アソシアトロンと連想記憶

森田 昌彦

はじめに

連想記憶とは，情報をパターンとして表現した上で分散的に記銘し，そのパターンの一部を与えると元のパターン全体が再現される，という形で情報を蓄え，読み出す記憶方式である．正確に言えば，これは「自己想起型」の連想記憶と呼ばれるものであり，この他に二つのパターンの一方から他方を読み出す「相互想起型」の連想記憶もあるが，後者は層状の神経回路によって実現され，パターン認識を行うのと本質的な差はない．

ここでは，回帰型(素子同士が双方向で結合している)神経回路によって実現される，自己想起型の連想記憶を扱うが，そのような神経回路モデルとして1969年に中野によって提案されたのがアソシアトロン[1]である．その後，1982年にHopfieldが類似のモデルを提唱し[2]，連想記憶だけでなく，巡回セールスマン問題のような最適化問題を解くのに使えることを示して多くの注目を集めた．このいわゆるHopfieldモデルに対し，スピングラスという物理系との類似性から多数の物理学者が興味を持ち，種々の数理的解析を行った．

その結果，記憶容量をはじめとして，このモデルが持つ種々の性質が1980年代に明らかになったが，後述するモデルの問題点はそのままであった．これらを解決する方法が提案されたのは1990年以降であり，その多くは筆者が行ったものである．本稿では，アソシアトロンの原理から出発して，問題点とその解決法，脳との関係など，その後の研究成果について極力わかりやすく紹介したい．

原　理

アソシアトロンの原理をなるべく直観的に理解しやすいよう，Hopfieldモデルで導入された「エネルギー」の概念を用いて説明する．これは，システム(神経回路網)の安定性を示す指標であり，「エネルギー」が低い状態ほど取りやすいことを意味している．

簡単のため，±1のいずれかを出力する素子(モデル化されたニューロン)が二つある場合で考える．この二つの素子が，双方向に同じ強さで結合していた場合，各素子の出力値を掛け，さらに結合の強度(興奮性の場合は正，抑制性の場合は負の値を取るものとする)を掛けた値の−1倍を，この2素子からなる系の「エネルギー」と定義する．そうすると，例えば結合強度が正，つまり互いに興奮性に結合していた場合，一方が1で他方が−1を出力する状態より両方が1または−1を出力する状態の方が低エネルギーとなる．実際，前者は不安定で，後者が安定である．結合強度が負の場合はその逆となる．

素子がn(>2)個あるシステムの場合には，うち二つの素子の組み合わせすべてについて同様にエネルギーの値を計算し，それらを全部足し合わせたものを，そのシステム全体のエネルギーと定義する．そうすると，系の状態は基本的にエネルギーが減る方向にしか変化しないため，やはりエネルギーが低い状態ほど安定となる．素子の出力が2値ではなく連続値の場合にも，ほぼ同様なエネルギーが定義できる．

システム全体が取りうるすべての状態について，エネルギーの値を計算してグラフの「高さ」として表現すると，図1のようなエネルギーの「地形」が描かれることになる(図では状態を横軸1次元で表しているが，実際にはn次元の空間である)．この地形中の一点がシステムの現在状態を表すが，この点は時間と共にエネルギーの谷の方へ移動していき，最終的に谷底で留まることになる．このような谷底のことを「アトラクタ」と呼ぶ．

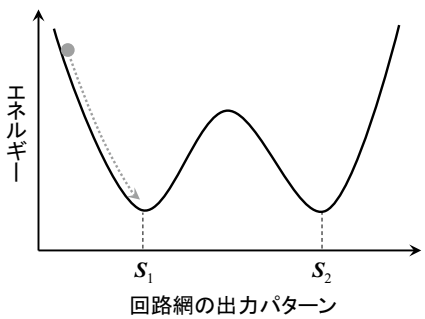

図 1

もりた まさひこ　筑波大学システム情報系教授

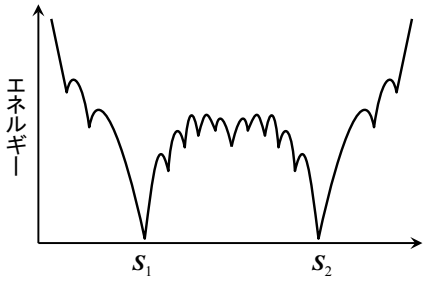

図 2

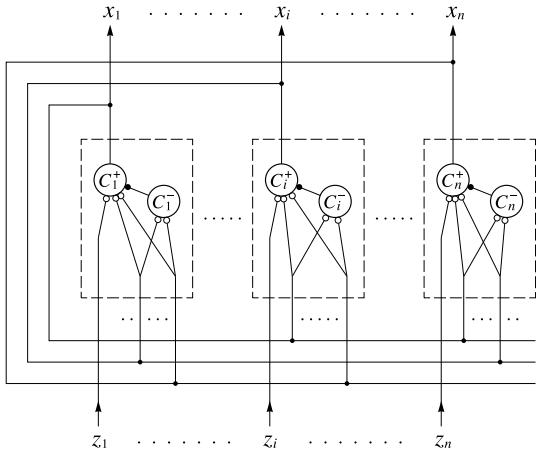

図 4

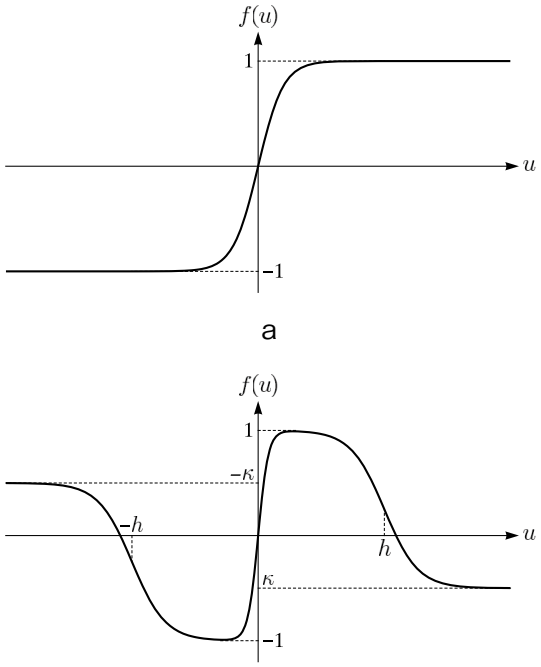

図 3

さて，アソシアトロンにおける記銘は，記憶したい n 素子の活動パターン S がアトラクタ，つまりエネルギーの谷になるよう素子間の結合強度を変えることによって行う．具体的には，1と1のように同じ値を出力する素子間の結合を強め，1と−1の素子間の結合を弱める(負の方向に強化する)のが最も単純な方法である(自己相関型連想記憶)．S がアトラクタになっていれば，その周囲の状態を与えたとき，システムは状態 S に変化する．これが想起の過程である．また，その後何も入力を与えなくても状態 S を保持するが，これは短期記憶に相当する．

問題点

アソシアトロンのような自己相関型の連想記憶は，脳の記憶のモデルとして考案された．実際，「一部の素子が壊れても特定の記憶が失われることはない」，「多くのパターンを記憶させるほど想起が難しくなる」など，脳の記憶に似た性質を持つ．また，各素子は他の素子からの信号に結合強度を掛けて足し合わせるという単純な計算をするだけで，学習アルゴリズム(結合強度の修正方法)も単純であり，生物学的に無理がない．しかし，一方で脳のモデルとしてはいくつかの問題点がある．

一つは，記憶装置としての性能がかなり低いことである．例えば，記憶できるパターンの数(記憶容量)は素子数の15%程度であるし，容量以下であっても記憶量がある程度増えると想起能力が大きく低下し，記憶したパターンとの誤差も増大する．また，記憶するパターン同士が無相関でないと，干渉が生じて容量や想起能力が著しく低下する．これらの問題は，学習アルゴリズムを変えることによってある程度解消できるが，そうすると計算が複雑になり生物学的な妥当性に疑問が生じる．

関連する問題点として，記憶したパターン以外のアトラクタである「偽の記憶」がある．偽の記憶は，複数のパターンを記憶すると必ず生まれ，例外的な状況を除いてその数は真の記憶の数よりも圧倒的に多い．この問題は，学習アルゴリズムをどのように変えても解消しない．かつて「夢は偽の記憶を想起した結果であり，それを反学習することによって偽の記憶を消去している」という仮説[3]に基づいてモデル[4]が提出されたが，実際にはそのような反学習にほとんど効果はない．

脳のニューロン活動との違いも問題点の一つである．例

えば，アトラクタはポジティブフィードバックによって生じると考えてよいので，記憶したパターンを想起した状態において，各素子は必ずほぼ取りうる最大または最小の値を出力する．しかし，実際のニューロン活動はそうなっていない．例えば，ある特定の図形を提示した後に持続的に活動するサル下側頭葉ニューロンの発火頻度は，取りうる最大値よりもかなり小さく，また図形によって違っていた[5]．また，サルに2枚の図形を連合学習させ，一方の図形から他方を想起させると，ニューロン群の活動が徐々にシフトする現象が観察された[6]が，このような形でのパターンからパターンへの連想は，自己相関型の連想記憶モデルでは実現できない．

非単調神経回路網

上記のような問題点の原因のほとんどは，学習アルゴリズムではなく，システムの動作規則，つまりダイナミクスにある．例えば，Hopfield モデルのエネルギーの地形は，実際には図1のような滑らかな形状ではなく，むしろ図2のような切り立った形状をしている．つまり，アトラクタに近いところほど傾斜が急になっており(アトラクタに引きつける力が強いことを意味する)，また強いアトラクタ(記憶したパターン)は孤立していて，その中間には偽の記憶だらけの領域が広がっている．

この問題を解決するために考案されたのが，各素子の入出力関係を表す関数(出力関数)を，これまで用いられてきた図3aのような単調増加飽和型(シグモイド)関数から，図3bのような非単調関数に変える方法である[7]．このようなシステム(非単調神経回路網)では，エネルギーが減る方向以外に変化することがあるものの，エネルギー地形が図1のような滑らかなものになる．これは，強いアトラクタの周辺では，各素子が受ける入力の絶対値が大きくなり，各素子の出力が0に近づいてアトラクタの吸引力が弱まるからである．

このようにエネルギー地形が滑らかになる結果，偽の記憶がほとんど存在しなくなり，正しい記憶パターンを想起する能力が大きく向上する．正しく想起できない場合には，しばしばカオス状態になって出力パターンがいつまでも変化しつづける．

非単調神経回路網のもう一つ重要な性質は，孤立した点アトラクタだけでなく，線状に連なるアトラクタ(エネルギーの溝)を形成できる，ということである．しかも，その中に緩やかな流れが生じるようにできるため，システムの状態は，溝の底に向かって強く引き寄せられた後，溝に沿ってゆっくり移動することになる．このようなエネルギーの溝のことを「軌道アトラクタ」と呼び，これによって時間的に徐々に変化する時空間パターンの記憶が実現できる[8]．

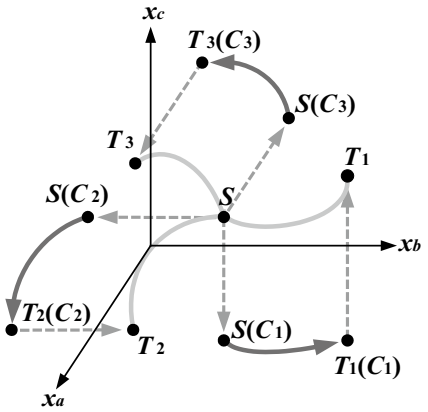

図 5

ただし，非単調出力関数を持つ素子というのは，脳のニューロンのモデルとしては不自然である．より生物学的妥当性の高いモデルとして，図4のような興奮性の細胞と抑制性の細胞のペアからなるユニットを相互結合したものが提案されている[9]．ユニットへの入力を0から徐々に増やしていったとき，最初は出力も増加していくが，ある程度まで増加すると抑制性細胞の働きによって減少に転じる．

このモデルは，比較的少数のユニットのみが出力を出す状態をアトラクタとして記憶することができるが，このときのモデルの振る舞いは，前述したサル下側頭葉ニューロン群の活動とよく合致する．例えば，ユニットの出力値は0と最大値に2極化するのではなく，その中間にも分布する．また，軌道アトラクタを形成することもできるため，あるパターンを与えると，別のパターンに徐々に変化するような連想が可能である[10]．

文脈依存的連想

人間は，ある一つの手掛かりから状況や文脈によって異なるものを連想することができる．ところが，連想記憶の場合にこれを実現するのは難しい．一般には，文脈を表す素子群を用意しておいて，「手掛かり＋文脈」を表すパターンと想起すべきパターンとを連合して記憶する方法が用いられるが，この方法は極めて効率が悪く，記憶パターンや文脈の数がごく少数の場合にしかうまくいかない．

理由を説明するために，手掛かりとなるパターン S_i と文脈を表現するパターン C_j の組み合わせに応じて異なるパターン T_{ij} を想起するという問題を考える．手掛かりと文

脈がそれぞれ m 種類あったとすれば，m^2通りの連想を記憶することになる．このとき，各手掛かりパターン S_iは，それぞれ m 個の異なるパターン T_{i1}，…，T_{im}と連合されることになるが，これは T_{i1}，…，T_{im}を平均したパターン T_iと連合されたのと同じことである．したがって，システムは手掛かり S_iから T_iを想起しようとする．同様に，文脈 C_jから，T_{1j}，…，T_{mj}を平均したパターンを想起しようとする．結果として，目的とする T_{ij}は想起されない．

この問題を解決するために考案されたのが，「選択的不感化法」である[11]．これは，S_iを表現する素子群の一部(通常は約半分)を C_jに応じて選び，それらを機能させない(常に中立値を出力させる)という手法である．S_iが±1 を成分とする 2 値パターンであった場合，選択的不感化を行うことにより {1，0，−1} の 3 値パターンに変化する．この 3 値パターンは文脈情報を含んでいるが，C_jは T_{ij}と直接連合されないため，前述の問題が生じない．

さて，この手法を非単調神経回路網に適用すると，かなり自在な連想が可能になる．例えば，パターン S_1，S_2，S_3，S_4を点アトラクタとして記憶した上で，C_1という文脈(約半数の素子が不感化された状態)において $S_1 \rightarrow S_2 \rightarrow S_3 \rightarrow S_1$という軌道アトラクタを形成し，同様に文脈 C_2，C_3，C_4において $S_1 \rightarrow S_3 \rightarrow S_4 \rightarrow S_1$，$S_1 \rightarrow S_4 \rightarrow S_2 \rightarrow S_1$，$S_2 \rightarrow S_4 \rightarrow S_3 \rightarrow S_2$という軌道アトラクタをそれぞれ形成すれば，文脈を切り替えることによって，どのようなパターン系列であっても想起することができる．その仕組みを模式的に示したのが図 5 である．

最初，何の文脈も与えない，すなわち選択的不感化を行わない状態でパターン S_1の一部を与えたとすると，システムは完全なパターン S_1を想起し，その状態を維持する．そこで文脈 C_1を与えると，システムの状態はある部分空間に射影され，そこで S_2に向かって遷移する．ただし，このとき約半数の素子は不感化されており，想起されるのは成分の約半数が 0 である 3 値パターンである．しかし，それが S_2に近づいた時点で選択的不感化を解除するとすぐに完全なパターン S_2が想起されるし，別の文脈 C_4に切り替えれば S_4に向かっての状態遷移が生じる．

このような仕組みは，実際の脳にもある可能性がある．実際，画面の色に応じて想起すべき図形が変化する課題を実行中のサル下側頭葉のニューロン活動[12]には，これまでの考え方では説明がつかないものがあったが，選択的不感化を取り入れたモデルはその点も含めて非常によく整合する[13]．

むすび

連想記憶装置アソシアトロンと，それを発展させた回帰型神経回路モデルについて解説した．紙面の関係で詳細には触れていないが，興味のある方は解説論文[14〜16]その他の文献を参照されたい．

残念ながら，多層神経回路に比べて連想記憶が工学的に応用されることはほとんどなかったし，近年の deep learning を中心とした人工知能ブームとも今のところ無縁である．しかし，人間の知性は脳の記憶の仕組みと密接に関係していると考えられるため，未経験の状況において適切に判断する能力などを持つ人工知能を実現するためには，連想記憶を用いる必要があると思われる．具体的にどう用いるかが難しいが，一つの試みとしては，非単調神経回路網と選択的不感化を用いたパターンベース推論[17]がある．

本解説が，連想記憶の新たな活用につながれば幸甚の至りである．

文献

1) 中野 馨．アソシアトロン―連想記憶のモデルと知的情報処理．昭晃堂；1979.
2) Hopfield JJ. Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci USA. 1982；79：2554-8.
3) Crick FC, Mitchison G. The function of dream sleep. Nature. 1983；304：111-4.
4) Hopfield JJ, Feinstein DI, Palmer RG. 'Unlearning' has a stabilizing effect in collective memories. Nature. 1983；304：158-9.
5) Miyashita Y. Neuronal correlate of visual associative long-term memory in the primate temporal cortex. Nature. 1988；335：817-20.
6) Sakai K, Miyashita Y. Neural organization for the long-term memory of paired association. Nature. 1991；354：152-5.
7) Morita M. Associative memory with nonmonotone dynamics. Neural Netw. 1993；6：115-26.
8) Morita M. Memory and learning of sequential patterns by nonmonotone neural networks. Neural Netw. 1996；9：1477-89.
9) Morita M. Computational study on the neural mechanism of sequential pattern memory. Brain Res Cogn Brain Res. 1996；5：137-46.
10) Morita M, Suemitsu A. Computational modeling of pair-association memory in inferior temporal cortex. Brain Res Cogn Brain Res. 2002；13：169-78.
11) 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．電子情報通信学会論文誌(D-Ⅱ)．2002；85：1602-12.
12) Naya Y, Sakai K, Miyashita Y. Activity of primate inferotemporal neurons related to a sought target in pair-association task. Proc Natl Acad Sci USA. 1996；93：2664-9.
13) 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．電子情報通信学会論文誌(D-Ⅱ)．2004；87：1665-77.
14) 森田昌彦．連想記憶の神経回路モデル．甘利俊一，酒田英夫，編．脳とニューラルネット．朝倉書店；1994．p. 127-42.
15) 森田昌彦．時系列パターンの学習・記憶の計算論．外山敬介，杉江 昇，編．脳と計算論．朝倉書店；1997．p. 54-69.
16) 森田昌彦．記憶と思考の神経回路モデル．丹治 順，吉澤修治，編．脳の高次機能．朝倉書店；2001．p. 211-29.
17) 山根 健，蓮尾高志，末光厚夫，他．軌道アトラクタを用いたパターンベース推論．電子情報通信学会論文誌(D)．2007；90：933-44.