神経システムの時系列学習則と学習の制御モデル

Temporal and Operant Learning Models by Neural System

重松　征史**
Yukifumi Shigematsu**

Abstract: Brain or central nerve system is a kind of information processing system which is differ from artificial computer one. A computer system has been made good progress by using the high-speed LSI technology and mass storage, however a nerve system has a slow-speed processing neuron and nonlinear element which can treats real-time information processing.　In this paper, we discuss about some dynamic or predictive functions in a nerve system , and propose models of a pulse neuron model, a temporal learning rule and a operant learning method of a neuron network. By using these rule and method, we proposed a temporal sequence associative memory network and an optimal action learning process.

Keyword:　Synaptic Plasticity,　Temporal Learning Rule,　Operant Learning,　Value Factor Controlling.

1. 緒　言

生物の脳神経系は、計算機とは異なる情報処理システムとして存在している。脳神経系が行っている情報処理は、単位要素として神経細胞であり、多数の神経細胞間を結ぶ複雑な神経線維の結合により情報が交換され判断された結果が出力される。神経線維を信号が伝達されて 1 段の神経細胞で判定処理する時間は約 10ms かかり、計算機の処理と比べると格段に遅い。それでも人間と計算機の能力の比較の一例をあげると、人間が人の顔や声を見聞きして人物を特定できるのは計算機よりも正確で早く 0.2 秒くらいである。この判定が処理速度の非常に遅い神経素子を約 20 段通過する時間内で計算機よりも早く判定が可能となるのが不思議であり、その仕組みを探求することから新しい情報処理の道が開けることも期待されている。そのために理化学研究所や産業技術総合研究所をはじめ多くの研究機関・大学で「脳を創る」プロジェクトの基で研究が進められている。

従来、工学的なニューラルネットワークの研究があったが、ここではもっと基礎的な脳神経生理学考えから取り組んで脳を理解するモデルを作ることが目的である。したがって複雑な脳神経系の持つ機能のうちで情報処理には欠かすことの出来ない原理を見つけて、脳神経系のモデルを構築することである。例えば、神経細胞が持つ機能をモデル化するとき、ニューラルネットワークで用いられるように細胞間の信号伝達をアナログ的な信号とするか、実際の神経細胞のようにパルス波形であるスパイク電位とするか決めるときスパイク電位が神経細胞の本質的な情報処理に欠かせない要素となっているならばパルス入出力の神経細胞モデルとしなければならない。そうしなければこの神経素子モデルを組み合わせて出来る神経回路網や神経システムが実際の脳神経システムが持つ高次の機能を実現することが出来ないからである。理想のモデル化とは、神経科学的な根拠を持って構築した神経情報処理モデルが、工学的にも有効であるとともに実際の脳生理学的な機能を実現してそれをフィードバックして生理学的な脳機能研究の推進にも役立てることである。

脳神経系の働きを検討すると、同時に起こる空間的な情報処理と、時間的な連続事象の関係を処理

** 愛媛大学大学院理工学研究科電子情報工学専攻情報工学コース
** Department of Computer Science, Graduate School of Science and Engineering, Ehime University, 3 Bunkyo-cho, Matsuyama, Ehime 790-8577, Japan.
email: sigematu@cs.ehime-u.ac.jp

平成 19 年 8 月 31 日 受付 , 平成 20 年 1 月 23 日 受理

することも同様に重要であることがわかる。例えば、経験的に原因と結果に気づき、次のことを予測しながら行動を取り対応するなどがある。脳神経系を理解しモデル化するうえで大切な原理・原則となる着目点を考えると、①外部世界から感覚器で捕らえた情報を実時間で認知・判定し行動を決定する機能、②経験を通して学習し記憶にとどめるための学習則、③取った行動の価値判断をして合目的な行動を習得するための学習の制御、④大脳皮質のコラム構造にあるような一つの機能を細胞の集団で作り上げるモジュール化、⑤モジュール化された機能を集めてさらに複雑な機能を構築し高次の管理機能を構築する階層構造などがあげられる。ここでは、神経細胞のダイナミック処理を考慮したパルス型神経素子モデル、神経接合部であるシナプスの可塑性から導かれる時系列学習則、賞罰のような価値判断により学習方向を制御し合目的な行動を取得する神経回路に関して今までの研究と考えをまとめて研究ノートとして紹介する。

2. パルス神経素子モデル

2．1　神経細胞の特性と機能

神経細胞（ニューロン）は、入力部である樹状突起と本体の細胞体、出力部の軸策（神経線維）とで出来ており、多数の神経細胞からのスパイク入力を樹状突起上の接合部（シナプス）の結合の強さにより入力の影響力を調整して受け取りそれらを集積して細胞体に伝える。細胞体では集積された活動電位を非線形演算してある閾値レベルを超えるとスパイク発生をし、出力の軸策を通してパルスの形で他の多数の神経細胞に信号を送る。神経細胞間の信号伝達はパルス波形のスパイク電位である。

これまでに述べてきたことから神経細胞素子モデルを作るとき、非常に複雑な神経細胞が持つ機能から出来るだけ簡略化したモデルとしなければならない。しかし、神経細胞の持つ本質的な情報処理にかかわる部分は省略することは出来ない。ここでは神経細胞の動的な機能を本質的な特徴と考えてモデル化した[1,2]。例えば、現時点の入力で発火しなくてもそれが内部ポテンシャルとして残留してその後に小さい入力があってもスパイク発火することや、複数入力スパイクの微妙なタイミングにより発火が異なることなどの機能を可能とした。また、入力が集積され細胞体で非線形演算がされパルス符号化されるとき、生理的に見ても符号化が妥当な方法で行われるモデルでなければならない。これらのことを考慮して神経素子モデルを作った。

2．2　神経素子のモデル

樹状突起は樹木の枝のように非常に複雑な形状をしておりその上に数千個のシナプスが葉のように分布している。それらを考慮してモデルを作ると非常に複雑となり計算機能力を超えてしまうので実用的でない。入力は思いきって簡略化してそれぞれの入力 X(t)とシナプスの結合加重 W との積を求め全ての入力積を加算したものが全体入力とする。神経素子 i の活性度 U は全ての入力の積和を求め前のステップの内部電位とから次のように求める。

$$U_i(t) = \sum_j W_{ij} X_i(t-1) + aV_i(t-1)$$

ここで、V は神経素子の内部電位、a はその減衰時定数であり、前の時間から減衰した内部電位が今のステップで活動度に加算され出力に影響することを示している。

活性度 U から非線形演算によって出力のスパイク発火が生じるかどうかは、ステップ関数 g(x)を用いて表現する。ステップ関数は数値 x が負のときは 0 でそれ以外は 1 となる。神経細胞は活性度がある閾値 θ を越えまで出力は無く閾値を越えるとスパイク発火をする。そして発火に伴ってしばらく応答を休む数 ms の不応期を持つので、発火に伴い活性度が低下するとみなした扱いをする。

$$X_i(t) = g(U_i(t) - \theta), \qquad V_i(t) = U_i(t) - pX_i(t)$$

ここで、内部電位は発火に伴い発火定数 p だけ減少し出力 X は[0,1]のスパイクとなる。このモデル

とＬＩＦ神経素子(Leaky Integrator and Fire neuron)との違いはＬＩＦ神経素子が発火に伴い内部電位 V が 0 にリセットされることだけであるが、動作としては違いがあり、この素子モデルは入力の少しの違いに対しても微妙な応答の違いが出てより実際の神経細胞の示す複雑な応答に近い特性を持つ。このことから計算の時間間隔をインパルスと不応期を含めた数 ms のように広く取って計算しても LIF 素子のような不都合は生じにくい。出力 X [1,0]は計算時間間隔内で細胞出力が有るか無いかを示す情報と解釈するとよい。

２．３　神経素子モデルの特性と検討

この神経素子モデルの活性度 U から出力スパイクのアナログ／パルス変換は、通信でΔ－Σ変調と呼ばれているパルス変調方式に近い特徴を持っている。Δ－Σ変調では、入力信号を積分(集積)して比較器で比較して零より大きければ正のパルスを出力し小さければ負のパルスを出力する。そして出力パルス分が入力から差し引かれる。上記の神経素子モデルはこの変調方式の正パルス発生部分の特性と同じ方式である。また、Δ－Σ変調のパルス信号は受信側でローパスフィルタを通せば素の信号に復調できる特性を持っており、神経素子モデルも同じように復調できる特性を備えている。生理学の知見でも眼の網膜出力である神経節細胞の実験で光刺激入力と細胞の活動電位からスパイク出力の関係を調べ比較的単純なパルス変換である結果が報告されている。この神経素子の各部分の波形を Fig.1 に示す。

神経素子モデルの動特性をみると、瞬時的な大きな入力に対して複数の連続スパイク発火となり、複数入力でそのパルス相互のタイミングにより出力が時間的に制御されることも確認できた。複数の神経素子間で同期発火や同期引き込み現象の動作もあり、神経細胞で見つかった確率共鳴現象である閾値以下の弱い入力に対しても適切な雑音レベルが加わり検出が可能となる機能も実現できた。このようにこの神経素子モデルは単純なモデルであるが神経細胞の持つ時間的処理の必要な基本動特性を実現していることが分かった。

このパルス型神経素子モデルを SAM 神経素子(Spike Accumulation and Modulation neuron model)と呼んでこれを用いてこれからの神経回路を構築してゆく。

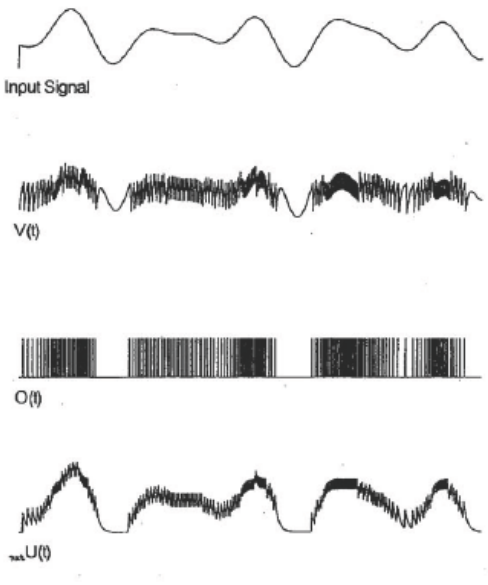

Fig.1 Waveform of input, V(t), output and accumulated pulse train

３．時系列学習則と時系列連想記憶モデル

３．１　脳神経系の時間情報処理と記憶

脳神経系では時間をどのように処理して記憶に留めているのであろうか。時間を 3 次元の空間座標に時間座標をもう一つ加えた 4 次元空間として取り扱う物理学的ニュートンの方法では、3 次元である脳神経系の中に時間的に変化する事象を記憶して留めることはできない。一方、時間に関して時間は出来事の前後関係に基づく事象の連鎖として存在するというライプニッツの考え方がある。実際日常生活の記憶は朝からの出来事の順に呼び出すことが容易であることからも納得できる。このことから脳神経系の記憶は時間的事象の連鎖として形成されると考える。

それでは神経系が時間を事象の連鎖として記憶するメカニズムは何かを求めることになる。ニューラルネットワークでは時間を扱う場合に神経回路に時間を扱う構造（例えばジョルダンやエルマーの出力・中間層から入力層へのフィードバックなど）を取り入れて時間の機能を作っている。しかし、

緒言でも述べたように脳神経系は時間情報を自由に処理できる機能を持っている。ベル音の後に食事が与えられることが続くとベル音で次の食事を予測することができる古典的条件付け学習のような学習は、アメフラシの単純な神経系でも可能である。したがって神経細胞自体が時間情報を処理できる能力を持っていると考えるのが自然である。

３．２ シナプスの可塑性

脳神経系では信号が神経細胞に興奮を引き起こし活動電位による動的な演算処理がされると同時に、神経間の結合も常に変化して新しい処理の機能を作っていることである。生理学者のヘブは 60 年も前に神経結合の変化を前段と後続の細胞が同時に発火すると結合が強められると仮定してヘブ学習則[3]を提案している。

近年、生理学実験でシナプスの強度が可塑性を示す変化をすることが、海馬における神経細胞のシナプスで見つけられ詳細な検証が行われた。それによると入力信号が強くまたは同期したタイミングによりシナプスの結合が変化して長期的に増強される長期増強（LTP）が起こり、入力が弱く位相のずれた入力では減弱する長期減弱（LTD）が起こることが知られている[4]。

また、細胞相互のスパイク発火のタイミングも詳しく調べられていて一つの神経細胞の発火時点と比べて後で発火する神経細胞への結合は時間差に応じて強化され、以前に発火した神経細胞への結合は時間差に応じて減衰すること(Fig.2)が確かめられSTDP(spike-timing dependent synaptic plasticity)と呼ばれている[5,6]。このように神経細胞は処理する信号の微妙なタイミングを用いて新しい処理回路を作り変えて時間情報処理をしている。これらのシナプスの可塑性の知見は、経験的な学習により記憶が神経系に形成される過程を示唆している。

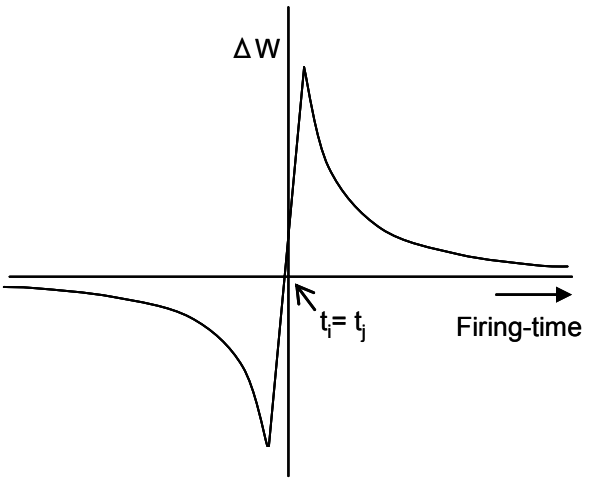

Fig.2 Weight modification of STDP

３．３ 時系列学習則

この生理的なシナプスの結合係数を変化させる LTP、LTD、STDP を含む可塑性の特性を考慮して時系列の学習則を考える。ヘブ学習則では同時発火の条件でであったが、時間関係を築くためにはシナプスに入力の影響がしばらく残留することが必要である。そこで活動の履歴値を導入した学習則を考えた。この履歴値と神経細胞の発火タイミングにより結合が変化すると仮定して改良ヘブ学習則を導いた[1,2]。また、STDP の結合の変化の前方への結合強化部分と後方への結合減衰部分の各成分は前者を興奮性シナプス、後者を抑制性シナプスによるものと仮定して学習則を作った。

興奮性結合の学習過程は、興奮性入力は前段からスパイク電位として伝わりシナプス内に入力履歴値 H^+が残留すると考えた。神経素子 i における j からの入力履歴値 $H_{ij}(t)$は、次のように決めた。

$$H^{+}{}_{ij}(t) = X_j(t) + q_1 H^{+}{}_{ij}(t)$$

ただし、q は履歴値の減衰時定数とする。この入力履歴が減衰しながらもしばらく保持される。

つぎに、神経細胞 i がある時に発火出力を出すと、各シナプスの入力履歴値と比較され学習される出力依存性の学習をする。結合 W の変化は次式のように入力履歴値に比例して増加すると同時に結合が入力履歴値に近づくように変化する学習則である。

$$\Delta W_{ij} = c_1 H^{+}{}_{ij}(t)\left\{ r_1 H^{+}{}_{ij}(t) - W_{ij} \right\} X_i(t)$$

この式で r は結合の大きさを決める定数で、 r ＝１－ q とすると W の最大値は１となる。W_{ij} の変化

はWがrHより小さいと増加してrHに接近して行き、rHより大きいと減少してrHに接近するHに関して2次関数の変化をし、最終的にはWがrHになるように学習を進めることになる。

一方、抑制性の学習過程は、興奮している神経細胞の活動を抑制するように働くので神経細胞がどのように興奮しているかを示す発火の履歴に関係すると考えて神経細胞 k の発火履歴 H^-を次のように決めた。

$$H^-_k(t) = X_k(t) + q_2 H^-_k(t)$$

ここで、q は履歴値の減衰定数で、もし履歴値の減衰を示す定数 q1 と q2 が等しいとすると、入力履歴H^+と神経細胞の発火履歴H^-は同じ細胞に関しては同じになる。

細胞 k への抑制性結合の学習則は、発火した細胞 i からの入力によって興奮を抑える入力依存性の学習とする。抑制性の結合強度 W がその細胞の履歴値－Hに比例し同時に－rHに近づくように負方向に変化する次式とする。

$$\Delta W_{ki} = -c_2(-H^-_k(t))\{-r_2 H^-_k(t) - W_{ki}\}X_i(t) = -c_2 H^-_k(t)\{r_2 H^-_k(t) + W_{ki}\}X_i(t)$$

この式からWは－rHより小さいと増加して－rHに接近して行き、－rHより大きいと減少して－rHに接近するので最終的にはWが－rHになるように学習を進めることになる。

興奮性結合と抑制性結合の学習則によってSTDPの特性が作られ、履歴値の2次関数による学習曲線の増加と減衰部分がLTDとLTPに関連づけたモデルとなっている。興奮性と抑制性の結合の大きさ比率はそれぞれの定数rによって決まる。これらの学習結果から興奮性・抑制性結合はそれぞれの履歴値に近づくように結合の学習が進み、未来(前)方向の結合は興奮性の結合が形成され、過去(後)方向の結合は抑制性の結合が形成されてゆく。このように時系列の入力による神経活動から時間流れの特性を持つ神経回路をつくる。

３．４　時系列学習と連想記憶回路

パルス駆動型のSAM神経素子モデルを用いて相互に結合可能な神経ネットワークを作り時系列入力信号を与えて上記の時系列学習則により時系列連想記憶ができるかを確かめた[1,2]。はじめに学習過程として時間的な順列入力でこの回路の神経素子をスパイク発火させながら時系列学習則で結合の学習を繰り返した。ある程度学習が進んだ時点で、最初の入力を与えると興奮性の結合によって次の神経素子のスパイク発火が起こりそれに続いて次々と順序に従った読み出しができた(Fig.3)。また順序入力が一部分共通になっている場合でA→B→C→D→EとM→N→C→O→Pとの間のCからの読み出しをみると、Aから始まるとC→Dと読み出し、Mから始まるとC→Oと読み出すことができた。これは前方の複数素子への結合強化があることで可能となった。これによって時系列の事柄を連想記憶する神経回路が入力刺激によって自動的にできることを示している。

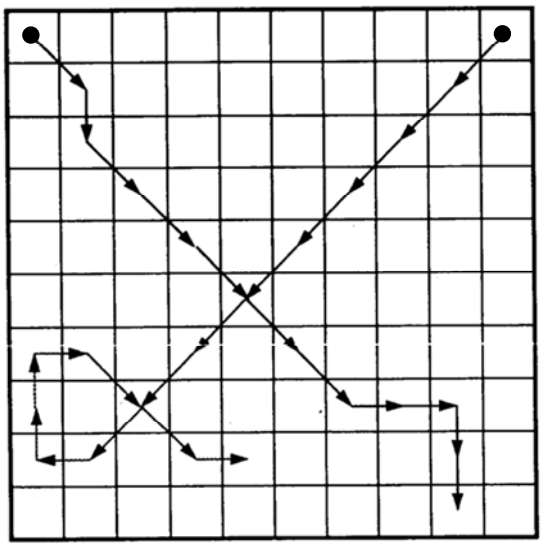

Fig.3 Recalling of a temporal associative memory

３．５　時系列処理の検討

ホップフィールドの相互結合形神経回路ではある素子から他の素子への結合とその逆の結合は同じ

結合強さでありすべての結合が対称結合である。しかしここで作られた素子間の結合は、時間的に前方には興奮性結合で後方には抑制性結合であり非対称結合となっている。非対称結合の神経回路はいくつかの特徴を持っている。非対称結合により、因果関係や三段論法の回路を作ることもできる。時間的な流れの中で記憶を読み出してつぎに何が起こるかを予測することができる。脳神経系では多くの場面で予測をしながら行動を選択していることからも大切な機能である。これらの時系列学習は時系列学習によって神経素子固有の結合の重さを変えることで時間処理を実現しているので、特別な構造を持った神経回路を必要としない。大脳皮質の神経回路のように自由に神経回路を配置して入力刺激によって自立的に時間情報処理の回路を作ることができる可能性を持っている。

4. 価値情報と学習の制御モデル

４．１　行動強化と学習

生物は日常環境の中で生活し経験するうちに、学習して自分に都合の良い合目的な行動がとれるようになる。その学習には、答えを教えられる教師あり学習もあるが、一般的には教師がいなくても学習はできる。例えば自由な行動を取っている間にある時、報酬や罰があると合理的な行動パターンを強化する。これはオペラント学習と呼ばれて動物の訓練でよく用いられている。すなわち、特別に教師が無くても報酬が与えられた時にはそれまでの行動が好ましいものとして強化され、罰の場合にはそれらの行動を回避して罰を避ける学習をして自然に合目的な行動が取れるようになる。この学習は環境や行動の諸条件を指定しなくても賞罰のみで行動を強化できることが優れた点である。上手に学習を進めることができれば未知の環境の中でもロボットが自律的に合目的な行動を取るようにできる。

この学習は、工学的には強化学習[7,8]と呼ばれるもので、出力の目標パターンは与えられず出力の良し悪しを示すスカラー値の評価信号だけが与えられる場合に、それを最大出力化するように学習することに相当すると定義することができる。その中でも評価信号が行動の後直ちに与えられず、一連の行動の後で与えられる場合には、これまでの行動の内でどの行動が最後の評価に結びつくものかを推定することが困難となり試行錯誤が増えてくる短所もある。

それを解決するアルゴリズムとして、最終的な評価信号を得るまでの各状態に対する評価関数を用いた TD 学習や Q 学習の方法が提案され制御やロボットの学習など[8,9]に活用されている。

生理学的にも、行動の制御として大脳基底核の細胞における報酬を予想させる入力信号とその選択的な反応が条件付け学習の進み具合により形成されることを示唆する実験結果も示されている。これをもとに大脳基底核においてドーパミンやセロトニンが関与した TD 学習を示唆する提案[10,11]もされているが、特別な構成の回路を仮定する必要があった。

一方、我々は入力履歴を考慮した時系列学習則を提案したが、これは細胞間の結合強度の学習則で局所的なことの処理でしかない。そこでこの時系列学習則に大局的な価値判断による評価信号を神経回路に与え方法により行動学習の制御をすることを考えた。そして行動の強化学習を可能とする神経ネットワークを構築する方法を提案する。

４．２　学習の制御と生理的知見

局所的な情報による時系列学習則では入力と出力の時間関係で学習をすることは出来ても、強化学習のように何段階かした行動の後で価値情報が与えられ、それまでの行動結果を好ましい方向に結びつける強化学習を進めることはできない。そこで価値情報または評価信号によって学習が進む方向を制御する方法を考える。

価値情報または評価信号は、好き嫌いの情動の活動によってもたらされるものであって、入力信号による神経の興奮活動とは異質のものである。評価信号の候補として、生理学的には情動活動によって神経伝達物質のドーパミンやセロトニンが放出されると言われている。これらの物質は、信号を伝達するグルタミン酸のようにシナプスの極限されたところに作用するのではなく、神経回路のもっと

広い部分に作用を及ぼすもので神経変調物質(neuro-modulator)とも呼ばれている。

Wickens らによる強化学習の細胞レベルでのモデル[12]によると、神経興奮によりシナプスの Ca^{+2} イオンが上昇し留まりそこにドーパミン細胞からドーパミンが放出されると結合重みを強化し、ドーパミンが無いと減弱するとした３要素が介入しているモデルを提案しているが時間関係は明確ではない。我々も神経変調物質ドーパミンが評価信号に関係し学習を目標の方向に制御する候補として考えてモデルを作った。

ここで空間を移動する場合を例にして、ある位置にいる状態とその時とる行動の関係を考えると、ある状態 Sa において行動 A1a を選択する事によって状態 Ｓｂ に移行し、次に行動 A2ｂ を選択すると状態 Sc に移行する。このように状態・行動の連鎖で表される。状態 Sa で別の行動 A3a を取ると状態 Ｓｃ になり遷移の異なる連鎖もでき、状態・行動の連鎖網ができる。この連鎖のある時点で状態 Sk になるとき評価信号（賞罰）が与えられると、今までの連鎖関係を時間順にたどって評価に従った強化か減衰かの行動の学習を制御することになる。

４．３　行動の強化学習の神経回路モデル

行動の強化学習のためには価値情報によって学習がどの様に制御されるかを考察する。時系列学習を強化学習にそのまま用いたときは出力細胞の発火時に学習が進むことになる。数ステップ行動した後で評価信号が与えられてもすでに学習が終わっていることになる。学習による結合係数の強化が行動の後にまで長期的に関わるためには、結合の強化が後で固定化され長期化する過程を加えた。そこで次のような仮説を考えてモデル化する。シナプスでの学習の効果は長期固定化せず学習結果が減衰はするけれどもその部分にしばらく残ると仮定する。また、価値評価により活性化するドーパミンなどの評価信号は、出力細胞層全体に分泌され、先に述べた学習効果を長期的に固定化する役割を果たしていると仮定した。以上の２つの仮定を基に行動の強化学習のモデルを作った。

行動の強化学習の方法を簡単な２層構造と価値評価細胞を有する Fig.4 の神経回路を用いて説明する。入力層は状態 S に関するＳ細胞層で、出力層は行動指令を出すＡ細胞層である。価値評価細胞はある状態Ｓｋ になったときに活性化され評価信号を出力のＡ細胞層全体に放出する。例えば、状態Ｓａ で行動Ａ1a を取ると入力が変化して状態Ｓｂ となる。このように状態と行動の時間連鎖が作られる。そしてＡ細胞の行動出力を出した細胞にはＳ細胞からの影響が減衰しながら残留する。その残留効果Ｒは次のように表される。

$$R_{ij}(t+1) = S_j(t) + \gamma R_{ij}(t)$$

ただし、Rij(t)は状態 j のとき行動 i を取った学習の残留効果であり、γは残留の減衰定数である。

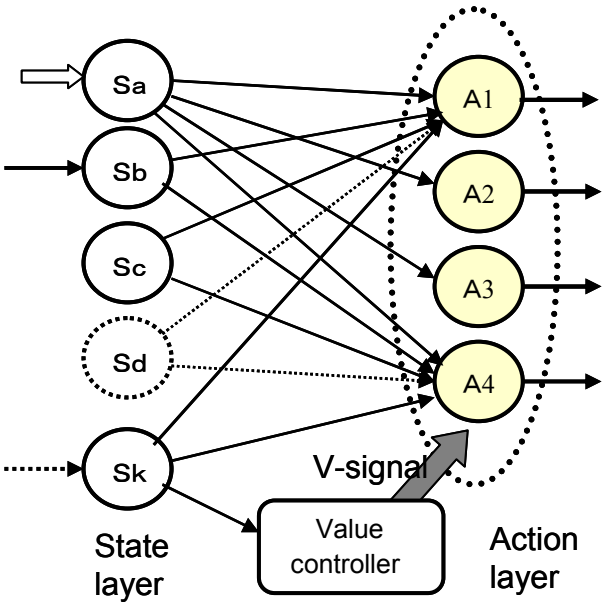

Fig.4 Neural network for Operant learning controlled by value signal

行動を次々と取った後で状態 Sk になると賞罰の価値評価が生じて価値評価細胞を活性化して評価信号が出力のＡ細胞層全体に放出され、これまでの残留効果Ｒによる状態から行動への結合の重みの固定化が次の式のように起こる。

$$\Delta W_{ij} = \alpha R_{ij}(t)$$

ただし、αは固定化定数で評価の賞罰で正数か負数を取るように考えることも出来る。
このことを、入力のS細胞から見ると、行動出力iとの結合重みが強化されると、状態Sjの時に行動Aiを取る確率が高くなるように行動の強化学習がされることを意味する。

４．４　モデルによる学習実験と結果

計算機により価値評価を加えた時系列学習則による強化学習を試した。ここでは一定のフィールドの中で目標点を探して到達する移動体を考えた。Fig.5のようにフィールド（10 x 10）を設け移動体は今いる位置の状態は認識でき、四方向(上下左右)に移動行動できるとする。神経回路との対応は、それぞれの位置を入力のS細胞とし、四方向の各行動を出力のA細胞と見立てて計算を行った。

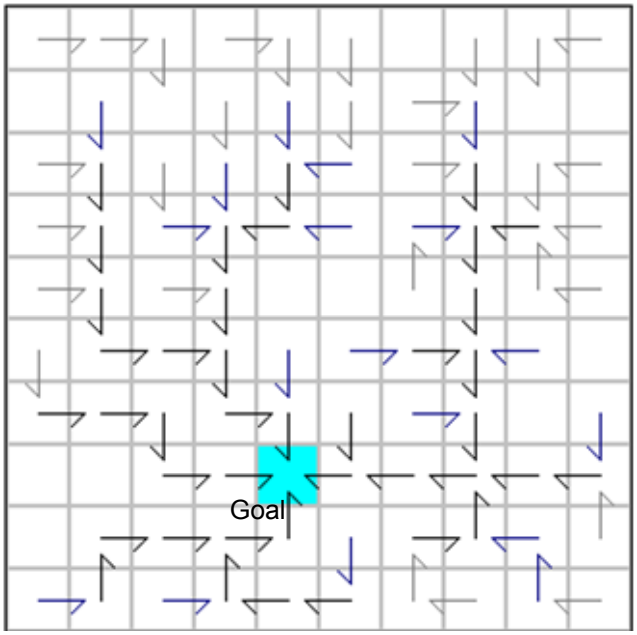

Fig.5 Enhanced direction after this learning process

始めは任意の出発点から状態ごとにランダムに行動を選択して移動して行くが目標点Skに到達すると、評価細胞が発火して評価信号がA細胞層内に与えられて、残留効果に比例してS細胞層からA細胞層への結合の重みが増強された。状態細胞から行動出力への結合重みは状態細胞Sjの出力の行動の確率和が一定となるように規格化した。

このような試行を繰り返す毎にしだいに状態から好ましい行動を取る方向への結合重みが形成されて行った。行動パターンもしだいに目標点に短いステップ数で達することが出来るようになった。Fig.5には、目標点にまで到達する各状態での最も選択確率の高い行動の方向を矢印で示した。また、Fig.6では450回学習に於いて試行毎の目標点に到達するまでのステップ数をグラフで示した。

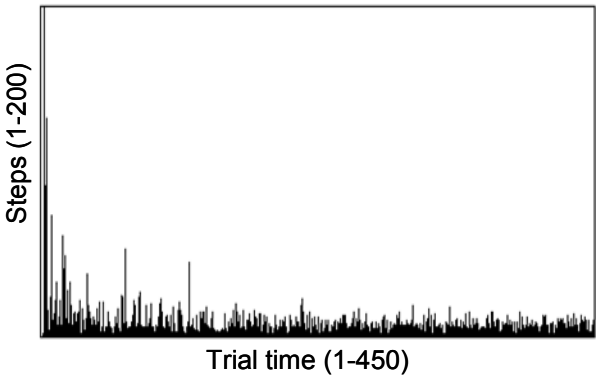

Fig.6 Trial time and necessary steps to a goal

学習固定化定数αと、残留減衰定数γによって、学習結果のパターンやその広がりが異なった。固定化定数αが大きいと学習が早く進むので試行回数が少なくすんだ。しかし、最初に目標点に到達して増強される効果が大きいので、次の試行でもその行動パターンの影響を受けた傾向になった。そのためにFig.5の矢印が必ずしも適正な方向を向いていないような行動パターンが所々に形成されやすかった。減衰定数γが小さいと目標点に到達した極近くの行動のみが強化されるために、目標から遠く離れた地点での行動はなかなか強化されなかった。そのため、全体のフィールドに行動選択の学習を広げようとすると、γもある程度大きい値を取る必要があった。

４．５　行動の強化学習の検討

価値情報による学習の制御を考慮して神経ネットワークのモデルで行動の強化学習を行う方法を提案し、その計算機による結果を示した。ここに示したのは最も単純な例として掲げたのでいくつかの問題点もあり改良すべき点もあるので以下で検討をする。

神経回路の構造を状態入力と行動出力の単純な二層構造としたが、実際には感覚入力がありそれから状態を識別して、行動を決める構造が必要である。感覚入力から状態を分類してゆく方法は、SOMによる自己組織化やARTによりクラス分けをすることが考えられる。ここでは状態が目標達成にどれだけ価値のある状態かの情報を使わない方法を取ったが、それを取り入れて価値評価細胞を一つ前の状態と今の状態の価値の差で出力を出すように機能を改めるなどをして、これまでのQ学習やTD学習の方法と比較検討をする必要もあるが、論理演算を導入しないできるだけ単純な回路構成を取ったモデルとした。また、価値情報には賞と罰があるように正反対の要素があるため、価値評価細胞部は複数の異なる評価をする部分を持ち、賞の評価信号と罰の評価信号で出力層細胞の結合を全く逆の方向に強化学習する様なモデルも機能の追加として当然考えられる。

さらに行動出力には興奮性のみでなく必ず抑制性も伴うので、出力細胞も興奮性と抑制性の拮抗した構造を取るようにした方が良いとも考えられる。この計算例では、上／下、左／右の各対はそれぞれ拮抗的な組み合わせであることは明らかであり、二つの対出力として回路構成をした方が良いと考えられる。

考察とまとめ

これまで脳神経系の情報処理システムを観察してそれが持っている情報処理機能を工学的に利用するために神経系の機能のモデル化を試みた。特に脳神経系が時間的な情報処理機能を持っていることに着目して、パルス形神経素子モデル、時系列学習則、行動の強化学習を実現するモデル述べた。

ここで用いた神経素子モデルは入出力がパルスで内部の演算はアナログの処理をする形となっている。これはアナログ入出力神経素子モデルと比べて伝達する情報量は小さい短所はあるが、タイミングを用いた演算処理など時間的に有効な機能を持ち実際の神経細胞の動作に近い形であることなど神経回路構築の基本素子モデルとして適切であると考えられる。またハードウェアで神経回路を作ることを考えると細胞間の信号伝達を 2 値信号にできるので細胞間の多数の結線を省略できて回路が作りやすい長所もある。

脳神経系が時間を事象の時間連鎖として認識しているという解釈は脳神経系の時間処理のために大切なことと考える。我々の日常生活で本質的には何時何分というような時刻を意識できないので、時計を使って時間を計り事象と時刻を結びつけて記憶に留めていると考えられる。時系列学習則により事象の連鎖が神経回路に記憶されると、事象の流れを読み出して次の事象を予測でき準備して対応できるので日常生活では不可欠な機能と考えられる。感覚器入力に近い部分では比較的実時間に近いタイミングで時系列の学習が進められると考えられるが、いくつかの階層を経て作られた高次の頭の中での記憶読み出しは、学習の時と同じ時間をかけて読み出すのではなく、興奮性結合により事象の順序だけを短時間で読み出すことができる。これは 1 日の出来事を短時間で振り返ることも可能であることを意味している。この階層化によって高次の履歴値の減衰定数は脳内でのリハーサルに必要な実際より小さい値でもかまわない。興奮性のシナプス可塑性は、ＬＴＰ、ＬＴＤの成立条件など生理的にも詳しく調べられているが、抑制性のシナプス可塑性は未知の部分が多くここでは入力依存性を仮定してモデルを作った。これは細胞の興奮が大きいと周囲からそれを抑制して回路全体が興奮しすぎないように抑制すると仮定している。

情動活動である好き嫌いの感情は扁桃体で認識されることが知られている。扁桃体は記憶の重要な役割をしている海馬の近くにあり海馬と神経連絡を持っているので学習の進む方向を制御していると考えてもおかしくはない。また、ドーパミン細胞系ともつながっており大脳新皮質にも影響を与えていることが考えられる。ここでは状態Ｓとそこでの行動Ａとの関係を学習するモデルであったが、状態Ｓが目標達成にどれくらい価値がある状態かを知って、行動の学習に役立てる方法も考えられる。このモデルでは価値評価細胞が目標達成の時にだけ活動すると仮定していたが、各状態から価値評価

細胞にも結線がありこの結合が入力履歴を持った時系列学習で強化される機能を追加すると状態と目標達成との関係ができて行動の学習機能を改善することも考えられる。

脳神経系の情報処理機能から基本原理を考慮して工学的に利用するためのモデルを紹介した。時間処理に関していくつかの処理方法を提案できたが、これらのモデルで用いる数々のパラメータをどのように決めるかという課題は残されている。例えば時間減衰定数は入力をどのような時間タイミングで処理するかによって適切に決める必要がある。また、興奮と抑制信号のバランスを考えて決めなければならないパラメータもある。神経系システムではこれらのパラメータが自律的にバランスよく調整できるどのような仕組みを持っているか知りたい点である。

謝　辞

これまでの成果は電子技術総合研究所と愛媛大学で研究して来たことを中心にまとめたものである。その間に多くの方々から教えていただき議論をしていただきました。特に故松本元博士はモデルに対して適切な評価と批評をしていただいたことを感謝します。また生理的な知見を教えていただいた電総研や理研の脳神経科学の研究者の方々にも感謝します。最後に、初めて網膜の情報処理に興味を持ち生理実験を自分で始めようとしたとき丁寧に指導し研究者としての研究態度を教えていただいた慶應義塾大学医学部の故冨田恒夫教授に感謝します。

参考文献

[1] 重松 征史“神経回路素子と学習・記憶“、pp.173-193.,日本物理学会編、”脳・心・コンピュータ”、丸善 k.k., 1996.

[2] Shigematsu, Y., Matsumoto, G. and Ichikawa, M. “A Temporal Learning Rule and Associative Memory” pp.164-172., in edited by Moreno-Diaz, R. and Mira-Mira, J. Brain Processes, Theory and Models, The MIT Press, Cambridge, Ma., 1996.

[3] Hebb, D.O. “The First Stage of Perception: Grouth of the Assembly”, The Organization of Behavior, (1949) pp.45-56., in edited by Anderson, J.A. and Rosenfeld, E., Neurocomputing: Foundation of Research, The MIT Press, Cambridge, Ma., 1989.

[4] Bliss, T.V. and Collingridge, G.L. “A Synaptic Model of Memory: Long-term Potentiation in the Hippocampus” pp.31-39, Nature, 361, 1993.

[5] Ganguly K., Kiss L. and Poo L. “Enhancement of Presynaptic Neuronal Excitability by Correlated Presynaptic and Postsynaptic Spiking” pp.1018-1026., Nature Neuroscience, 3, 2000.

[6] Froemke, R.C. and Dan, Y. “Spike-timing-dependent Synaptic Modification induced by Natural Spike Train” pp.433-438. Nature, 416,　2002.

[7] Barto, A.G. "Adaptive Critics and the Basal Ganglia", pp.215--232., in edited by Houk, J.C., Davis, J.L. and Beiser, D.G. Models of Information Processing in the Basal Ganglia, The MIT Press, Cambridge, Ma., 1995.

[8] 山村雅幸、宮崎和光、小林重信“エージェントの学習“、人工知能学会誌、9、pp.683-689. 1995.

[9] Barto, A.G., Sutton, R.S. and Anderson, C.W. (1983) "Neuronlike Adaptive Elements That Can Solve Difficult Learning Control Problems" pp.834-846., IEEE, SMC-13, 5, 1983.

[10] Schltz,W., Dayan, P. and Montague, P.R. ”A Neural Substrate of Prediction and Reward”, pp.1593-1599. Science, 275, 1997.

[11] Montague, P.R., Dayan, P. and Sejnowski, T.J. “A Framework for Mesencephalic Dopamine Systems Based on Predictive Hebbian Learning”, pp.1936-1947, J. Neuroscience, 16(5), 1996.

[12] Wickens, J. and Koetter, R. "Cellular Models of Reinforcement", pp.188-214., in edited by Houk, J.C., Davis, J.L. and Beiser, D.G. Models of Information Processing in the Basal Ganglia, The MIT Press,