

軌道アトラクタを用いたパターンベース推論

山根 健^{†a)} 蓮尾 高志[†] 末光 厚夫^{††} 森田 昌彦^{†b)}

Pattern-Based Reasoning Using Trajectory Attractors

Ken YAMANE^{†a)}, Takashi HASUO[†], Atsuo SUEMITSU^{††}, and Masahiko MORITA^{†b)}

あらまし シンボルグラウンディング問題やフレーム問題に起因する古典的人工知能の限界を超えるには、もともとパターンで表現される外界の情報をパターンそのまま処理するパターンベースの推論が有効だと考えられるが、シンボルやそれに類するものを全く用いる必要のない推論エンジンはこれまでなかった。本論文では、非単調神経回路網が構成する大自由度力学系のダイナミックスを利用して、完全なパターンベースの推論を行うモデルを提案する。このモデルでは、情報はすべてパターンとして分散的に表現され、適切な推論結果を表すパターンへの状態遷移が生じるよう、力学系のある部分空間に軌道アトラクタを形成することが知識の学習に相当する。簡単な推論システムを構築したところ、全く未知の問いに対しても類推によって適切に答え、非単調推論も自然な形で実現できるなど、従来の推論方式にはない特徴が示された。まだ研究の初歩的段階ではあるが、本モデルは推論方式や性質が脳に似ており、大きな可能性をもつと考えられる。

キーワード 分散表現, 非単調ニューラルネット, 選択的不感化, 神経力学系, 常識推論

1. ま え が き

シンボル操作に基づく古典的人工知能には、シンボルグラウンディング(記号接地)問題[1]及びフレーム問題[2]という未解決の難問があることが古くから指摘されている。そのため、ごく限られた状況でしか適切に動作せず、現実の複雑な世界ではしばしば役に立たないとされる。例えば、人や車が行き交う街で適切に盲人を誘導するシステムを実現することは、現在の技術では不可能である。

しかしこれらの問題は、情報をシンボルで表現し、それを操作することによって推論を行うというシンボルベース推論の枠組みの中では避けることができない上に、将来的にも解決の見込みは全く立っていない。また、シンボルではなくファジー集合によって情報を表現し、それをファジー論理に従って操作するファジー

推論も、これらと本質的に同じ問題を抱える。

古典的人工知能は人間の論理的思考をモデル化したものとされるが、人間の知能は明らかに上記の限界を超えている。これは、人間は常に論理的に思考しているわけではなく、推論の際に必ずしも直接シンボルを操作しているわけではないことの反映である。また、一部の類人猿を除く動物は、言語をもたないにもかかわらず、やはり現実世界で役に立つ知能をもっている。例えば盲導犬は、シンボル操作を行っていないが、未経験の状況であっても適切に盲人を誘導するだけの思考能力をもつ。

こうしたシンボル操作によらない思考のメカニズムは未解明であるが、脳の基本的な計算様式と関係していることは間違いないであろう。すなわち、脳では情報はシンボルとして集約的に表現されるのではなく、多数のニューロンの興奮のパターンとして分散的に表現され、推論の過程はそれらのニューロンの並列の相互作用のダイナミックスとして展開される[3]。このようなシンボルを全く用いない推論方式のことを、ここでは「パターンベース推論」と呼ぶ。

パターンベース推論では、もともとパターンとして表現されている現実世界の情報を、シンボル化することなく処理するのであるから、少なくともシンボル

[†] 筑波大学システム情報工学研究所, つくば市
Graduate School of Systems and Information Engineering,
University of Tsukuba, Tsukuba-shi, 305-8573 Japan

^{††} 島根大学総合理工学部, 松江市
Interdisciplinary Faculty of Science and Engineering,
Shimane University, Matsue-shi, 690-8504 Japan

a) E-mail: yamane@bcl.esys.tsukuba.ac.jp

b) E-mail: mor@bcl.esys.tsukuba.ac.jp

グラウンディングは回避できる。また、パターンの類似性などを利用した類推によって、フレーム問題についてもある程度対処できる可能性がある。つまり、パターンベース推論には古典的人工知能の限界を超える可能性があり、逆に現時点でその可能性があるのはパターンベース推論だけであろう。本研究の最終的な目標は、このような脳を模範とする推論方式を確立し、それによって人間や高等動物のような優れた知能を実現することである。

しかしながら、真にパターンベースといえる推論方式、つまりシンボルやそれに類するものを全く用いず、かつ現実的（規模が大きくなっても計算量が爆発しないなど）な推論エンジンは、これまで知られていない。特に、代表的なパターンベースの情報処理系である従来型のニューラルネット（いわゆるコネクショニストモデル）の場合、そのような推論の実現は、次のような理由で原理的に困難であった。

まず、推論を行うのであれば少なくとも有限オートマトンと同等の能力は必要だと考えられるが、リカレント型ニューラルネットは、理論上は任意の有限オートマトンを模擬する能力があるとされてきた。ところが実際には、オートマトンの各状態をニューロンの活動パターンとして分散表現すると、状態数が増えたときに1対多対応による荷重の平均化という問題が生じるため、原理的に局所表現を介在させる必要がある[4], [5]。実際、従来のニューラルネットを用いた推論システムは、ごく小規模なものを除き、すべてシンボルベースのシステムと組み合わせる（内部状態に応じて異なる回路網を切り換える場合も含む）か、部分的に局所表現を用いている。しかし、前者ではパターンをシンボルに対応づけることが必要であるし、後者の局所表現はシンボルまたはファジー集合による表現とほぼ等価であるから、いずれにせよシンボルグラウンディング問題は避けられない。

これに対して、筆者らのグループは最近、選択的不感化という方法によって上記の平均化の問題を解消できること、それを非単調神経回路網[6]に適用したモデルが、局所表現を用いることなく任意の大規模有限オートマトンを模擬する能力をもつことを示した[4]。このモデルは、理屈上は十分なサイズの記憶装置（チューリングマシンのテープに相当するもの）と組み合わせることによって、古典的人工知能の推論エンジンを模擬することができる。しかし、そのような方法は現実的でないし、パターンベースである意味も

ほとんどない。

そこで、本論文では、このモデルを用いて、分散表現の利点を生かした推論、すなわちパターン間の類似性に基づく類推や、パターン構造に埋め込んだ暗黙の知識を利用した推論を行うための基本的な方法を提案する。以下では、まずモデルの動作原理について説明し、その汎化能力について議論する。次いで、実際に簡単な推論システムを構築することによって推論の方法を示すとともに、パターンベース推論の可能性を検証する。

2. 軌道アトラクタを用いた連想モデル

本章では、推論に用いるモデルについて簡単に述べる。詳細については文献[4]を参照されたい。

2.1 非単調神経回路網

非単調神経回路網[6]は、ある種の非単調な出力特性をもつ神経素子が相互に結合した回路網である。ここでは、非単調出力関数として

$$f(u) = \frac{1 - e^{-cu}}{1 + e^{-cu}} \cdot \frac{1 - e^{c'(|u|-h)}}{1 + e^{c'(|u|-h)}} \quad (1)$$

を用い（ c, c', h は正の定数）、連続時間のダイナミクスに従って動作するものとする。すなわち、 i 番目の素子の出力 y_i は次式で与えられる。

$$\tau \frac{du_i}{dt} = -u_i + \sum_{j=1}^n w_{ij} y_j + z_i \quad (2)$$

$$y_i = f(u_i) \quad (3)$$

ここで、 n は素子の数、 u_i は素子の内部電位、 w_{ij} は j 番目の素子からの結合荷重、 z_i は回路外部からの入力信号、 τ は時定数を表す。

このモデルでは内部電位 u_i の符号が重要なので、 $x_i = \text{sgn}(u_i)$ を考え（ $\text{sgn}(u)$ は $u > 0$ のとき $1, u \leq 0$ で -1 をとる符号関数）、ベクトル $\mathbf{x} = (x_1, \dots, x_n)$ のことを回路網の状態と呼ぶ。ある時刻における回路網の状態 \mathbf{x} は、 2^n 個の可能な状態からなる状態空間中の1点で表される。また、各 x_i は非同期的に変化するから、 \mathbf{x} が別の状態に遷移する際には、ほとんどの場合隣の点（成分が一つだけ違う状態）に移動する。したがって、状態空間中には、時間とともに \mathbf{x} の連続的な軌跡が描かれることになるが、これを \mathbf{x} の軌道と呼ぶ。

非単調神経回路網は、一般的なニューラルネットにない多くの優れた性質をもつが、中でも重要なのは、

状態空間中のある軌道に沿って安定に状態遷移することが可能ということである．このような軌道のことを軌道アトラクタと呼び，これを用いて様々な時空間パターン処理が実現できる．

2.2 学習と連想

軌道アトラクタを用いて，ある空間パターン群 S^1, \dots, S^m から対応する別の空間パターン群 T^1, \dots, T^m を連想することを考えよう．ここで S^μ と $T^\mu (\mu = 1, \dots, m)$ はいずれも ± 1 を成分とする n 次元ベクトルであり，前者を cue パターン，後者を target パターンと呼ぶことにする．そのためには， S^μ を始点とし， T^μ を終点とする m 本の軌道アトラクタを形成すればよいが，その方法は次のとおりである．

まず， S^μ から T^μ へ至るある連続的な軌道に沿って変化する時空間パターンを作成し，これを学習信号 r として回路網の学習を行う．具体的には， $x = S^\mu$ を初期状態として式 (1)~(3) に従って回路網を動作させた上で， r を $z_i = \lambda r_i(t)$ の形 (r_i は r の成分， λ は入力強度) で各素子に入力する．これと並行して，すべての結合荷重 w_{ij} を

$$\tau' \frac{dw_{ij}}{dt} = -w_{ij} + \alpha r_i y_j \quad (4)$$

に従って更新する．ここで， τ' は学習の時定数であり， $\tau' \gg \tau$ である． α は学習係数であって定数でもよいが， $|u_i|$ の増大とともに α を減少させると学習性能が高まるため， $\alpha = \alpha' x_i y_i$ とした方がよい (α' は正定数)．

直観的にいうならば，式 (4) の学習則は r 周辺における回路網のエネルギーを低くする．したがって， r が S^μ から T^μ まで連続的に変化することによって，エネルギーの溝が刻まれることになる．更に， r が x よりも少し先行することによって， x から r の方向，すなわち S^μ から T^μ へ向かう緩やかな流れが溝の底に作られる．すべての μ についてこれを行うが， r の入力強度 λ を徐々に減らしながら学習を何度か繰り返すと，回路網の初期状態として cue パターン S^μ やそれに近いパターンを与えるだけで， x は r の軌道にほぼ沿って target パターン T^μ に達するようになる．

2.3 選択的不感化による文脈修飾

軌道アトラクタを用いると，cue パターン間の直交性が低い (パターン同士が類似している) 場合でも，正しく連想することが可能である．しかし，上記の方法では，同一の cue から異なる target への連想はできない．cue パターンとは別の情報 (ここでは「文脈」

と呼ぶ) に応じて別々の target パターンを想起するためには，文脈に応じて軌道アトラクタの始点が変わるよう，回路網に修飾を加える必要がある．

その手段として最も簡単なのは，cue パターン S^μ と文脈パターン C^ν を連結したベクトル (S^μ, C^ν) を軌道アトラクタの始点とすることである．しかしながら，このような方法では，前述した平均化の問題が解消されないため，素子数をいくら増やしても扱えるパターン数に限りがある．そこで，本モデルでは選択的不感化法を用いる．これは，文脈ごとに異なる素子群を不感化する (入力に関係なく中立の値を出力させる) 手法である．

具体的には，回路網の i 番目の素子の出力が，式 (3) ではなく

$$y_i = g_i f(u_i) \quad (5)$$

によって与えられるものとする．ここで， g_i はこの素子の感度を表し，通常は 1 であるが不感化されるときには 0 となる．一般に g_i は文脈パターン C に依存して決まるが，ここでは最も単純に $g_i = (1 + c_i)/2$ とする．ただし， C は n 次元のパターンであって，成分 c_i は ± 1 を確率 $1/2$ でとるものとする．

このような修飾を加えると，回路網の状態が n 次元の状態空間から $n - k$ 次元 (k は不感化される細胞数で $k \simeq n/2$) のある部分空間に射影される．同時に，その部分空間における (不感化されない素子によって決まる) ダイナミクスに従って回路網が動作することになる．

図 1 に模式的に示すように，これを利用して，同一

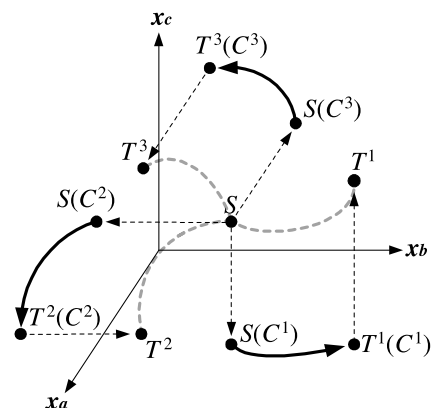


図 1 選択的不感化による文脈依存の状態遷移の模式図
Fig.1 Schematic representation of the process of context-dependent association.

の cue パターン S から文脈ごとに異なる target を連想することができる．すなわち、文脈パターン C^ν によって射影される部分空間における軌道アトラクタによって、 $S(C^\nu)$ から $T(C^\nu)$ へ状態遷移する．ここで、 $S(C^\nu)$ 及び $T(C^\nu)$ は C^ν による修飾を受けた S 及び T を表す 3 値 (1, 0, -1) パターンである．target パターン T^ν が点アトラクタになっていれば、この状態で修飾を外す (すべての素子を不感化から解放する) ことによって T^ν が想起される．

3. モデルの汎化能力

パターンベースの情報処理の最も大きな利点の一つに、汎化が生じること、すなわち未学習のパターンであっても処理できることが挙げられる．次章で述べる推論システムの類推機能も、軌道アトラクタと選択的不感化を用いた連想モデルの優れた汎化能力に基づいている．しかし、文献 [4] ではそれが十分に示されていないかった．そこで、本章では、このモデルの汎化能力を数値実験により検証する．

3.1 実験方法

素子数 $n = 2000$ のモデルを計算機上で構成し、以下の方法で実験した．

まず、cue パターン S^1, \dots, S^{10} 、文脈パターン cue パターン C^1, \dots, C^{10} 、target パターン T^1, \dots, T^{19} をすべてランダムに作成する．これらについて、表 1 のような連合をモデルに学習させる．この表は、パターン C^ν で表される文脈 (以下、文脈 C^ν と表記) において、cue パターン S^μ が与えられたとき、想起すべき target パターンを示したものである．

ここで、cue パターン $S^\mu (\mu \geq 2)$ は μ 個の異なる文脈において同一の target パターンと対応し、また文脈 $C^\nu (\nu \geq 3)$ では、異なる $\nu - 1$ 個の cue パターンが同一の target パターンと対応するようにした．これは、近接した複数の軌道アトラクタが汎化能力に与える影響を見るためである．学習回数は 10 回、モデルのパラメータは $c = 50$ 、 $c' = 10$ 、 $h = 0.5$ 、 $\tau' = 5000\tau$ 、 $\alpha' = 2$ である．

次に、学習後のモデルに様々なノイズを加えた cue パターンを与え、各文脈において正しい target パターンが想起できるノイズの最大量 (臨界ノイズ量) を求めた．同様に、文脈パターンにノイズを加えた (cue パターンにはノイズを加えない) 場合の臨界ノイズ量も調べた．

なお、以下においてパターン間の「類似度」をバ

表 1 文脈ごとの cue パターンと target パターンの関係
Table 1 Relation between cue and target patterns in each context.

	C^1	C^2	C^3	C^4	C^5	C^6	C^7	C^8	C^9	C^{10}
S^1	T^1	T^{11}	T^{12}	T^{13}	T^{14}	T^{15}	T^{16}	T^{17}	T^{18}	T^{19}
S^2	T^2	T^2	T^{12}	T^{13}	T^{14}	T^{15}	T^{16}	T^{17}	T^{18}	T^{19}
S^3	T^3	T^3	T^3	T^{13}	T^{14}	T^{15}	T^{16}	T^{17}	T^{18}	T^{19}
S^4	T^4	T^4	T^4	T^4	T^{14}	T^{15}	T^{16}	T^{17}	T^{18}	T^{19}
S^5	T^5	T^5	T^5	T^5	T^5	T^{15}	T^{16}	T^{17}	T^{18}	T^{19}
S^6	T^6	T^6	T^6	T^6	T^6	T^6	T^{16}	T^{17}	T^{18}	T^{19}
S^7	T^7	T^7	T^7	T^7	T^7	T^7	T^7	T^{17}	T^{18}	T^{19}
S^8	T^8	T^8	T^8	T^8	T^8	T^8	T^8	T^8	T^{18}	T^{19}
S^9	T^9	T^9	T^9	T^9	T^9	T^9	T^9	T^9	T^9	T^{19}
S^{10}	T^{10}	T^{10}	T^{10}	T^{10}	T^{10}	T^{10}	T^{10}	T^{10}	T^{10}	T^{10}

ターンベクトル間の方向余弦によって定義し、ノイズ量は類似度の変化量と定義した．したがって、もとのパターンとの類似度が 0 となる、すなわちもとの成分の半分が反転するようなノイズ量が 1 となる．また、臨界ノイズ量は想起の成功率が 50% (10 回中 5 回) 以上となる値としたが、10 回中 10 回成功する値との差は 0.1 以下であった．

3.2 結果

まず、図 2 に cue パターンに関する臨界ノイズ量を示す．グラフの横軸は cue パターンであり、4 本の線は実験に用いた四つの文脈を表す．

この図から、文脈 C^1, C^4, C^7 では、どの cue パターンかに関係なく、したがって同じ target パターンに至る軌道アトラクタの数に関係なく、臨界ノイズ量はほぼ一定であることが分かる．しかも、その値は 0.8 以上であるから、 S^μ との類似度が 0.2 しかない cue パターンを与えても、 S^μ を与えたときと同じ target パターンを想起できることになる．

ただし、文脈 C^{10} では cue パターンごとのばらつきがやや大きく、特に S^9 に関する臨界ノイズ量は約 0.6 とほかよりも小さい．一方、この文脈で唯一 T^{10} を target パターンとする cue パターン S^{10} に関する臨界ノイズ量は 0.8 であり、他とあまり変わらない．したがって、 S^9 に関する臨界ノイズ量が小さいのは、文脈 C^1 から C^9 においてすべて T^9 を想起するよう学習したためと思われる．また、この実験では、学習に用いた cue パターン間の類似度はすべてほぼ 0 であるが、cue パターン同士が類似していると、臨界ノイズ量のばらつきは大きくなる (4.3 でそのような例を扱う)．しかし、全体としてみると、cue パターンに関する汎化能力は一樣に高いといえる．

これに対して、図 3 は文脈パターンに関する臨界ノ

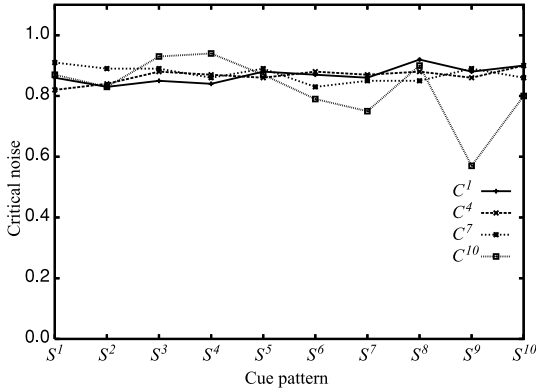


図 2 cue パターンに関する汎化能力

Fig.2 Generalization ability for cue patterns.

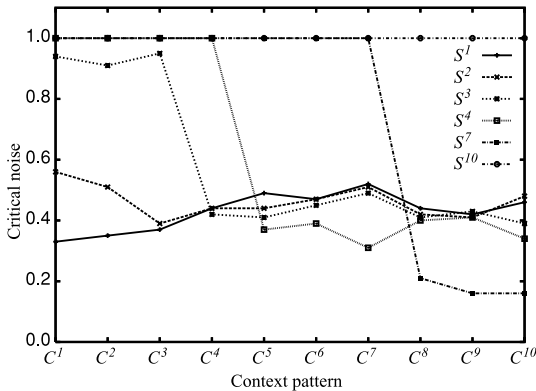


図 3 文脈パターンに関する汎化能力

Fig.3 Generalization ability for context patterns.

イズ量を示したものである．図 2 に比べて S^μ や C^ν による違いが大きく，1 またはそれに近い値と，0.2～0.5 程度の低い値に分かれていることが分かる．このとき，大きな値をとるのは，すべて target パターンが $T^3 \sim T^{10}$ のいずれか，すなわち三つ以上の文脈で同じ連合を学習した場合である．これを詳細に見てみよう．

まず， S^{10} が cue パターンのとき，臨界ノイズ量はすべて 1 である．これは，どんな文脈パターンを与えても必ず T^{10} を想起するからである．また， S^7 が cue パターンのとき，文脈 $C^1 \sim C^7$ において臨界ノイズ量が 1，文脈 $C^8 \sim C^{10}$ では臨界ノイズ量が 0.2 以下となっている．このことは，文脈パターンが $C^8 \sim C^{10}$ にかなり近い場合を除く，ほぼすべての場合において T^7 を想起することを意味している． S^4 が cue パターンの場合も同様であるが， T^4 以外を想起する文脈パ

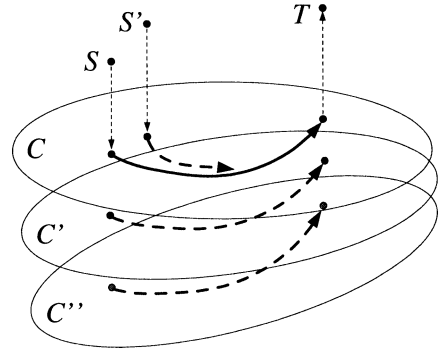


図 4 軌道アトラクタによって生じる 2 種類の汎化

Fig.4 Two kinds of generalization caused by trajectory attractors.

ターンの範囲がやや広くなり， S^3 が cue パターンの場合には，全く新しい文脈で T^3 を想起するのは困難になる．このように，同じ連合を学習する文脈の数が多いときに限り，文脈パターンに関して極めて強力な汎化が生じる．

cue パターンと文脈パターンでこのような違いが生じる理由を模式的に示したのが図 4 である．図中の平面は，ある文脈に対応する部分空間，すなわちその文脈における力学系の状態空間を表す．

さて，cue パターンに関する汎化は，学習時の cue パターン S と類似した別の cue パターン S' が与えられたとき，文脈 C において $S(C)$ から $T(C)$ へ至る軌道アトラクタに引き込まれることによって生じる．これに対して，文脈パターンに関する汎化が生じるのは， C とは別の文脈 C' に対応する部分空間にも $S(C')$ から $T(C')$ へ至る軌道ができるためだが，そのでき方には 2 通りある．

一つは， C' が C に十分近い場合である．このとき，二つの文脈では不感化されない素子の多くが共通するから，それらの力学系はよく似た流れをもつ．その結果，基本的に文脈 C で形成されたすべての軌道アトラクタに対応する軌道アトラクタが文脈 C' にも存在することになる．

これに対して，文脈 C から離れた文脈 C'' では， C において軌道アトラクタを形成しても対応する軌道アトラクタはできない．しかし，たとえ文脈パターン間の類似度が 0 であっても，不感化されない素子の半分は共通であるため， $S(C)$ から $T(C)$ への軌道アトラクタの影響はかなり受ける．そのため，数個の異なる文脈 C^ν ($\nu = 1, 2, \dots$) で $S(C^\nu)$ から $T(C^\nu)$ への軌道

アトラクタを形成したとき、それらの総合的な影響によって、文脈 C'' にも軌道アトラクタができると考えられる。

以上から、このモデルでは cue パターンだけでなく、文脈パターンについても強い汎化が生じることが示された。

4. 推論システムの構成

4.1 推論の方法

上述のモデルにおいて、文脈 C^1 で cue パターンとして S^1 を与えたとき、状態が S^2 を経由して T^1 へ遷移したとしよう。このとき、例えば S^1 が「スズメ」を、 S^2 が「鳥類」を、 T^1 が「飛ぶ」を表すものとするならば、これはモデルが「スズメは鳥類であるから飛ぶ」と推論したとみなすことができる。この場合、 C^1 は『飛ぶか』について問っている文脈を表すことになり、モデルに cue パターン S^1 及び文脈パターン C^1 を与えることは「スズメは飛ぶか」と問っていると解釈できる。ただし、このような言葉による表現は分かりやすく説明するためであって、 S^1 や C^1 自体は情報を分散表現したパターンであることに注意されたい。

このような考えに従ってモデルに推論を行わせるためにはあらかじめ知識を与えておく必要があるが、それに相当するのは文脈パターン C^1 (飛ぶか) による修飾を加えた状態で S^1 (スズメ) から S^2 (鳥類) 及び S^2 から T^1 (飛ぶ) への軌道アトラクタを形成することである。同様に、文脈 C^1 において S^3 (ウマ) から S^4 (哺乳類) 及び S^4 から T^2 (飛ばない) への軌道アトラクタを形成すれば、「ウマは哺乳類なので飛ばない」という知識を学習したことになるし、別の文脈 C^2 (動くか) における S^1 から S^5 (動物) 及び S^5 から T^3 (動く) への軌道アトラクタは「スズメは動物なので動く」という知識に対応する。

さて、このようにして必要な知識を獲得したシステムは、学習した軌道アトラクタに沿って状態遷移をすることによって、演繹的推論ができるというだけではない。知識を適切に表現すれば、3. で述べた汎化能力によって、未学習の問いに対してももっともらしい結論を導出する、いわゆる常識推論の能力をもつと考えられる。

例えば、文脈 C^1 (飛ぶか) において、 S^1 (スズメ) に似たパターンである $S^{1'}$ (ツバメ) が cue パターンとして与えられると、モデルに形成された $S^1 \rightarrow S^2$

(鳥類) $\rightarrow T^1$ (飛ぶ) という軌道アトラクタに引き込まれる可能性がかなり高い。つまり、「スズメは鳥類であるから飛ぶ」という知識を利用して推論が進み、「ツバメは飛ぶ」と結論を導くことが期待される。同様に、『翼をもつか』という文脈が C^1 (飛ぶか) に似たパターン $C^{1'}$ で表されるならば、「スズメは翼をもつ」と類推することが期待される。

4.2 知識の表現方法

実際に推論システムを構成するには、知識を何らかの方法でパターンとして表現してモデルに与えなければならない。この方法がシステムの推論能力を大きく左右するが、具体的にどうするのが最適であるかは不明である。

一方、本システムが模範とする脳の場合、外界から得られるすべての情報は、パターンとして感覚器から取り込まれる。その後、特徴抽出など様々な処理を経て前頭葉などの連合野に送られるが、ここが推論エンジンに相当すると考えられる。また、推論の結果もパターンとして表現され、それがそのまま次の推論に用いられ、運動系などに送られて行動を引き起こしたりすると考えられる。このように、脳では推論がパターンベースで行われているだけでなく、その入出力を担う部分もパターンベース処理がなされており、これらが一体となって知能が実現されている。

しかしながら、脳の情報処理メカニズムには不明な点が多く、例えば前頭葉における情報表現の詳細は分かっていない。したがって、現段階では、脳のようにパターン入出力装置を組み込んだシステムの構築を目指すのは現実的でない。かといって、意味付けされていないパターンを用いたのでは、どんな推論が行われたのか理解不能であろうし、能力の評価も困難である。

そこで、本論文では理解のしやすさを優先し、言葉で表現可能な推論を本システムに実行させることにする。すなわち、「スズメ」などのシンボルに対応するパターン(以下、「スズメ」のコードパターンと呼ぶ)を人工的に作成し、推論の結果として得られるパターンについても、あらかじめ決められたコードパターンとの類似度を計算し、類似度が 0.75 以上で最も近いパターンに対応づけられた意味によって解釈する。

このようなやり方をとるにしても、個々の事物を具体的にどのようなパターンで表現するか、という問題が残る。分散表現においてはパターン間の相関が情報の本質であるから、似たものは類似したパターンに対応づけるべきである。しかし、あらかじめすべての事

物の間の類似性を考慮してパターンを作成するのは、事物の数が増えると無理である。そこで、ここでは次のような簡便な方法を用いるが、これが最適というわけではない。

まず、文脈パターンと target パターンについては、基本的に全くランダムに作成する。したがってパターン間の類似度はほぼ 0 となるが、似ている文脈における類推能力をみるために、一部のパターン間についてはある程度の類似度をもつようにした。具体的には、『飛ぶか』と『翼をもつか』に対応する文脈パターン間のみ、類似度が 0.5 になるよう設定した。なお、この二つの文脈では target パターンも共通のものを用いるが、文脈パターンによって修飾されたパターンは異なり、その解釈も変わる（文脈『飛ぶか』において「飛ぶ」を表す target パターンは、文脈『翼をもつか』では「翼をもつ」と解釈される）。

一方、cue パターンについては、同じカテゴリーに属するもの同士はすべて一定の類似度をもつようにする。具体的には、各カテゴリーに対するコードパターンをランダムに生成し、それを親パターンとして一定量のノイズを加えることによって子パターンを作成する。これらをそのカテゴリーに属する cue のコードパターンとした。このとき、親子パターン間の類似度が a （下記の実験では $a = 0.5$ ）ならば、同じカテゴリーの cue パターン同士は平均で a^2 （下記実験では 0.25）の類似度をもつことになる。

また、図 5 に示すような上位概念の階層を考え、一つ上の概念を表すパターンとの類似度が a になるようにする。このとき、異なるカテゴリーに属する cue パターン同士の類似度は、カテゴリー同士の近さに応じて決まる。例えば、「鳥類」と「哺乳類」のコードパターン間の類似度は $a^2 = 0.25$ なので、それぞれに属する「スズメ」と「ウマ」のコードパターン間の類似度は $a^4 = 0.0625$ となる。

図 5 のような階層構造は、cue パターンの作成に用いるだけでなく、推論に必要な知識を学習する際にも用いる。すなわち、学習させる知識は、できるだけ上位の概念を用いて一般性のある形で表現する。ただし、その上位概念は文脈ごとに異なってよい。例えば、『飛ぶか』という文脈では「スズメ」の上位概念として「鳥類」を用いるが、『動くか』という文脈では「動物」を用いる方がよい。その方が、「スズメは鳥類なので動く」、あるいは単に「スズメは動く」という形で学習するよりも、「ウマは動くか」や「スズメは餌

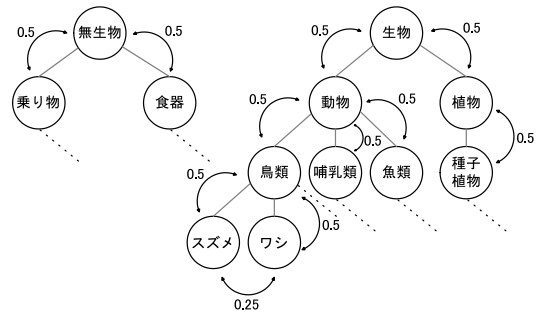


図 5 パターン間の類似度の階層構造

Fig. 5 Hierarchical structure of similarities among patterns.

を食べるか」といった別の未学習の問いでの利用可能性が高いからである。

このように上位概念を用いて知識を表現する際、一般に例外的な知識の扱いが難しい。例えば、「鳥類は飛ぶ」、「ペンギンは鳥類である」と「ペンギンは飛ばない」は論理的に矛盾するからである。しかし、本システムの場合、多少の例外があっても、積極的に上位概念を用いて表現してよい。また、例外についてもそれほど特別な扱いは要しない。「ペンギン」の例でいうならば、『飛ぶか』という文脈では上位概念を用いずに「ペンギンは飛ばない」と教えるが、それ以外の文脈では、他の鳥と全く同じ扱いをしてよい。

4.3 計算機シミュレーション

素子数 $n = 2000$ のモデルにいくつかの知識を学習させ推論を実行させてみた。モデルのパラメータは、すべて 3. の数値実験と同じである。基本的な学習方法については既に述べたが、実際にシステムを構成する場合にはいくつか注意すべき点がある。

一つは、一部の例外的知識だけ学習が難しい（対応する軌道アトラクタを形成しにくい）場合がある、ということである。これは、一般的な知識に対応する軌道アトラクタの影響を強く受けるためと考えられるが、学習できなかった知識だけを重ねて学習することによって対処可能である。

また、一般にニューラルネットでは、新しいパターンを追加で学習しようとする、既学習のパターンをすべて再学習しなければならない場合が多い。本システムでも、非常に多量の知識を追加学習する場合には、既学習知識についても再学習が必要である。しかし、比較的少量の知識を追加するだけで、全体を再学習するのでは非常に効率が悪い。そこで、ここでは追加す

表 2 システムに与えた知識
Table 2 Knowledge given to the system.

文脈	与えた知識
『飛ぶか』	「スズメ」⇒「鳥類」⇒「飛ぶ」 「コウモリ」⇒「飛ぶ」 「ウマ」⇒「哺乳類」⇒「飛ばない」 「サクラ」⇒「植物」⇒「飛ばない」 「ヒコウキ」⇒「飛ぶ」 「ヘリコプタ」⇒「飛ぶ」 「クルマ」⇒「飛ばない」 「バイク」⇒「飛ばない」 「ボート」⇒「飛ばない」 「コップ」⇒「食器」⇒「飛ばない」
『子供は』	「ツバメ」⇒「鳥類」⇒「卵」 「イヌ」⇒「哺乳類」⇒「胎児」 「アサガオ」⇒「種子植物」⇒「種子」
『呼吸は』	「ハト」⇒「鳥類」⇒「肺臓」 「クマ」⇒「哺乳類」⇒「肺臓」
『動くか』	「ワシ」⇒「動物」⇒「動く」 「タンポポ」⇒「植物」⇒「動かない」 「バイク」⇒「乗り物」⇒「動く」 「フォーク」⇒「無生物」⇒「動かない」
『餌は』	「ライオン」⇒「動物」⇒「食べる」 「ツツジ」⇒「植物」⇒「食べない」 「スプーン」⇒「無生物」⇒「食べない」
『葉緑体は』	「キリン」⇒「動物」⇒「なし」 「ユリ」⇒「植物」⇒「あり」 「ポット」⇒「無生物」⇒「なし」

る知識だけを学習し、その影響を確かめることにした。

具体的には、まず、表 2 に示す 25 個の知識を与えて学習を行った。これらは最終的にすべて正しく学習することができたが、それまでの学習回数は、一般的知識も例外的知識もともに平均 10 回であった。この段階で、様々な未学習の問いを与えて推論させるテストを行った。

次いで、表 3 に示す知識を追加して与えた。これらの学習回数も平均 10 回であった。追加学習後、先ほどと同じ問いについて再テストするとともに、追加した知識に関する問いについてもテストを行った。

推論の結果を表 4 に示す。この表で、網が掛かっている部分は、表 3 の知識を追加した後に新たにテストした問いに関連する部分、及び追加学習後に変化した答を示す（それ以外の部分は追加学習の前後で同じ）。

太字の部分は学習した問いに対する答を表すが、当然ながらすべて知識として与えたものと一致している。

表 3 追加した知識
Table 3 Added knowledge.

文脈	追加した知識
『飛ぶか』	「ペンギン」⇒「飛ばない」 「マグロ」⇒「魚類」⇒「飛ばない」
『翼は』	「ヘリコプタ」⇒「もたない」
『子供は』	「ヤマメ」⇒「魚類」⇒「卵」 「サメ」⇒「胎児」
『呼吸器は』	「サケ」⇒「魚類」⇒「エラ」
『死ぬか』	「カモ」⇒「生物」⇒「死ぬ」 「ウシ」⇒「生物」⇒「死ぬ」 「ナノハナ」⇒「生物」⇒「死ぬ」 「ボート」⇒「無生物」⇒「死なない」 「ナイフ」⇒「無生物」⇒「死なない」

それ以外はすべて未学習の問いに対する推論結果であるが、これらは与えられた知識を前提とするならば、どれも妥当な答だと考えられる。

4.4 考 察

上記の結果を詳細に見ると、以下のことが分かる。

(a) cue パターンの類似性に基づく類推

本システムは、例えば「ワシは飛ぶか」という問いに対して、学習していないにもかかわらず、「飛ぶ」と答えている。これは、「ワシ」は「スズメ」と類似したパターンによって表現されているため、「スズメ」→「鳥類」→「飛ぶ」という軌道アトラクタに引き込まれるからである。また、各カテゴリーに属する未学習の cue パターンを多数作成して実験した結果（表 4 で「…」で表されている部分）、このような cue パターンの類似性に基づく類推は、すべてのカテゴリーのすべての問いに関してなされることが確認された。

(b) 文脈パターンの類似性に基づく類推

『飛ぶか』と『翼をもつか』は類似した文脈パターンで表現されるため、後者の文脈において何も学習しなくても、前者の文脈で形成された軌道アトラクタと並行した流れが生じる。これによって、基本的に「飛ぶ」ものは「翼をもつ」という類推がなされる。

(c) 非単調推論

パターンの類似性やカテゴリーに基づく類推では、例外をどのように扱うかが大きな問題となる。例外的な知識を追加することによって矛盾が生じてしまう場合の推論（非単調推論）は、古典的人工知能の一つの重要な研究分野となっているが、決定的な解決策があるとはいえない。しかしこの問題にも、本システムは

表 4 推論結果
Table 4 Reasoning results.

	飛ぶか	翼は	子供は	呼吸は	動くか	餌は	葉緑体は	死ぬか
スズメ	飛ぶ	もつ	卵	肺臓	動く	食べる	なし	死ぬ
ワシ	飛ぶ	もつ	卵	肺臓	動く	食べる	なし	死ぬ
ツバメ	飛ぶ	もつ	卵	肺臓	動く	食べる	なし	死ぬ
カモ	飛ぶ	もつ	卵	肺臓	動く	食べる	なし	死ぬ
ペンギン	飛ぶ	もつ	卵	肺臓	動く	食べる	なし	死ぬ
	飛ばない	もたない						死ぬ
ハト	飛ぶ	もつ	卵	肺臓	動く	食べる	なし	死ぬ
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
コウモリ	飛ぶ	もつ	胎児	肺臓	動く	食べる	なし	死ぬ
ウマ	飛ばない	もたない	胎児	肺臓	動く	食べる	なし	死ぬ
イヌ	飛ばない	もたない	胎児	肺臓	動く	食べる	なし	死ぬ
ウシ	飛ばない	もたない	胎児	肺臓	動く	食べる	なし	死ぬ
クマ	飛ばない	もたない	胎児	肺臓	動く	食べる	なし	死ぬ
ライオン	飛ばない	もたない	胎児	肺臓	動く	食べる	なし	死ぬ
キリン	飛ばない	もたない	胎児	肺臓	動く	食べる	なし	死ぬ
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
マグロ	飛ばない	もたない	卵	エラ	動く	食べる	なし	死ぬ
ヤマメ	飛ばない	もたない	卵	エラ	動く	食べる	なし	死ぬ
サメ	飛ばない	もたない	胎児	エラ	動く	食べる	なし	死ぬ
サケ	飛ばない	もたない	卵	エラ	動く	食べる	なし	死ぬ
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
サクラ	飛ばない	もたない	種子		動かない	食べない	あり	死ぬ
タンポポ	飛ばない	もたない	種子		動かない	食べない	あり	死ぬ
アサガオ	飛ばない	もたない	種子		動かない	食べない	あり	死ぬ
ナノハナ	飛ばない	もたない	種子		動かない	食べない	あり	死ぬ
ツツジ	飛ばない	もたない	種子		動かない	食べない	あり	死ぬ
ユリ	飛ばない	もたない	種子		動かない	食べない	あり	死ぬ
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
ヒコウキ	飛ぶ	もつ			動く	食べない	なし	死なない
ヘリコプタ	飛ぶ	もつ			動く	食べない	なし	死なない
		もたない						死なない
クルマ	飛ばない	もたない			動く	食べない	なし	死なない
バイク	飛ばない	もたない			動く	食べない	なし	死なない
ボート	飛ばない	もたない			動く	食べない	なし	死なない
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
コップ	飛ばない	もたない			動かない	食べない	なし	死なない
フォーク	飛ばない	もたない			動かない	食べない	なし	死なない
ナイフ	飛ばない	もたない			動かない	食べない	なし	死なない
スプーン	飛ばない	もたない			動かない	食べない	なし	死なない
ポット	飛ばない	もたない			動かない	食べない	なし	死なない
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

うまく対処している。

例えば哺乳類である「コウモリ」に関して、『飛ぶか』という問いに対しては「飛ぶ」と答えるが、それ

以外の問いに対しては「ウマ」や「イヌ」など他の哺乳類と同じような推論が行われている。つまり、「コウモリ」→「飛ぶ」の軌道がアトラクタとなるのは、

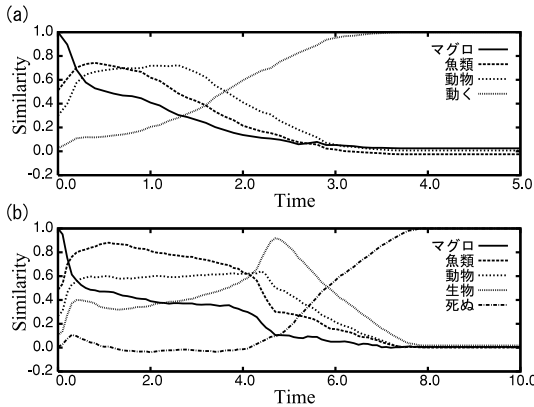


図6 パターン構造を利用した推論の過程

Fig. 6 Process of reasoning using structure of code patterns.

文脈『飛ぶか』のときに限られ、それ以外の文脈では「コウモリ」から「哺乳類」へ向かう流れの方が強い。

また、文脈『飛ぶか』において哺乳類に属する未学習の cue パターンを与えたときも、それが「ウマ」や「イヌ」よりも「コウモリ」にかなり近い限り、「哺乳類」の近くを通して「飛ばない」へ達する。これは、「ウマは哺乳類なので飛ばない」と学習しているのみならず、様々な文脈で「イヌ」→「哺乳類」や「クマ」→「哺乳類」を学習しているために、「コウモリ」の近傍以外では「哺乳類」へ向かう流れの方が強いからである。

(d) 暗黙的知識の利用

分散表現では、パターン間の距離（類似度）を用いて、図5のような関係を暗黙的に表現することができる。一方、シンボル表現または局所表現では、すべての関係を明示的に表現しなければならない。ここに両者の最も本質的ともいえる違いがある。本システムは、このような分散表現の特質を生かした、より高度な類推も可能である。

例えば、すべての魚類に関して、動くかどうかは全く学習しておらず、「動物」であるということも明示的には教えられていない（どの文脈でも「マグロ」→「動物」等の学習はしていない）にもかかわらず、「マグロは動く」と推論している。この過程における回路網の状態 x と各コードパターンとの類似度を図6(a)に示す。横軸は時定数 τ を単位とする時間をとったグラフである。

この図から分かるように、 x は最初「マグロ」から「魚類」の近くへ状態遷移するが、これは知識の学習に

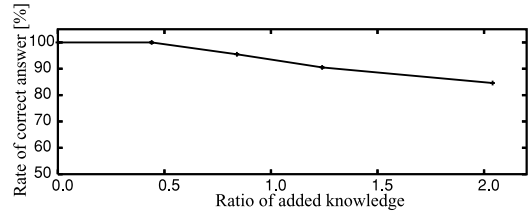


図7 追加学習による既学習知識への影響

Fig. 7 Interference to learned knowledge by additional learning.

より自然に生じた流れによる。それから「動物」の近くを通して「動く」へ達するが、これは「ワシ」に関する学習によって作られた「動物」→「動く」の軌道アトラクタに引き込まれるためである。「マグロ」と「ワシ」はほとんど似ていない（コードパターン間の類似度は 0.0625）が、「魚類」と「動物」のコードパターンは比較的似ている（類似度 0.5）のために、このような類推が可能となっている。

同様に、図6(b)は「マグロは死ぬ」と推論する過程を示したものであるが、「マグロ」→「魚類」→「動物」という流れに乗って状態遷移した後、「カモ、ウシ、ナノハナは生物なので死ぬ」という知識の学習によって作られた軌道アトラクタに引き込まれることによって行われる。

(e) 追加学習

追加学習した表3の知識のうち、「マグロ」をはじめとする魚類に関する知識は、既に学習した知識とはあまり関係がなく、正しい推論に失敗するといった悪影響は見られない。同様に『死ぬか』という問いに関する知識は、他の問いに関する推論にほとんど悪影響を及ぼさない。

一方、「ペンギンは飛ばない」といった例外的知識を追加学習した場合には、他の鳥に関する推論への悪影響が懸念される。しかし、実際には、他の鳥に関する知識を再学習する必要は全くなかった。また、「飛ぶ」ものの中で「ヘリコプタ」は例外的に「翼をもたない」が、このことを追加学習しても、「飛ぶ」ものは「翼をもつ」という文脈パターンの類似性に基づく類推の能力は損なわれなかった。

追加学習の影響をより詳しく見るために、更に多くの知識を追加する実験を行った。具体的には、例外的知識の割合を約 20% に保ったまま、追加する知識を表3のものから増やしていき、最初に学習した知識に関する推論（表4の網がかかってない部分）が正しくて

きるかどうかテストした。結果を図7に示す。グラフの横軸は、既学習知識に対する追加知識の比率、縦軸は既学習知識に関する正答率である。

グラフから分かるように、追加する知識が増えるにつれて正答率が若干下がるが、下がり方は非常に緩やかであり、既学習知識の2倍以上を追加しても80%以上であった。このことは、既学習知識の再学習をそれほど頻繁に行うことなく、必要な知識を次々と追加できることを意味する。

5. む す び

非単調神経回路網が構成する力学系の軌道アトラクタを用いて、局所表現を全く用いることなく純粋にパターンベースの推論を行うモデルを提案した。また、実際に簡単な推論システムを構築し、分散表現の利点を生かした推論能力をもつことを示した。

本モデルはまだ単純な推論のみを扱う初歩的段階にあるが、古典的人工知能や従来型のニューラルネットにはない次のような特性を備えている。

(1) これまで、一般に少数の一般的知識を組み合わせて推論しようとする例外的知識の扱いが難しく、逆に例外の扱いを容易にするためには多数の推論規則または知識が必要であった。これに対して本モデルは、分散表現がもたらす強力な類推能力にほとんど影響を与えず、例外を学習することができる。そのため、非単調推論や常識推論を自然な形で実現することが可能である。

(2) 代表的なニューラルネットの一つである多層パーセプトロンでは、ごく少数の知識を追加学習しただけで学習済みの知識が急激に失われるという、カタストロフィック干渉と呼ばれる現象[7]が知られており、大きな問題となってきた。これに対して本モデルでは、そのような現象が生じないため、知識の追加が容易である。

(3) 従来のニューラルネットモデルの多くは、ごく小規模な例題に関してはうまく動作するが、モデルを大規模化すると必要な素子数や学習時間が爆発的に増大するといった問題を抱える。これに対して本モデルは、原理的にはいくらかでも大規模化でき、それに比例していくらかでも知識を増やすことができる。もしハードウェア化して全素子を並列に動作させることができれば、大規模化しても推論に要する時間はほとんど変わらないし、学習時間も知識の量に比例してしか増えない。

(4) 本モデルは、単にパターンベース推論を行うだけでなく、それを力学系によって実現している点も重要である。自律ダイナミクスをもつ情報処理系の意義については文献[8]などを参照されたいが、その具体的な利点の一つとして、推論結果は静的な空間パターンで表される事物や概念に限られず、時空間パターンで表されるもの、例えば運動の系列であってもよいことが挙げられる。

(5) 本モデルは脳を直接モデル化したものではないが、本モデルの基本原理である軌道アトラクタと選択的不感化法は、それぞれ脳で用いられていることが示唆されている[9]~[11]。推論機構が同じかどうかは今のところ不明だが、少なくとも従来方式よりずっと脳に近いことは間違いないであろう。

以上のことから、筆者らは本モデルには古典的人工知能の限界をも超えるような大きな可能性があると考えている。

ただし、可能性があるだけに残された課題も多い。まず、パターンベースの推論システムの真価は、本来シンボル化が困難な情報を扱ってこそ発揮されるはずであるが、4.2で述べたように、今回はシンボル化可能な知識しか扱っていない。今後、センサからの情報を推論に適したコードパターンに変換するシステムや推論結果を表す出力パターンに従ってアクチュエータを動かす制御系などの開発が必要である。また、コード化の方法ないし暗黙的知識の表現方法についても、より深く研究する必要がある。

もう一つの大きな課題は、より複雑な推論が行えるよう、モデルを拡張することである。例えば、本モデルの回路網を複数運動させることによって、多段階の推論が実現できると考えている。更に、このようなモデルを脳の推論や思考のメカニズムと比較することも重要で興味深い課題である。

謝辞 本研究は、文部科学省科学研究費補助金基盤研究(B)(No.15300068)、特定領域研究「情報学」(No.16016207)及び「情報爆発」(No.18049008)の補助を受けた。

文 献

- [1] S. Harnad, "The symbol grounding problem," *Physica D*, vol.42, pp.335-346, 1990.
- [2] J. McCarty and P.J. Hayes, "Some philosophical problems from the standpoint of artificial intelligence," *Machine Intelligence*, vol.4, pp.463-502, 1969.
- [3] 甘利俊一, 酒田英夫(編), 脳とニューラルネット, 朝倉

書店, 1994.

- [4] 森田昌彦, 松沢浩平, 諸上茂光, “非単調神経素子の選択的不感化を用いた文脈依存的連想モデル,” 信学論 (D-II), vol.J85-D-II, no.10, pp.1602-1612, Oct. 2002.
- [5] 森田昌彦, 村田和彦, 諸上茂光, 末光厚夫, “選択的不感化法を適用した層状ニューラルネットの情報統合能力,” 信学論 (D-II), vol.J87-D-II, no.12, pp.2242-2252, Dec. 2004.
- [6] M. Morita, “Memory and learning of sequential patterns by nonmonotone neural networks,” Neural Networks, vol.9, pp.1477-1489, 1996.
- [7] M. McCloskey and N. Cohen, “Catastrophic interference in connectionist networks: The sequential learning problem,” The Psychology of Learning and Motivation, vol.24, pp.109-164, 1989.
- [8] 松山隆司, 杉本晃宏, 佐藤洋一, 川島宏彰, “人間と共生する情報処理システムの実現を目指して,” 人工知能誌, vol.19, no.2, pp.257-266, March 2004.
- [9] 末光厚夫, 諸上茂光, 森田昌彦, “下側頭葉における文脈依存的連想の計算論的モデル,” 信学論 (D-II), vol.J87-D-II, no.8, pp.1665-1677, Aug. 2004.
- [10] 末光厚夫, 諸上茂光, 森田昌彦, “下側頭葉ニューロンの想起活動に関する計算論的考察,” 日本神経回路学会誌, vol.9, no.3, pp.174-180, 2002.
- [11] 宮澤泰弘, 末光厚夫, 森田昌彦, “選択的不感化理論に基づく海馬ニューロン活動のモデル化,” 日本神経回路学会誌, 印刷中.

(平成 18 年 4 月 25 日受付, 8 月 25 日再受付)



末光 厚夫 (正員)

平 10 筑波大・工学システム学類卒・平 15 同大学院博士課程了。同大学院システム情報工学研究科研究員を経て, 平 18 島根大学総合理工学部教務職員, 現在に至る。神経回路モデルによる脳の記憶機構の研究に従事。



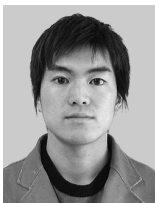
森田 昌彦 (正員)

昭 61 東大・工・計数卒・平 3 同大学院博士課程了。日本学術振興会特別研究員, 東京大学工学部助手を経て, 平 4 筑波大学電子・情報工学系講師。現在, 同大学院システム情報工学研究科助教授。脳の情報処理機構及び神経回路網による情報処理の研究に従事。平 5 日本神経回路学会研究賞, 平 6 同学会論文賞, 平 11 日本心理学会研究奨励賞受賞。



山根 健

平 17 筑波大・工学システム学類卒。現在, 同大学院博士課程システム情報工学研究科在学中。神経回路モデルの研究に従事。



蓮尾 高志

平 18 筑波大・工学システム学類卒。現在, 同大学院博士課程システム情報工学研究科在学中。神経回路モデルの研究に従事。