

自己組織化ニューラルネットによる概念形成モデル

Self-organizing Neural Networks for Concept Formation Model

小谷中 洋介, 本間 経康, 酒井 正夫, 阿部 健一

Yosuke Koyanaka, Noriyasu Homma, Masao Sakai, Kenichi Abe

東北大学大学院工学研究科

Graduate School of Engineering, Tohoku University

キーワード : ニューラルネットワーク (Neural networks), 概念形成 (concept formation),
自己組織化 (self-organizing), Hebbの学習則 (Hebbian rule)

連絡先 : 〒980-8579 仙台市青葉区荒巻字青葉 東北大学 大学院 工学研究科 電気・通信工学専攻 阿部研究室
Tel.: (022)217-7074, Fax.: (022)263-9290, E-mail: koyanaka@abe.ecei.tohoku.ac.jp

1. はじめに

概念の形成は生体の脳における重要な機能の一つである。我々が『リンゴ』の特徴を挙げよと言われた場合「球体」「赤」「おいしい」などと挙げることができるのは、我々の中に【形】【色】【味】といった概念が形成されており、さらに『リンゴ』に関するそれらの情報（「球体」「赤」「おいしい」）を記憶として既に取得しているからだといえる。一方、我々が全く概念を持たない状態において初めて『リンゴ』と出会った場合には、例え観測対象である『リンゴ』から「球体」「赤」「おいしい」などの特徴的な情報が観測されていたとしても、それらを同時に気付くことはおそらく不可能である。なぜならば、感覚的なものとして入力される情報を、そのような論理的情報として記憶するためには【形】【色】【味】といった概念が必要不可欠だからである。このような概念は、おそらく観測対象から得られる膨大な量の感覚的情報のうち、最も印象的だった情報を論理的

に処理しようとする過程で、発見的に一ずつ形成されると考えられる。

また【形】【色】【味】といった概念が既に形成されている段階で、初めて『バナナ』に出会った場合を考える。このとき、脳は『バナナ』より観測される膨大な感覚的情報から「棒状」「黄」「おいしい」といった論理的な情報を容易に取得し記憶するはずである。すなわち、観測的对象から、既に形成されている概念に関する情報を同時に取得することは、比較的容易であると考えられる。

さらに、観測対象と過去に取得した記憶の同一性の判別には、概念ごとの論理的な情報の比較が用いられていると考えられる。例えば【形】という概念しか形成されていない段階で『リンゴ』と『ボール』の同一性を判別した場合には、ともに「球体」という論理的情報を持っており、両者を異なるものと判別することが困難である。一方で、【味】という概念を形成することが出来れば「おいしい」「おいしくない(食べられない)」という

情報の比較によって，両者を論理的に区別することが可能となるはずである．

このような脳における概念形成の特徴は，以下のようにまとめることができる．

- 1) 新しい概念は一つずつ形成される．
- 2) 既形成の概念に関する論理的情報は一度にまとめて修得できる
- 3) 概念ごとの論理的情報を比較することで，観測対象と過去に記憶した対象の同一性を比較できる．

本研究では，外界からの情報に対し自己の構造を適切に変える自己組織能力を持つニューラルネットワーク¹⁾を基にして，先にまとめた特徴を考慮した概念形成モデルを提案し，また，簡単な追加学習問題に対する動作検証を行うことでその有効性を示す．

2. 概念形成モデル

2.1 観測対象の定義

観測対象より取得される情報の扱いを定義する．まず， N 個の観測対象の集合を S ，

$$S = \{x^1, x^2, \dots, x^i, \dots, x^N\}, \quad (1)$$

と定義する．ここで， x^i ， $i = 1, 2, \dots, N$ は i 番目の観測対象ベクトルであり， n 個の概念より，

$$x^i = [x_1^i, x_2^i, \dots, x_n^i]^T \in \mathbb{R}^n, \quad (2)$$

のように定義できる．ここで， x_j^i ， $j = 1, 2, \dots, n$ は i 番目の観測対象の j 番目の概念に関する情報であり，本研究では簡単のためその概念の有効性の有無だけを考慮して， $x_j \in \{0, 1\}$ のように2値で表す．

2.2 ニューロンモデル

提案ニューロンモデルをFig. 1に示す．また，二

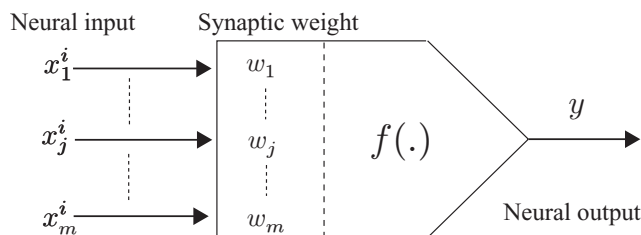


Fig. 1 提案ニューロンモデル²⁾．

ニューロンの結合重み w を，

$$w = [w_1, w_2, \dots, w_m]^T \in \mathbb{R}^m, \quad m < n, \quad (3)$$

のように定義する．ここで， m は既形成の概念の数であり， $w = x^i \in S$ が成り立つならば，提案モデルにおいて観測対象 x^i がニューロンに記憶されたと考える．すなわち，観測対象の概念ごとの情報 x_j^i が，ニューロンの結合重み w_j として記憶されることを意味する．

また，対象 x^i より観測した場合のニューロンの出力を，次式のようなRBF関数により定義する．

$$y = f(x^i, w), \\ \equiv \exp\left(-\sum_{j=1}^m (w_j - x_j^i)^2\right). \quad (4)$$

これより，対象 x^i より観測される情報とニューロンが取得している記憶 w の差が小さいほど出力がニューロンの出力が大きくなり，全て一致した場合は $y(x^i) = 1$ を出力することが分かる．

2.3 競合学習

複数のニューロンより構成されるニューラルネットワークにおいて，観測対象 x^i に最も近い記憶を持つニューロンを認識対象の候補として選択する．本研究では，そのための手法として一般的な自己組織化マップ (Self-Organizing Map: SOM)³⁾における競合学習を用いる．以下に，本研究で用いた競合学習のアルゴリズムを示す．

始めに，観測対象に対して式 (4) の出力が最も

大きいニューロンを，候補ニューロン c とする．

$$c = \arg \max_k y^k(x^i). \quad (5)$$

ここで， y^k は k 番目のニューロンの出力を意味する．このとき，式(5)を満たすニューロンが複数($c_l, l = 1, 2, \dots$)存在した場合は，それぞれのニューロンの結合重みの総和 $W^{c_l} = \sum_{j=1}^m w_j^{c_l}$ が最も大きいニューロンを候補とする．

$$c = \arg \max_{c_l} W^{c_l}. \quad (6)$$

さらに，式(6)を満たすニューロンが複数存在した場合はランダムで選択する．

2.4 結合重みの改変

候補として選択されたニューロン c の結合重み w^c を，Hebbの学習則⁴⁾を基にして改変する．Hebbの学習則とは，二つ以上のニューロンが同時に発火した場合は，そのニューロン間の結合が強化されるというものである．本研究では，観測対象の概念ごとの情報 x_j^i の有無により，候補ニューロン c の対応する結合重み w_j^c 次のように改変する．

$$w_j^c = \begin{cases} 1, & (x_j^i = 1), \\ 0, & (x_j^i = 0). \end{cases} \quad (7)$$

2.5 概念形成アルゴリズム

以下に，本研究で用いた概念形成アルゴリズム例を示す．

- 1) 初期状態 $t = 0$ において，ニューロンユニット数を $M(t) = M(0) = 1$ ，即形成概念数を $n(t) = n(0) = 1$ ，そして，その結合重みベクトルを $w^1(t) = [w_1^1(0)] = [0]$ と設定する．
- 2) 観測対象の集合 S より，任意の対象 $x^{i(t)} \in S$ ， $i(t) \in \{1, 2, \dots, N\}$ を選択し，その対象を観測した場合のニューロンの出力 $y^k(x^{i(t)})$ ， $k = 1, 2, \dots, M(t)$ と2.3節で示した競合アルゴリズムにより候補ニューロン c を選択する．

3) 以下の場合ごとに結合重みを改変する．

- (a) 候補ニューロンの出力が $y^c(x^{i(t)}) = 1$ であり，かつ観測対象の情報が完全にニューロンの結合重みに記憶されている($x^{i(t)} = w^c$)場合は，改変を行わない．
- (b) 候補ニューロンの出力が $y^c(x^{i(t)}) = 1$ あり，かつ観測対象の情報がまだ完全にニューロンの結合重みに記憶されていない($x^{i(t)} \neq w^c$)場合は，新たな概念を形成する．

$$n(t+1) = n(t) + 1. \quad (8)$$

また，その新たなる概念に関する情報を，候補ニューロンは観測対象から取得し，その他のニューロンの記憶は便宜的に0とする．

$$w_{n(t)+1}^k(t+1) = \begin{cases} x_{n(t)+1}^{i(t)}, & (k = c), \\ 0, & (\text{otherwise}). \end{cases} \quad (9)$$

- (c) 候補ニューロンの出力が $y^c < 1$ の場合は，新たに $M(t) + 1$ 番目のニューロンを生成する．

$$M(t+1) = M(t) + 1. \quad (10)$$

また，観測対象から取得した情報 $x^{i(t)}$ を，新たに生成したニューロンの結合重み $w^{M(t)+1}$ として取得する．

$$w_j^{M(t)+1}(t+1) = x_j^{i(t)}, \quad (11)$$

$$j = 1, 2, \dots, n(t).$$

- 4) $t = t + 1$ とし 2) に戻る．

3. 動作検証

提案モデルと2.5節で示した概念形成アルゴリズムの有効性を検証するために，簡単な追加学習問題に対する動作検証を行う．以下に，用意した3つ

の観測対象

$$\boldsymbol{x}^1 = [0, 0, 1]^T, \quad (12)$$

$$\boldsymbol{x}^2 = [0, 1, 1]^T, \quad (13)$$

$$\boldsymbol{x}^3 = [1, 0, 1]^T, \quad (14)$$

を認識・学習することで，概念が形成される一連の流れを示す．

まず初期状態 $t = 0$ における，ネットワークモデルは以下のような状態である．

$$M(0) = 1,$$

$$n(0) = 1,$$

$$\boldsymbol{w}^1(0) = [0].$$

ここで，最初の観測対象として $\boldsymbol{x}^{i(0)} = \boldsymbol{x}^2$ を選択した場合，ニューロンの出力は以下のように決まる．

$$y^1(\boldsymbol{x}^2) = 1.$$

これより，概念形成アルゴリズムの条件3)-(b)が適用され $n(1) = 2$ となり，ネットワークの結合重みは，

$$\boldsymbol{w}^1(1) = [0, 1]^T,$$

のように改変される．

次に， $t = 1$ における観測対象として $\boldsymbol{x}^{i(1)} = \boldsymbol{x}^1$ を選択した場合，ニューロンの出力は以下のように求まる．

$$y^1(\boldsymbol{x}^1) = 0.3679.$$

これより，概念形成アルゴリズムの条件3)-(c)が適用されることから，新たにニューロンが生成され $M(2) = 2$ となり，ネットワークの結合重みは次のように改変される．

$$\boldsymbol{w}^1(2) = [0, 1]^T,$$

$$\boldsymbol{w}^2(2) = [0, 0]^T.$$

さらに， $t = 2$ における観測対象として $\boldsymbol{x}^{i(2)} = \boldsymbol{x}^3$ を選択した場合，ニューロンの出力は以下のように

に求まる．

$$y^1(\boldsymbol{x}^3) = 0,$$

$$y^2(\boldsymbol{x}^3) = 0.3679.$$

これより，概念形成アルゴリズムの条件3)-(c)が適用されることから， $M(3) = 3$ となり，ネットワークの結合重みは次のように改変される．

$$\boldsymbol{w}^1(3) = [0, 1]^T,$$

$$\boldsymbol{w}^2(3) = [0, 0]^T,$$

$$\boldsymbol{w}^3(3) = [1, 0]^T.$$

同様に， $t = 3$ における観測対象として $\boldsymbol{x}^{i(3)} = \boldsymbol{x}^1$ を選択した場合，ニューロンの出力は，

$$y^1(\boldsymbol{x}^1) = 0.3679,$$

$$y^2(\boldsymbol{x}^1) = 1,$$

$$y^3(\boldsymbol{x}^1) = 0.3679,$$

であり，概念形成アルゴリズムの条件3)-(b)が適用されることから， $n(4) = 3$ となり，ネットワークの結合重みは次のように改変される．

$$\boldsymbol{w}^1(4) = [0, 1, 0]^T,$$

$$\boldsymbol{w}^2(4) = [0, 0, 1]^T = \boldsymbol{x}^1,$$

$$\boldsymbol{w}^3(4) = [1, 0, 1]^T = \boldsymbol{x}^3,$$

この段階で，観測対象 \boldsymbol{x}^1 ， \boldsymbol{x}^3 が，ニューロンの結合重み \boldsymbol{w}^1 ， \boldsymbol{w}^3 としてそれぞれ記憶されたことになる．従って，以後，学習済みの観測対象 \boldsymbol{x}^1 ， \boldsymbol{x}^3 が選択された場合は，概念形成アルゴリズムの条件3)-(a)が適用され，ネットワークの改変は行われない．

その後，観測対象として $\boldsymbol{x}^{i(t)} = \boldsymbol{x}^2$ が選択された場合，ニューロンの出力は，

$$y^1(\boldsymbol{x}^2) = 0.3679,$$

$$y^2(\boldsymbol{x}^2) = 0.1353,$$

$$y^3(\boldsymbol{x}^2) = 0.1353,$$

となり，概念形成アルゴリズムの条件3)-(c)が適用されることから，ニューロンが新たに生成され ($M(t+1) = 4$)，ネットワークの結合重みは，

$$\begin{aligned}w^1(t+1) &= [0, 1, 0]^T, \\w^2(t+1) &= [0, 0, 1]^T = \boldsymbol{x}^1, \\w^3(t+1) &= [1, 0, 1]^T = \boldsymbol{x}^3, \\w^4(t+1) &= [0, 1, 1]^T = \boldsymbol{x}^2,\end{aligned}$$

のように改変される．これより，残っていた観測対象 \boldsymbol{x}^2 もニューロンの結合重み w^4 として記憶され，全ての観測対象が提案モデルに完全に学習されたことになる．

4. まとめ

本研究では，自己組織化ニューラルネットワークによる概念形成モデルと，その認識・学習アルゴリズムを提案した．また，簡単な追加学習問題に対する動作検証を行うことで，脳における概念形成の特徴を考慮した認識と追加学習を提案モデルによって実現できることを示した．

動作検証では，概念形成を行う過程において，どの観測対象の情報とも一致しない記憶 w^1 が生成されたが，このような冗長な記憶が生体の脳での概念形成における何に対応し，どのような意味を持つかの考察は，今後の課題である．また，提案モデルの特性や能力をより詳細に考察するためには，概念の情報を2値ではなく連続値で扱った場合や，観測対象から得られる情報に誤差や欠落を生じさせた場合などの，より難しい問題における動作検証を試みる必要がある．

参考文献

- 1) 福島邦彦：神経回路と自己組織化 - 自己組織化のモデル - ，共立出版株式会社 (1979)
- 2) N. Homma, M. Sakai, K. Abe, and H. Takeda:

A Self-Organizing Concept Formation Network, Proc. 2003 SICE Annual Conf., 2003

- 3) T. Kohonen: Self-organized formation of topologically correct feature maps, Biol. Cybern., vol. 43, pp. 59-69, 1982
- 4) D. O. Hebb: The Organization of Behavior, Wiley, New York, 1949