原著論文

非一様な相関を持つパターンを記憶した 神経回路モデルによる典型性効果の再現†

平原 誠*1・永野 俊*2

概念階層に関する重要な心理学的知見に典型性効果がある。これがもたらす心理現象としては、ある概念(上位概念)に属するインスタンス(下位概念)を列挙させると典型的なインスタンスほど先に想起されること、あるインスタンスがある概念に属するかどうかを判断させるとそのインスタンスが典型的であるほど判断時間が短くなること、などが知られている。本研究では、典型性効果の理解への第一ステップとして、典型的なインスタンスほど高確率かつ高速に想起する神経回路モデルを提案する。

キーワード：神経回路モデル，典型性効果，概念階層，非一様な相関を持つパターン

1. はじめに

概念階層に関する重要な心理学的知見に典型性効果がある。これは、“果物”の概念において“イチジク”よりも“オレンジ”のほうが典型的であるというように、ある概念(上位概念)に属するインスタンス(下位概念)は典型性にしたがって順序付けられていることを主張している[1]。

インスタンスの典型性はさまざまな認知タスクにおけるヒトの振る舞いに影響を与えることが知られている[2]。例えば、ある概念に属するインスタンスを列挙させると典型的なインスタンスほど先に想起されること[3]、あるインスタンスがある概念に属するかどうかを判断させるとそのインスタンスが典型的であるほど判断時間が短くなること[4]、などがその例である。

CollinsとLoftus[5]は典型性効果の説明を可能とするモデル(spreading activation model)を提案している。このモデルでは、概念とインスタンスはそれぞれ単一ノードで表現され、概念とインスタンスとの関係はノード間のリンクで表現される。インスタンスの典型性を規定しているのはリンクの長さであり、インスタンスが典型的であるほどその概念と短いリンクで結び付けられる。ある概念ノードの活性はリンクを通じ

てインスタンスに伝播される。ここでリンクが長いほど活性伝播に時間がかかることを仮定すれば、典型性効果の説明が可能となる。

本研究はこれをより脳に近い神経回路モデルで実現することを目的とする。本モデルでは、概念とインスタンスは発火パターン(非一様な階層的相関パターン)で表現され、インスタンスの典型性はその概念との相関で表現される。これらを記憶した後、ある概念の入力から典型的なインスタンスを高速かつ高確率で想起するモデルを提案する。“果物”が入力されると“イチジク”よりも“オレンジ”のほうを好んで想起し、“イチジク”を想起する場合(想起してしまう場合)には長い想起時間を必要とするようなモデルである。

2. 概念とインスタンスの表現

本研究では概念とインスタンスを2階層の相関パターンにおける第1と第2レベルの N 次元パターンとして扱うことにする。そして、インスタンスが典型的であるほど、概念との相関が強いものと仮定する。したがって、パターン空間において典型的なインスタンスは非典型なものに比べて概念の近くに分布することになる。

まず、 P_1 個の概念を $\{\xi_i^\mu\}$ ($\mu=1, \dots, P_1; i=1, \dots, N$)で表す。各要素 ξ_i^μ は等確率で ± 1 の値をとるものとする。

次に、 μ 番目の概念 $\{\xi_i^\mu\}$ に属する P_2 個のインスタンスを $\{\xi_i^{\mu\nu}\}$ ($\nu=1, \dots, P_2; i=1, \dots, N$)で表す。各要素 $\xi_i^{\mu\nu}$ は次の確率で ± 1 の値をとるものとする：

$$\Pr(\xi_i^{\mu\nu} = \pm 1) = \frac{1 \pm \xi_i^\mu b_{\mu\nu}}{2}.$$

† Reproduction of the Typicality Effect by a Neural Network Storing Nonuniformly Correlated Patterns
Makoto HIRAHARA and Takashi NAGANO

*1 法政大学工学部
Hosei University

*2 玉川大学学術研究所脳科学研究施設
Tamagawa University Research Institute

このようにインスタンスを生成すると、親との相関 $E[\xi_i^\mu \xi_i^{\mu\nu}] (= N^{-1} \sum \xi_i^\mu \xi_i^{\mu\nu})$ は $b_{\mu\nu}$ となる。本研究ではこの $b_{\mu\nu}$ を 1 未満の正数に設定し、インスタンス ν の典型性として解釈することとする。

生成されたインスタンスの間には、

$$E[\xi_i^{\mu\nu} \xi_i^{\mu'\nu'}] = \begin{cases} 1, & \mu = \mu' \text{ かつ } \nu = \nu' \\ b_{\mu\nu} b_{\mu'\nu'}, & \mu = \mu' \text{ かつ } \nu \neq \nu' \\ 0, & \mu \neq \mu', \end{cases}$$

という関係が成立する。同じ概念 μ に属するインスタンス ν と ν' との相関は $b_{\mu\nu} b_{\mu\nu'}$ であり、異なった概念に属するインスタンス同士は無相関である。ある概念 μ において、典型的なインスタンスは非典型的なものに比べて、その概念や別の典型的なインスタンスと多くの特徴を共有することを仮定していることになる。

3. モデル

モデルの構造を図1に示す。本モデルは2つの連想記憶モデル (AM1とAM2) をカスケード接続したものであり (3つ以上の接続への拡張は容易)、非一様な階層的相関パターンの取り扱いが可能な従来モデル [6] に変更を加えたものである。この他にも階層的相関パターンを扱う従来モデルは数多くあるが [7,8,9,10], これらは相関 $b_{\mu\nu}$ が μ や ν に依存しないことを前提にしている。以下では、モデルの動作を、学習フェーズと想起フェーズとに分けて説明する。

図1 モデルの構造。概念 $\{\xi_i^\mu\}$ を記憶する自己想起型連想記憶モデル AM1 と、インスタンス $\{\xi_i^{\mu\nu}\}$ と差分パターン $\{\eta_i^{\mu\nu}\} (\equiv \{\xi_i^\mu \xi_i^{\mu\nu}\})$ のペアを記憶するリカレントな相互想起型連想記憶モデル AM2 から構成される。⊗はAM1で想起された概念 $\{\xi_i^{(\mu)}\}$ とAM2 (の第2層) で想起された差分パターン $\{\eta_i^{\mu\nu}\}$ とからインスタンス $\{\xi_i^{\mu\nu}\}$ を生成し、AM2 (の第1層) にフィードバックする。

3.1 学習フェーズ

第1連想記憶AM1は P_1 個の概念 $\{\xi_i^\mu\}$ を記憶する自己想起型モデルである。AM1は、要素が等確率で ± 1 の値をとるパターンを記憶する“通常の連想記憶モデル” (例えば、ホップフィールドモデル [11]) と同じであるため、記述を省略する。なお、記憶時に概念が与えられない場合にはAmariの概念形成モデル [12] などを導入すればよい [9]。

第2連想記憶AM2は2層からなるリカレントな相互想起型モデルであり、 $P_1 P_2$ 個の入出力ペア $(\{\xi_i^{\mu\nu}\}, \{\eta_i^{\mu\nu}\})$ を記憶する。ここで、 $\{\eta_i^{\mu\nu}\}$ は $\{\eta_i^{\mu\nu}\} \equiv \{\xi_i^\mu \xi_i^{\mu\nu}\}$,

で定義される差分パターンである [6,9,10]。 $\xi_i^\mu = \xi_i^{\mu\nu}$ の場合に $\eta_i^{\mu\nu} = 1$ となり、そうでない場合は $\eta_i^{\mu\nu} = -1$ となるから、差分パターンは概念の情報を含まず、インスタンス特有の情報だけを含んでいる。そのバイアス $(N^{-1} \sum_i \eta_i^{\mu\nu})$ は典型性 $b_{\mu\nu}$ に等しいので、インスタンスの典型性が高いほどその差分パターンはスパースになる。第1層の素子 j から第2層の素子 i への重みを w_{ij} で表すと、入出力ペアは

$$\begin{aligned} w_{ij} &= \sum_{\mu\nu} \beta_{\mu\nu} \eta_i^{\mu\nu} (\xi_j^{\mu\nu} - \xi_j^\mu b_{\mu\nu}) \\ &= \sum_{\mu\nu} \beta_{\mu\nu} \eta_i^{\mu\nu} (\eta_j^{\mu\nu} - b_{\mu\nu}) \xi_j^\mu, \end{aligned} \quad (1)$$

$$\beta_{\mu\nu} \equiv \frac{1}{N(1 - b_{\mu\nu}^2)}, \quad (2)$$

と記憶される。ただし、 $w_{ij} = 0$ とする。このように、第2層の素子 $i (i=1, \dots, N)$ は第1層の全ての素子 (素子 i を除く) からの入力を受けている。

3.2 想起フェーズ

階層的相関パターンを扱った上記従来研究 [6, 7, 8, 9, 10] では、インスタンスにノイズを載せたパターンをモデルに入力し、そのインスタンスが安定に想起されるかをテストしていた。これは従来研究が、“通常の連想記憶モデル”では記憶不可能な階層的相関パターンの記憶そのものを問題にしていたからである。

これに対して本研究では、概念が与えられた時にどのインスタンスが想起されるのかという一對多の問題を扱うことになるから、モデルへの入力には概念 $\{\xi_i^\mu\}$ にノイズ $\{\rho_i\}$ を載せたパターン $\{\xi_i^\mu\}$ とする (図1)：

$$\{\xi_i^\mu\} = \{\xi_i^\mu \rho_i\}. \quad (3)$$

ここで、ノイズ $\{\rho_i\}$ はバイアス c ($0 \leq c \leq 1$)を持つランダムパターンであり、以下の確率 $\Pr(\rho_i = \pm 1)$ で各要素 ρ_i に ± 1 の値を割り振ることによって生成されるものとする：

$$\Pr(\rho_i = \pm 1) = \frac{1}{2}(1 \pm c). \quad (4)$$

入力 $\{\zeta_i^\mu\}$ は $c = 1$ の場合にノイズのない概念 $\{\xi_i^\mu\}$ となり、 $c = 0$ の場合に概念と相関のない全くのランダムノイズとなる。なお、インスタンスにノイズを載せたパターンが入力された場合には、従来モデルと同様にそのインスタンスを想起することができる。

まずAM1は入力 $\{\zeta_i^\mu\}$ を受け、時刻 t_0 で概念 $\{\xi_i^\mu\}$ を想起したものとする。上述のとおり、AM1は“通常の連想記憶モデル”であるのでダイナミクスの記述を省略する。なお、この想起は、概念の数 P_1 がAM1の記憶容量($\approx 0.15N$ [11])よりも小さく、入力と概念との相関($=c$)が十分に大きければ保証される[13, 14]。以下では概念と、AM1が想起した概念(AM1の安定状態)とを区別するため、後者を $\{\xi_i^{(\mu)}\}$ で表すことにする(図1)。

次に、リカレントな相互想起型モデルAM2はAM1と同じ入力 $\{\zeta_i^\mu\}$ を受け、AM1で想起された概念 $\{\xi_i^{(\mu)}\}$ を利用して、第1層でインスタンス $\{\xi_i^{\mu\nu}\}$ を、第2層で差分パターン $\{\eta_i^{\mu\nu}\}$ を想起する(図1)。時刻 t における第1層の出力を $\{x_i(t)\}$ で表すと、AM1が概念 $\{\xi_i^{(\mu)}\}$ を想起した時刻 $t=t_0$ に

$$x_i(t_0) = \xi_i^\mu,$$

と初期化され、重み w_{ij} を介して第2層に伝播される。次時刻 $t+1$ における第2層の出力 $\{y_i(t+1)\}$ は次のように与えられる：

$$h_i(t) = \sum_j w_{ij}(x_j(t) - \xi_j^{(\mu)}\theta(t)), \quad (5)$$

$$\theta(t) = \frac{1}{N} \sum_i y_i(t),$$

$$y_i(t+1) = f(h_i(t)).$$

ここで出力関数 $f(u)$ は $u > 0$ の場合に $+1$ 、 $u < 0$ の場合に -1 であり、 $u = 0$ の場合は ± 1 の値を等確率でとる。このように第2層の内部状態 $h_i(t)$ は第1層の出力 $x_j(t)$ を $\xi_j^{(\mu)}\theta(t)$ だけシフトした値の重み付け総和で与えられるが、このシフトの生理学的妥当性を裏付ける根拠は今のところない。シフトを導入しなくても本モデルは所期の動作を実現するが、導入した

実験で確認されている(結果省略)。次に図1の⊗は、AM1で想起されている概念 $\{\xi_i^{(\mu)}\}$ を用いて第2層の出力を

$$z_i(t+1) = \xi_i^{(\mu)} y_i(t+1), \quad (6)$$

と変換し、第1層にフィードバックする：

$$x_i(t+1) = z_i(t+1). \quad (7)$$

以上のダイナミクスにより、第1層ではインスタンス $\{\xi_i^{\mu\nu}\} = \{\xi_i^{(\mu)} \eta_i^{\mu\nu}\}$ が、第2層では差分パターン $\{\eta_i^{\mu\nu}\}$ が想起される。

このフィードバックは第1層の次時刻の出力 $\{x_i(t+1)\}$ をAM1で想起された概念 $\{\xi_i^{(\mu)}\}$ に近づけるように働く。このため、AM2で安定に想起され得るのは概念 $\{\xi_i^{(\mu)}\}$ に属する P_2 個のインスタンス $\{\xi_i^{\mu\nu}\}$ と差分パターン $\{\eta_i^{\mu\nu}\}$ のペア($\nu=1, \dots, P_2$)のみとなる。このことを理解するために、AM1が例えば1番目の概念 $\{\xi_i^{(1)}\}$ を想起した場合を考える。この場合、第1層の出力 $\{x_i(t)\}$ はフィードバック(式6と7)により、

$$x_i(t) = \xi_i^{(1)} y_i(t),$$

となるから、式5は次のように書き換えられる：

$$\begin{aligned} h_i(t) &= \sum_j w_{ij}(1) \xi_j^1 \xi_j^{(1)} (y_i(t) - \theta(t)) \\ &\quad + \sum_{\mu \neq 1} \sum_j w_{ij}(\mu) \xi_j^\mu \xi_j^{(1)} (y_i(t) - \theta(t)), \\ w_{ij}(\mu) &\equiv \sum_\nu \beta_{\mu\nu} \eta_i^{\mu\nu} (\eta_i^{\mu\nu} - b_{\mu\nu}). \end{aligned} \quad (8)$$

ここで概念どうしは互いに独立であり、どの概念もバイアス($N^{-1} \sum_i \xi_i^\mu$)がゼロであるから、第2項を無視すると、

$$h_i(t) \approx \sum_j w_{ij}(1) (y_i(t) - \theta(t)), \quad (9)$$

となる。これは P_2 個のパターンを記憶した通常の自己想起型連想記憶モデルのダイナミクスと同じである。

すなわち、AM1が概念 μ を想起した時、 $P_1 P_2$ 個の入出力ペアを式1で記憶した相互想起型モデルAM2は、その概念 μ に属する P_2 個の差分パターンを式8で記憶した“概念 μ 専用の自己想起型モデル”となる。このため、AM1で想起された概念の差分パターン(とインスタンス)のみが想起可能となり、他の概念の差分パターンは想起不可能となる。

4. 数値実験

モデルに概念を入力し、AM2(の第1層)でどのインスタンスが想起されるかを確認した。“果物”の入力により何を連想するか、という問いを想定している。本数値実験での入力は、一般性を失うことなく、概念1にノイズを載せたパターン $\{\xi_i^1\}$ とした(式3)。また、“通常の連想記憶モデル”であるAM1の動作に興味はないため、ここではAM1が時刻 $t_0=0$ で概念 $\{\xi_i^{(1)}\}$ を想起したものとして、AM2の動作確認のみを行った。ここで、AM2の第1層の出力 $\{x_i(t)\}$ とインスタンス $\{\xi_i^{\mu\nu}\}$ との類似度を

$$S^{\mu\nu}(t) \equiv \beta_{\mu\nu} \sum_i (\xi_i^{\mu\nu} - \xi_i^{\mu} b_{i\nu})(x_i(t) - \xi_i^{(1)} \theta(t)), \quad (10)$$

で定義すると、 $S^{\mu\nu}(t) > 0.99$ となったとき、モデルはインスタンス $\{\xi_i^{\mu\nu}\}$ をステップ数 t で想起したものと解釈した。なお、 $S^{\mu\nu}(t)$ は後述のSN解析において、差分パターン $\{\eta_i^{\mu\nu}\}$ の想起に貢献するシグナルに相当するものである。これは式5が

$$\begin{aligned} h_i(t) &\approx \sum_{mn} S^{mn}(t) \eta_i^{mn} \\ &= S^{\mu\nu}(t) \eta_i^{\mu\nu} + R^{\mu\nu}(t), \end{aligned} \quad (11)$$

と書き換え可能であるためである。ここで、 $R^{\mu\nu}(t)$ は差分パターン $\{\eta_i^{\mu\nu}\}$ の想起を妨げるクロストークノイズである。

まず、ノイズのバイアス c (式4)に4つの条件(0.0, 0.3, 0.6, 0.9)を設け、 $N=1000$, $P_1=8$, $P_2=8$ 、インスタンス $\{\xi_i^{\mu\nu}\}$ の典型性を $b_{i\nu}=0.05(\nu-1)+0.55$ として数値実験を行った。よって、概念1に属する最も典型的なインスタンスは $\{\xi_i^{18}\}$ ($b_{18}=0.9$)となり、最も非典型的なインスタンスは $\{\xi_i^{11}\}$ ($b_{11}=0.55$)となる。各試行では、式3で生成した概念 $\{\xi_i^1\}$ を入力し、想起されたインスタンスとその想起に要したステップ数を記録した。想起に失敗した場合には“失敗”としてカウントした。以上の試行を、ノイズバイアス c の各条件につき1万回繰り返した。

実験結果を図2に示す。図2Aは概念1に属するインスタンスの想起確率を典型性 $b_{i\nu}$ ($\nu=1, \dots, 8$)の関数としてプロットしたものであり、右端は想起失敗確率である。他の概念に属するインスタンスが想起されることはなかった。ほぼ失敗なく概念1のインスタンスが想起されていること、そして、典型的なインスタンスほど想起確率が高くなっていることがわか

図2 ノイズのバイアス c を変化させた場合の実験結果：(A) 概念1に属するインスタンス $\{\xi_i^{1\nu}\}$ の想起確率と想起失敗確率；(B) その想起に要した平均ステップ数。横軸はインスタンスの典型性 $b_{i\nu}$ ($\nu=1, \dots, 8$)を示す。他の概念のインスタンスが想起されることはなかった。

る。さらに、想起確率はノイズバイアス c によって変化していないこともわかる。

図2Bは概念1に属するインスタンスの想起に要した平均ステップ数(±1SEM)である。典型性が高いインスタンスほど短いステップ数で想起されていることがわかる。この場合も、ノイズバイアス c によって変化していないことがわかる。

次に、 $c=0.9$ に固定し、各概念に属するインスタンスの数 P_2 に3つの条件(8, 16, 24)を設けて同様な数値実験を行った。他の条件については上と同様である。

実験結果を図3に示す。図3Aは概念1に属するインスタンスの想起確率と想起失敗確率である。他の概念に属するインスタンスが想起されることはなかった。 P_2 が増すにつれて、想起確率が全体的に下がっていくこと、それに伴って想起失敗確率が高くなって

図3 各概念に属するインスタンスの数 P_2 を変化させた場合の実験結果: (A) 概念1に属するインスタンスの想起確率と想起失敗確率; (B) その想起に要した平均ステップ数. 横軸は典型性 b_{1v} ($v=1, \dots, 8$)である. 他の概念のインスタンスが想起されることはなかった.

くことがわかるが, 典型的なインスタンスほど想起確率が高くなっていることには変わりがない. モデルの限界に近づくにつれて, 典型性の低いインスタンスから先に記憶が失われていくようである.

図3 Bは想起に要した平均ステップ数(± 1 SEM)である. P_2 が増すにつれて, 平均ステップ数が長くなることがわかるが, 典型性が高いインスタンスほど短いステップ数で想起されていることには変わりがない.

5. 議論

典型性が高いインスタンス(とその差分パターン)ほど高い確率かつ短いステップ数で想起されることが確認された. この動作の理解を深めるためにAM2を解析する.

式11から, シグナル $S^{\mu\nu}(t)$ (式10)が大きくなるほど, そしてクロストークノイズ $R^{\mu\nu}(t)$ が小さくなるほど,

ど, 差分パターン $\{\eta_i^{\mu\nu}\}$ が想起されやすくなることがわかる. 以下では, ある入力を与えられた場合のシグナルとノイズを評価することでAM2の振る舞いを考察する.

まず, 想起の開始時点(時刻 $t=t_0$)について考える. 数値実験と同様に入力を $\{\zeta_i^1\}$ とし, AM1が時刻 t_0 で概念1 $\{\xi_i^{(1)}\}$ を想起したものとすると, 差分パターン $\{\eta_i^{\mu\nu}\}$ (他概念の差分パターンも含む)の想起に貢献するシグナル $S^{\mu\nu}(t_0)$ は,

$$\begin{aligned} S^{\mu\nu}(t_0) &= \beta_{\mu\nu} \sum_i (\xi_i^{\mu\nu} - \xi_i^{\mu} b_{\mu\nu})(\zeta_i^1 - \xi_i^{(1)} c) \\ &= \beta_{\mu\nu} \sum_i \xi_i^{\mu} (\eta_i^{\mu\nu} - b_{\mu\nu}) \xi_i^{(1)} (\rho_i - c), \end{aligned}$$

と書ける. ここで, 概念 $\{\xi_i^{\mu}\}$, 差分パターン $\{\eta_i^{\mu\nu}\}$, AM1が想起した概念 $\{\xi_i^{(1)}\}$, そして入力ノイズ $\{\rho_i\}$ は互いに独立であり, それぞれの平均(バイアス)は $E[\xi_i^{\mu}] = 0$, $E[\eta_i^{\mu\nu}] = b_{\mu\nu}$, $E[\xi_i^{(1)}] = 0$, $E[\rho_i] = c$ であるから, シグナルの平均 $E[S^{\mu\nu}(t_0)]$ はゼロとなる. したがって, その分散 $V[S^{\mu\nu}(t_0)]$ は $S^{\mu\nu}(t_0)$ の自乗の平均で与えられるから,

$$\begin{aligned} V[S^{\mu\nu}(t_0)] &= E \left[\beta_{\mu\nu}^2 \sum_i \xi_i^{\mu} (\eta_i^{\mu\nu} - b_{\mu\nu}) \xi_i^{(1)} (\rho_i - c) \right. \\ &\quad \times \left. \sum_j \xi_j^{\mu} (\eta_j^{\mu\nu} - b_{\mu\nu}) \xi_j^{(1)} (\rho_j - c) \right], \end{aligned}$$

となる. ここで, $i \neq j$ の場合には上と同様な理由で値が消え, $i = j$ の場合にのみ値を持つから,

$$\begin{aligned} V[S^{\mu\nu}(t_0)] &= E \left[\beta_{\mu\nu}^2 \sum_i (\eta_i^{\mu\nu} - b_{\mu\nu})^2 (\rho_i - c)^2 \right] \\ &= (1 - c^2) \beta_{\mu\nu}, \end{aligned}$$

と見積もられる.

シグナルの平均はどの差分パターンに対しても同じゼロであるが, その分散は典型性 $b_{\mu\nu}$ (式2の $\beta_{\mu\nu}$ 参照)が高いほど大きくなることがわかる. 言い換えれば, 典型性 $b_{\mu\nu}$ が高いほど, 大きなシグナル $S^{\mu\nu}(t_0)$ を生み出す確率が高いことになる. このことは, 典型的な差分パターン(典型的なインスタンスと対になる差分パターン)が非典型的な差分パターンよりも想起されやすいことを示唆する.

クロストークノイズ $R^{\mu\nu}(t_0)$ (式11)については

$$E[R^{\mu\nu}(t_0)] = 0,$$

$$V[R^{\mu\nu}(t_0)] = (1 - c^2) \left(\sum_{mn} \beta_{mn} - \beta_{\mu\nu} \right),$$

と見積もられる．典型性 $b_{\mu\nu}$ が高いほど分散が小さくなることがわかる．言い換えれば，典型性 $b_{\mu\nu}$ が高いほど，大きなクロストークノイズを受ける確率が低くなる．このことも，典型的な差分パターンが非典型的な差分パターンよりも想起されやすいことを示唆する．

以上から，AM1が時刻 t_0 で概念1 $\{\xi_i^{(1)}\}$ を想起した時点では，各概念に属する最も典型的な差分パターン $\{\eta_i^{\mu P_2}\}$ が最も想起されやすくなっていることがわかる．典型性 $b_{\mu\nu}$ が高い差分パターンは，どの概念に属そうとも，想起されやすいことになる．

にもかかわらず，数値実験でAM2が想起したのは，AM1が想起した概念1のインスタンス(第2層では差分パターン)だけであった．これは，上述のとおり，図1の⊗を介したフィードバック(式6と7)によって，AM2が概念1専用の自己想起型モデルになるためである(式9)．このため，他の概念に属するインスタンスの想起は不可能となる．

次に想起の終了時点，すなわち，第1層が概念1のインスタンス ν を想起した場合 $\{x_i(t)\} = \{\xi_i^{1\nu}\}$ について考える．この場合のシグナル $S^{1\nu}(t)$ は ν によらず1となる．クロストークノイズについては

$$E[R^{1\nu}(t)] = 0,$$

$$V[R^{1\nu}(t)] = (1 - b_{1\nu}^2) \left(\sum_{mn} \beta_{mn} - \beta_{1\nu} \right),$$

と見積もられる．この場合も，典型性 $b_{1\nu}$ が高いほど分散が小さくなることがわかる．インスタンスが典型的であるほど，大きなクロストークノイズを受ける確率が低く，安定に想起されることがわかる．なお，分散は $P_1 P_2$ が増した場合にも大きくなる．このことは，数値実験で P_2 が大きい場合に非典型的なインスタンスがほとんど想起されなかったことに一致する(図3A)．

最後に想起の開始と終了の中間時点での議論が残されている．開始時でも終了時でもそうであったように，中間時点でも典型性 $b_{1\nu}$ が高いほどクロストークノイズの分散が小さくなっていることが予想される．そのため，非典型的なインスタンスに向かう状態遷移は，典型的なインスタンスに向かう状態遷移に比べて大きなクロストークノイズに妨げられる確率が高いのだろう．非典型的なインスタンスの想起確率が低く，想起ステップ数が長かったのはこのためであると考えられる．さらに，クロストークノイズの分散は $P_1 P_2$ が増した場合にも大きくなるだろう． P_2 が大きい場合の想起ステップ数がどのインスタンスに対しても上昇したのはこのためであろう(図3B)．これらの予想を

裏付けるには上と同様な解析が必要であることは言うまでもないが，この解析は今までのような単純な方法では済まない．なぜならば，第1層や第2層の出力は記憶した全てのパターンと相関を持っているはずだからである[13, 14]．本モデルにおいては，インスタンスの相関(典型性) $b_{\mu\nu}$ が一樣でないことやダイナミクスに動的閾値 $\theta(t)$ (式5)を導入していることなどから，その解析は困難であり今後の課題としたい．

この他にも数多くの問題点が残されている．例えば，数値実験の $c=0$ の場合，入力は何のインスタンスに対しても無相関となるが，相関のある場合と全く同等に動作することが確認された(図2)．初期(時刻 t_0)のシグナル $S^{\mu\nu}(t_0)$ (初期オーバーラップ)がどのインスタンスに対してもほぼゼロであったにも関わらず，本モデルはほとんどの場合で，AM1で想起された概念に属する何らかのインスタンスを想起した．想起に必要な最低の初期シグナル(臨界オーバーラップ)がほぼゼロであったことは，これまでの連想記憶モデルに関する常識[13, 14]に反している．

6. まとめと今後の課題

典型性効果の神経回路レベルでの理解に向けて，ある概念を入力すると，その概念に属する典型的なインスタンスを高確率かつ高速に想起するモデルを提案し，数値実験によって動作確認を行った．“果物”が入力されると，“イチジク”に比べ，“リンゴ”を高確率で想起する，というようなモデルである．

今後の課題として，“果物”の入力から，まず“リンゴ”や“バナナ”を想起し，後になって“イチジク”や“ザクロ”を想起するようなモデルに展開することがあげられる．インスタンスを典型性の順序で次々と想起していくことは，ヒトの振る舞いにもよく見られることである[3]．

しかし，ヒトの振る舞いはこのような“かたい想起”に留まらない．例えば，“火事のときに持ち出すもの”という概念を予め持っているヒトは少ないだろう．その場で概念を生成し，“子供”，“預金通帳”，“家の権利証”，“絵画”，“アルバム”など，普段は全く別の概念に属しているようなインスタンスを次々に想起することもある[15]．このような想起は本モデルの枠組みを超えている．かたい概念階層で縛られたボトムアップ的な想起システムだけを考えている限りでは，ヒトの柔軟な想起を説明しきれない．これは今後の大きな課題である．

謝辞

有益な議論をして頂いた玉川大学教授の塚田稔先生

と元教授の斎藤秀昭先生に感謝致します。また、本論文に対して種々のご指摘をくださいました査読者の方々に感謝致します。本研究の一部は日本学術振興会科学研究費補助金奨励研究A(13780304)の援助を受けて行われました。

参考文献

- [1] Rosch, E. (1973) On the internal structure of perceptual and semantic categories. In T.E. Moore (ed.) Cognitive development and the acquisition of language, New York: Academic Press.
- [2] Roth, I., and Bruce, V. (1995) Perception and representation (Second Edition). Open University Press.
- [3] Rosch, E. (1978) Principles of categorization. in E. Rosch and B.B. Lloyd (eds.) Cognition and Categorization, Hillsdale, NJ: Erlbaum.
- [4] Smith, E.E., Shoben, E.J., and Rips, L.J. (1974) Structure and process in semantic memory: A featured model for semantic decisions. Psychological Review, 81, 214-241.
- [5] Collins, A.M., and Loftus, E.F. (1975) A spreading activation theory of semantic processing. Psychological Review, 82, 407-428.
- [6] Hirahara, M., Oka, N., and Kindo, T. (2000a) Cascade associative memory storing hierarchically correlated patterns with various correlations. Neural Networks, 13, 51-61.
- [7] Gutfreund, H. (1988). Neural networks with hierarchically correlated patterns. Physical Review A, 37, 570-577.
- [8] Morita, M. (1993) Associative memory with nonmonotone dynamics. Neural Networks, 6, 115-126.
- [9] Hirahara, M., Oka, N., and Kindo, T. (1997) Associative memory with a sparse encoding mechanism for storing correlated patterns. Neural Networks, 10, 1626-1637.
- [10] Hirahara, M., Oka, N., and Kindo, T. (2000b) A cascade associative memory model with a hierarchical memory structure. Neural Networks, 13, 41-50.
- [11] Hopfield, J.J. (1982) Neural networks and physical systems with emergent collective computational abilities. Proceedings of the National Academy of Sciences USA, 79, 2554-2558.
- [12] Amari, S. (1977) Neural theory of association and concept-formation. Biological Cybernetics, 26, 175-185.
- [13] Amari, S., and Maginu, K. (1988) Statistical neurodynamics of associative memory. Neural Networks, 1, 63-73.
- [14] Okada, M. (1995) A hierarchy of macrodynamical equations for associative memory. Neural Networks, 8, 833-838.
- [15] Barsalou, L.W. (1983) Ad hoc categories. Memory & Cognition, 11, 211-227.

(2005年9月31日 受付)

(2006年1月8日 採録)

【問い合わせ先】

〒184-8585 東京都小金井市梶野町3-7-2

法政大学工学部経営工学科

平原 誠

TEL: 042-387-6367

FAX: 042-387-6367

E-mail: hirahara@k.hosei.ac.jp

著者紹介

ひらはら まこと
平原 誠 [正会員]

1992年法政大学大学院修士課程修了。同年松下技研(株)勤務(松下電器産業(株)からの出向)。2000年法政大学助手。ニューラルネット、視覚心理、最適化などの研究に従事。博士(工学)。電子情報通信学会、日本神経回路学会等会員。

ながの たかし
永野 俊 [非会員]

1968年早稲田大学大学院博士課程修了。工学博士。1967~1969年早稲田大学助手。1969~1987年通産省(現経産省)電気試験所(現産総研)研究員。1987~2005年法政大学教授。2005~現在玉川大学学術研究所特別研究員。この間、米国、カーネギー・メロン大学客員研究員、オーストラリア、ウィーン工科大学客員教授歴任。ニューラルネット、視覚心理などの研究に従事。電子情報通信学会(フェロー)、日本神経回路学会等会員。

Reproduction of the typicality effect by a neural network storing nonuniformly correlated patterns

by

Makoto HIRAHARA and Takashi NAGANO

Abstract :

The typicality effect is one of important findings on concept hierarchies in the field of cognitive psychology. The followings are some of well-known phenomena: if asked to list up all the instances of a concept, we recall them in descending order of their typicality; if asked whether an instance belongs to a concept, we can judge faster as its typicality is higher. As the first step toward a neuronal basis of the typicality effect, this paper proposes a neural network model which recalls a typical instance of a concept with a higher probability and also in a shorter period of recalling time than an atypical one.

Keywords : neural network, typicality effect, concept hierarchy, nonuniformly correlated patterns

Contact Address : **Makoto HIRAHARA**

Hosei University

3-7-2 Kajinchotyo, Koganei-shi, Tokyo, 184-8584 JAPAN

TEL : 042-387-6367

FAX : 042-387-6367

E-mail : hirahara@k.hosei.ac.jp