずしも必要ではない．一般にネットワーク内の個々の素子はロボットの特定の状態において反応し，その状態に対する適切な出力の形成に寄与する．よってそれらの素子は，素子の代表している個々の状態において評価可能である．SONE ではネットワークの各素子に関する評価値を算出し，その評価値によってネットワークトポロジーを決定する手法をとることで，タスク・環境に応じた評価時間の設定を必要としないオンライン・リアルタイムな学習を実現している．

従来 NN の分野においても，このようなネットワークの各素子を評価してトポロジーを決定する手法が存在する [3, 4, 5]．しかしながら，このような手法の多くは強化信号を用いた学習が困難である，オンライン・リアルタイムな学習が困難である等の問題があるため，stand alone な自律ロボットへの応用は難しい．強化学習の枠組みでは，ロボットが実際に環境で探索

を行うことで報酬・罰に相当する強化信号を獲得し，学習器はその強化信号によって学習を行うことができる．そこで SONE では，この外部から与えられる強化信号を各素子毎の評価値に反映するために強化信号伝播規則を導入し，ネットワークの出力層から各素子へと強化信号を伝播することで各素子の評価を行っている．各素子はその評価に応じて新しい素子の生成，または自己解体を行い，ネットワークトポロジーの自動決定を促す．

SONE の特徴は，ネットワークの各素子が独立して活動することでローカルネットワークが構成され，そのローカルな活動がグローバルなネットワークを成長させることである．そして，ネットワークの各素子が報酬に対し貪欲（Greedy）に設計され，各素子の受け取る報酬量を増加させるようにネットワークが構築される枠組みであるため，SONE の設計は報酬量を安定的に増大させる回路素子の設計に帰着し，その設計がタスク・環境と独立したものとして扱える．筆者らは論理回路素子をベースとして SONE を構成しており，以下ではその具体的な実装法を述べる．

2.1 自己組織化論理回路

SONE を実装するための素子には多様な選択が考えられる．著者らは，その中でも比較的容易な対象である二値による論理回路素子へ SONE を実装した．SONE の概念を論理回路へ適用した自己組織化論理回路は AND ノード，OR ノード，反転リンク，非反転リンクより構成される．本稿ではこの中でも基本となる OR ノード (Fig.1)，非反転リンク (Fig.2) に関する素子の設計法を述べた後，それらの素子を用いたネットワークの構成法を述べる．

2.2 OR ノード

Fig.1 に示される OR ノードは，1 本のテストリンクと 2 本の実リンクを持っており，それぞれのリンクからの入力は X_T，$X(1)$，$X(2)$ で与えられる．テストリンクは各 OR ノードにつき必ず 1 本存在するが，実リンクの本数には制限が無く可変であり，一般に N 本の実リンクを持つことができる．出力生成時（出力フェイズ）には各 OR ノードは実リンクに対し OR 演算を行うことで出力 $Y = \bigcup_{i=1}^{N} X(i)$ を計算する．強化信号伝播時（伝播フェイズ）には各 OR ノードは Table1 に示されるルールに従って各リンクとその入力側ノードに対しそれぞれ強化信号 R_1, R_2 を伝播する．Table1 の各 Case は次のように機能する．例えば OR ノードが 3 本の実リンクを保持しており，それらの入力が $\{X(1), X(2), X(3)\} = \{T, T, F\}$ であるとする．このとき，OR ノードの出力は $Y = T$ となる．OR ノードがこの出力を行った結果負の強化信号 $R < 0$ を受け取った場合，F の出力を行っているリンク 3 は Case1 に相当し，T の出力を行っているリンク 1, 2 は Case5 に相当する（Table1 において N_T は実リンクのうち T を出力するリンクの数として計算される）．各リンクは各 Case に従って算出された R_1 を受け取り，各リンクの入力側ノードに R_2 を伝達する．R_1, R_2 を伝達された素子はそれらの信号を蓄え，自らの評価値である R 値にこれを加える．また，強化信号を伝達したノードは R 値を 0 にリセットする．テストリンクには，テストリンクが昇格，実用化された場合を想定して OR ノードの出力の算出を行い，Table1 を適用する．ただし，テストリンクの昇格によって OR ノードの出力が反転する場合には Table1 の結果算出される R_1 に-1 を乗じ，また R_2 は常に 0 とする．構造変更時（構造変更フェイズ）にはテストリンクの昇格判定が行われ，テストリンクの R 値がある閾値 $Th1$ を上回る場合，テストリンクの昇格，実用化を行う．また，OR ノードの保持するリンク数が 1 以下である場合には OR ノードは入出力の演算を保つようにネットワークを適宜つなぎ直し，自己解体する．

2.3 非反転リンク

Fig.2 に示される非反転リンクは，一つのテストノードを持っている．このテストノードは AND ノード，OR ノードの二通りの中から，非反転リンクの出力側ノードの種類と逆のノードを備えるように生成される．テストノードはさらに二本のテストリンク（TL1,TL2) を保持しており，TL1 は非反転リンクと同じノードに結合することで，非反転リンクと同様の入力を得る．また，TL2 はネットワーク内にある他のノードと結合している．出力フェイズには各非反転リンクは入力をそのまま出力として伝える（$Y = X$）．伝播フェイズには各非反転リンクは Table2 に従って R_T を計算し，テストノードへと伝播する．テストノードは伝播された R_T を用いて自らの伝播規則（例えば，テストノードが OR ノードをテストしているならば OR ノードの強化信号伝播規則）を用いて TL2 に伝播する．TL2 は伝播された強化信号を自らの R 値に加える．構造変更フェイズにはテストノードの昇格判定が行われ，非反転リンクの R 値がある閾値 $Th2$ を上回りかつ TL2 の R 値がある閾値 $Th3$ を上回れば，テストノードを昇格，実用化し，不要となった非反転リンクは自己解体を行う．また，非反転リンクの R 値が 0 を下回った場合には非反転リンクは自己解体する．

2.4 ネットワーク

これらの素子の出力フェイズ，学習フェイズ，構造変更フェイズを用いてネットワークの自己組織化を行うことができる．ネットワークの初期状態としてセンサ入力を受け付けるための入力ノードとモータ出力を行うための出力ノードを，ロボットに応じて必要な数用意する．これらのノードは OR ノードを用いて構成し，自己解体は不可とする．リスト構造を用いてこれらノードの管理を行い，リストの前方には入力ノード，後方には出力ノードを配置する．新しくできたノード（中間ノード）はその出力側に位置するノードの直前に挿入されることで，リストへ登録される．

ネットワークの出力計算の際には，このリストの前方から順に出力フェイズによってノードの起動を行い各ノードの出力を計算する．ネットワークに強化信号を伝播する際には，リストの後方から前方へ向かって順に伝播フェイズによってノードの起動を行い各ノードによる信号の伝播を行う．ネットワークの構造変更を行う際には，全てのノード，リンクにおける構造変更フェイズを起動するが，この際にはその順序を問わない．ネットワークの計算は出力計算，強化信号伝播，構造変更の順に繰り返し行われ，ロボットの行動する全てのタイムステップに関してネットワークの出力，学習，構造変更が行える．

このような学習フェイズ（強化学習フェイズ）に加え，著者らは教師あり学習フェイズも提案している [7]．こちらの教師あり学習フェイズは，ロボットの実際の行動以前に基本となるネットワークを構築することで，ロボットの学習初期における探索コストを低減することができる他，強化学習と較べネットワークの学習性能を定量的に評価しやすいという利点がある．

教師あり学習フェイズでは，外部より入出力ベクトルを受け取り，それに応じて学習を行う．SONE はネットワークの各素子が互いに評価値を算出し合い，素子の生成・淘汰を繰り返すことで学習する学習器であるため，外部からの評価も同様にネットワークの各素子に付与することで教師あり学習も可能である．

そこで，出力端子の出力と教師信号に誤差が無い場合にはその出力端子へ報酬として 1 を出力端子の出力と教師信号誤差が有る場合にはその出力端子へ罰として-1 を付加することで教師データの学習を実装している．この実装では，ネットワーク内部の報酬系は従来の強化学習フェイズと共有となるため，教師あり学習と強化学習の系をひとつのネットワークに共存することができると考えられる．

本稿ではこの教師あり学習フェイズを用いて時系列学習に対する評価を行う．

3. SONE へのフリップフロップ素子の導入

SONE は各素子の持つローカルルールによる逐次的なネットワーク構造自己組織化によって，ネットワークの受け取る強化信号を増加させる枠組みである．各素子は強化信号を伝播し合うことで互いの価値評価を行い，その評価に応じてネットワーク構造を変更する．ネットワークの各素子が強化信号に対し貪欲（Greedy）に設計され，各素子の受け取る報酬量を増加させるようにネットワークが再構築されるため，SONE の設計は報酬量を安定的に増大させる回路素子の設計に帰着し，

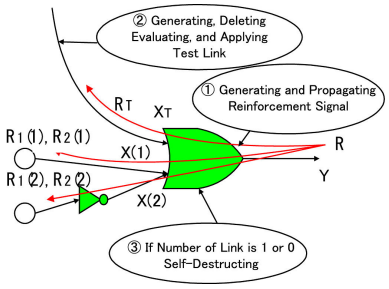


図 1: Or ノード

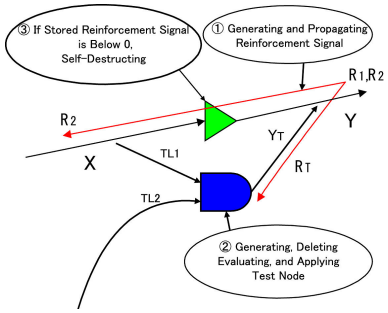


図 2: 非反転リンク

表 1: OR ノードの強化信号伝播規則

$Case1:$	$(Y = T) \wedge (X(k) = F)$
	$R_1(k) = 0, R_2(k) = 0$
$Case2:$	$Y = F$
	$R_1(k) = R/N, R_2(k) = R/N$
$Case3:$	$(Y = T) \wedge (N_T = 1) \wedge (X(k) = T)$
	$R_1(k) = R, R_2(k) = R$
$Case4:$	$(Y = T) \wedge (N_T \neq 1) \wedge (R \geq 0) \wedge (X(k) = T)$
	$R_1(k) = -R \times (N_T - 1)/N, R_2(k) = 0$
$Case5:$	$(Y = T) \wedge (N_T \neq 1) \wedge (R < 0) \wedge (X(k) = T)$
	$R_1(k) = R \times N_T/N, R_2(k) = 0$

その設計がタスク・環境と独立したものとして扱える．筆者らは論理回路素子を基本として SONE を構成しており，And ノード，Or ノード，反転リンク，非反転リンクを用いたネットワークを，ネットワーク外部からの強化信号に従って成長させることで，より多くの強化信号が得られるネットワークが形成されることを示している．

このようなネットワークにおいて時系列問題を扱う場合，主に二つの方法が考えられる．リカレントニューラルネットワーク（RNN) のようにフィードバックループを用いた学習と，メモリ機能を備えた素子による学習である．

従来 SONE ではフィードバックループの形成により，単純な時系列問題を扱えている．しかしながら，長期的な隠れ状態を記憶・保持することは困難であった．そこで本稿では後者の方法を用い，SONE によるネットワーク内にメモリ機能を備えた素子を自己組織的に獲得させることとした．今回使用した素子は表 3 の真偽表に従ってその出力を決定するフリップ・フロップ素子（FF ノード）である．本稿では SONE によりネットワーク内の各素子が評価できることを利用し，FF ノードを自己組織的に獲得させた．FF ノードの生成は，ネットワークの各リンクにテスト用 FF ノードを備えることで，テスト用 FF ノードの評価値が閾値を上回った際に昇格・実用化することとした．また FF ノードの解体には，FF ノードの出力側リンクが 0 本となった際に解体することとした．

表 2: 非反転リンクの強化信号伝播規則

$Case1:$	$(R > 0) \wedge (Y_T = Y)$
	$R_T = 0$
$Case2:$	$(R > 0) \wedge (Y_T \neq Y)$
	Reconstructing TL2
$Case3:$	$(R \leq 0) \wedge (Y_T = Y)$
	$R_T = R_1$
$Case4:$	$(R \leq 0) \wedge (Y_T \neq Y)$
	$Y_T = -R_1$

表 3: FF ノード真偽表

入力 $1(I_1(t))$	入力 $2(I_2(t))$	出力 $(O(t))$
T	T	T
T	F	$O(t-1)$
F	T	$O(t-1)$
F	F	F

4. 実験

SONE の学習方式には強化学習の他に教師あり学習もあり，ネットワークへの強化信号の与え方によっていずれの学習も可能である [7]．本稿では FF ノードの効果を検証するため，教師あり学習による軌道学習を扱った．実験に用いた軌道は図 3(a) に示される軌道である．この軌道は，左側から中央に進入した場合には右側に，右側から進入した場合には左側に抜ける重複軌道（軌道中央部）を含んだ軌道である．この軌道を学習するためには，重複軌道の入り口でどちらの方向から進入したかを記憶し，重複軌道通過時点まで隠れ状態として記憶を保持し，重複軌道出口においてその記憶を活用する必要がある．

SONE には教師データより各点の座標を 16bit の入力として与え，SONE の出力が教師データと等しい出力ノードには正の強化信号（1）を，出力が異なる出力ノードには負の強化信号 (-1) をそれぞれ与えて強化信号の伝播とネットワーク構造の変更を行った後，次点の教師データの学習へ移行する．SONE による教師あり学習では，このプロセスを繰り返すことでオンラインに学習を行う．この実験を図 3(a) の軌道に対し各 10 回繰り返し行った．同様の実験を RNN でも行った．実験に用いた RNN は入力層，出力層のノード数を 2 とし，中間層 (5〜20)，コンテキスト層のノード数（1〜10），さらには学習率（0.001〜0.2）を様々に変えて実験を行った．ただし，RNN による実験では学習則としてバックプロパゲーション・スルータイム（BPTT）を利用し，80 万ステップを上限としてオフラインで学習を行った．

この実験の結果，FF ノード導入前の SONE には学習できなかった，図 3(a) に示されるような軌道が学習できるようになった．また，RNN による実験では学習率 0.2 以上では学習が安定せず，学習率 0.01 未満では 80 万ステップ学習時点で有効な結果は得られなかった．ノード数に関しては，中間層 10，コンテキスト層において良好な結果が得られた．表 4 に SONE による結果と，RNN による代表的な結果を示す．またこれらの学習に失敗したケースでは図 3(b) に示されるような局所解軌道への収束が多数確認されている．

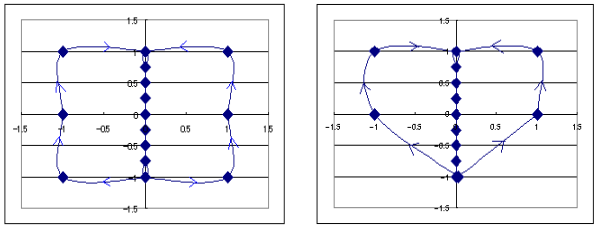
5. 考察

従来の SONE や RNN が収束した局所解軌道は重複軌道の出口の点と，その 1 ステップ先に到達すべき点が重なってしまっている．これは出口において記憶の活用が行われておらず，左右どちらの軌道へ移動するかの判断ができていないためである．その後左右の軌道への復帰が見られるが，これは教師データからの入力を得てからの復帰であり，学習による隠れ状

表 4: 学習結果

学習器	収束ステップ数	平均誤差	収束時平均中間ノード数	収束時平均 FF ノード数	隠れ状態の学習確率 [%]
SONE	8947.6	2.41×10^{-1}	208.8	-	0
SONE（FF）	2751.0	4.25×10^{-2}	190.2	38.6	90
RNN	−*	2.49×10^{-1}	10	-	0
RNN	−*	2.15×10^{-1}	10	-	0
RNN	−*	2.27×10^{-1}	10	-	0

*80 万ステップ学習時



(a) 正解軌道　　(b) 局所解軌道

図 3: 軌道

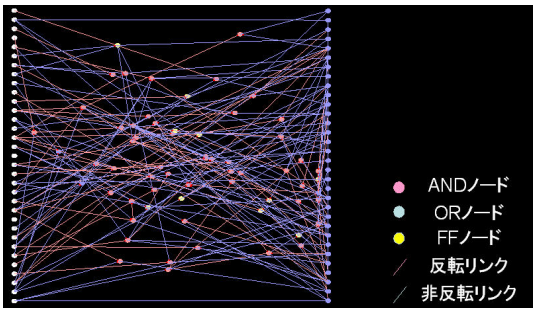


図 4: SONE によって獲得されたネットワーク

態の利用は行えていない．

従来の SONE や RNN の場合，記憶の保持はフィードバック結合によって行われるため，減衰や外乱の影響が生じ易く長期間の記憶保持は困難であると考えられる．一方で FF ノードを導入した場合，FF ノード自身がスイッチのように動作し，この動作には減衰を伴わないため，比較的に長期間の記憶保持が可能であると考えられる．

6. おわりに

本稿では，SONE による長期間の隠れ状態を含んだ時系列問題学習を実現するために，FF ノードを導入することを提案した．提案した自己組織化手法によって，FF ノードを含んだネットワークが自己組織的に獲得できるようになり，従来の SONE や RNN には困難であった，長期間の記憶保持が必要な隠れ状態を伴った軌道のオンライン学習が可能となった．しかしながら，本稿で行われた評価実験は教師あり学習によるものであるため，さらに応用範囲を広げるためには，強化学習においても同様なネットワーク自己組織化が可能であることを確かめる必要がある．今後は SONE による強化学習によって同様の効果を検証していきたい．

参考文献

[1] Katsunari SHIBATA, Yoichi OKABE, and Koji ITO :"Direct-Vision-Based Reinforcement Learning Using a Layered Neural Network-For the Whole Process from Sensors to Motors-", SICE Vol.37 No.2, 2001.

[2] Kenneth O. Stanley and Risto Miikkulainen: "Efficient Reinforcement Learning Through Evolving Neural Network Topologies", In Proceedings of the Genetic and Evolutionary Computation Conference, 2002.

[3] Masumi Ishikawa:"Structural Learning with Forgetting" , Neural Networks, Vol.9, No.3, pp.509-521, 1996.

[4] Jie Ni and Qing Song:"Dynamic pruning algorithm for multilayer perceptron based neural control systems", Neurocomputing, 2005.

[5] Gang Leng, Girijesh Prasad, and Thomas Martin McGinnity: "An on-line algorithm for creating self-organizing fuzzy neural networks", Neural Networks 17 1477-1493, 2004.

[6] Chyon Hae KIM, Tetsuya OGATA, and Shigeki SUGANO: "Self-Organizing Algorithm for Logic Circuit based on Local Rules", *Transaction* of the Society of Instrument and Control Engineerings (SICE) Vol. 42 No.4, 2006.

[7] Chyon Hae KIM, Tetsuya OGATA, and Shigeki SUGANO:"Enhancement of Self-Organizing Network Elements for Supervised Learning", In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), 2007.