

忘れる…その大切さ

～脳の数理モデルの幾何的研究～

金道敏樹

■研究のねらい

覚えていたはずのことを忘れることは悔しく、また嫌なことなので、私たちは「忘れる」ことはよくないことだと思いがちです。でも、脳という私たちの心と知を支える情報処理機械にとって、「忘れる」ことはどうなのでしょう。

やはり、よくないことなのでしょう。それとも、何かもっと別の大切な役割があるのでしょうか。

この研究のねらいは、脳の数理モデルに対してこのような問いに答えられるような方法を提示することにあります。もちろん私の興味はここで紹介する「忘れる」ことにだけにあるのではなく、もう少し一般的に、

- ① 外界からの刺激が与えられたとき、脳はそれをどのような表現で記憶し、どのようなプロセスで記憶を呼び覚ますか？
- ② 逆に、脳の情報処理のプロセスに注目したとき、それに対応する情報表現はどのようなものか？
- ③ そして、どういう方法であれば、こうした問題を取り扱えるか？

というものです。表題の「忘れる…」は、その中の一つのテーマです。

さて、人を初めとする生体の認知や行動のメカニズムを探求する研究に、脳が行なっている情報処理の様式／原理の解明があります。この研究のいくつかあるアプローチの一つが、脳の活動（機能）の本質を捉えている（と思われる）脳のモデルを作り、モデルが内包する脳の性質を微視的、巨視的の両面から解明する理工学的なアプローチです。

私の研究は、この理工学的なアプローチにあって、非線型性の強い脳の数理モデルを幾何的に捉えようとする点に特徴があります。具体的には、幾何的な解析方法を与え、その方法に基づいて、脳に見出された現象が数理モデルのどんなメカニズムとプロセスに対応するのかを明らかにしようとするものです。

■研究成果

今回取り上げた問題の一つが、「忘れる」ことであり、それは「脳は時間的に変化する刺激をどのようにして記憶するか」に関係する研究です。この研究では、過去を忘れながら刺激を記憶することで、脳は時間的に離れた刺激であっても関係づけることができることを示し、さらに、時間的に変化する刺激の中にノイズが入ってきたときでも、このノイズ刺激を除いて時間的に変化する刺激を思い出せるという一種の選択的な時系列刺激の記憶が可能であることを明らかにします。

1. 外からの刺激をどうやって記憶するか？

記憶は、脳の中の神経細胞と神経細胞をつなぐシナプスの結合の強さとして表現されている、と考えられています。そして、シナプスは、人の声など外部から刺激に応じて興奮した神経細胞のパターンにしたがって、その結合の強さを変化させます。こ

のシナプスの結合の変化の規則としてよく知られているものに、2つの神経細胞が同時に興奮したとき、その2つの神経細胞をつなぐシナプスの結合は強化されるというヘップの学習則があります。ヘップの学習則の言うことは、神経細胞 A と神経細胞シナプス B とが S_{AB} でつながっているけど、神経細胞 C はそれらとつながっていないとき、第 3 の神経細胞 C が興奮したかどうかはシナプス S_{AB} の変化に影響を与えない（空間的局所性）。また、神経細胞 A が興奮して随分時間が経ってか

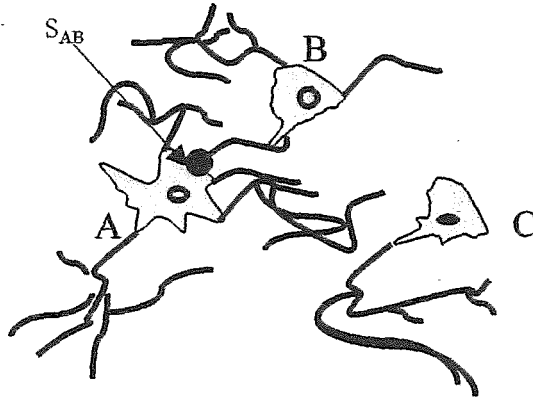


図 1. シナプス結合 S_{AB} でつながれた 2 つの神経細胞 A、B と、独立な神経細胞 C。

時間遅れの
大きい経路



図 2. 時間遅れの大きい経路と小さい経路を別々に通った刺激が脳のある場所で合流する場合の概念図

ら神経細胞 B が興奮しても、シナプス S_{AB} は変化しない（時間的局所性）ということです（図 1 参照）。俗な言い方をすれば、この規則は、よそで誰が何をしようとか関係ない、タイミングを外すと効果がない、ということですからとても自然な考えです。

しかし、シナプスの変化に時間的局所性あるとすれば、どうやって脳は時間とともに変化する刺激を記憶できるのでしょうか。

2. 脳は、時間とともに変化する刺激をどのように記憶するか？

ふつうの考えは、刺激が引き起こした神経細胞の興奮が、脳の中で時間遅れの大きい経路と時間遅れの小さい経路を別々に通って、脳のある場所で合流することによって、2つの異なる時刻の刺激の時間的な関係がシナプスに記憶されるというものです（図 2 参照）。以下、これを「時間遅れ学習」と呼ぶことにします。

時間遅れ学習とは、脳は時刻 t の刺激が入力されたとき時刻 $(t+\tau)$ の刺激を思い出せるように記憶する方法です。だから、外部から $\blacksquare \rightarrow \blacklozenge \rightarrow \blacktriangle$ という刺激が来たとき、この学習則によって記憶されるのは、 \blacksquare の次は \blacklozenge ($\blacksquare \rightarrow \blacklozenge$)、 \blacklozenge の次は \blacktriangle ($\blacklozenge \rightarrow \blacktriangle$) という時間 τ だけ離れて入ってきた 2 つの刺激の組になります。いくつもの刺激が連続する時系列刺激は、この 2 つの刺激の組に分解されて記憶されるのです。

この方法は、時間的な関係を記憶する方法として、一見とてももったもんな方法です。しかし、この時間遅れ学習にはいくつかの不自然な性質があります。その一つが、途中にノイズ刺激が入ってきた時におきる現象です。例えば、外部から $\blacksquare \rightarrow \blacklozenge \rightarrow \blacktriangle$ であるはずのものが、 $\blacksquare \rightarrow \circ \rightarrow \blacktriangle$ という具合に途中に本来の刺激とは全く無関係の刺激（ノイズ刺激） \circ が割り込んでしまったとき、時間遅れ学習はそのノイズ刺激をも意

味のあるものとして記憶してしまいます。しかも、ノイズ刺激○は刺激■の次に来るという順番も正確に記憶する。ノイズ刺激自体を記憶することは変ではないけれども、「ノイズ刺激○は刺激■の次に来る」という順番も正確に記憶するというのは、少々奇妙な話です。

私たちは、むしろ「時系列刺激■→?→▲の途中に、ノイズ刺激○が入ってきた」と記憶しているのではないのでしょうか。この疑問が、私たちを次のような研究テーマに導きます：

- ①脳の中には、時系列刺激の中からノイズ刺激を選び分けることができるような、もっと適切な記憶方法があるのではないだろうか。
- ②しかも、時間遅れメカニズムを使うことなく、場所局所性と時間局所性を満たしているような記憶法はないのだろうか。

この問題を、古典的な脳の数理モデルの一つ、連想記憶を使って考えてみましょう。

3. 「忘れる」連想記憶

3.1. 連想記憶の概略

まず、先ほどシナプスの変化に関するとても自然な仮定（場所局所性と時間局所性）を満たしているような記憶法を用いている連想記憶について説明をします。

まず、話をできるだけ簡単にするために「忘れる」ことがない場合について考えましょう。

初めの方で話したように、脳を形作っている神経細胞には、興奮している状態と興奮していない状態があります。そして、神経細胞は、入力される刺激が一定の値を越えた場合に興奮状態になります。これに対応して、連想記憶では、神経細胞を、シナプス結合を介した入力の総和が一定の値を越えたとき興奮状態に対応する1を出力し、その他の場合には非興奮状態に対応する-1出力する非線型素子としてモデル化します。

連想記憶とは、こうした非線型素子を相互に結合したものです。そして、連想記憶のシナプス結合は、外部からの刺激■で連想記憶の神経細胞の状態が図3のような興奮パターン（薄い灰色が興奮している、黒が興奮していない神経細胞を示す）となったとき、太線で示したシナプスが強化され、細線で示したシナプスが弱化されます。これは、先に触れたヘップの学習則を修正したもので、コバリアンス学習則と呼ばれる標準的なシナプスの変化の規則です。この規則は、時間についても（ある時刻の神経細胞の興奮パターンだけから決まる）、場所についても（直接つながっている2つの神経細胞の状態だけから決まる）、局所的にシナプスを変化させる規則です。

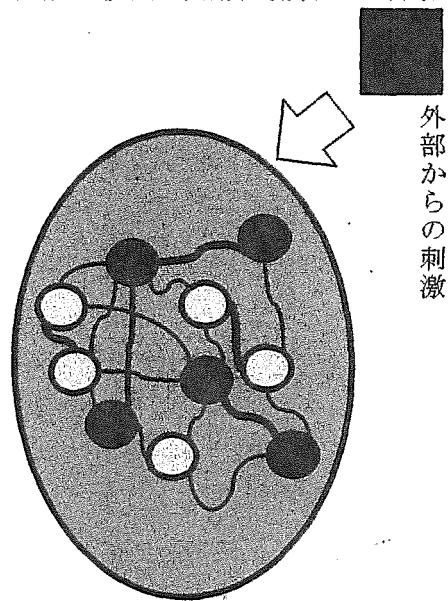


図 3. 連想記憶の構造と学習アルゴリズムをの概略を支援す概念図

3.2. 連想記憶の幾何的性質

こうした連想記憶の性質を幾何的視点から眺めてみましょう。

まず、状態が±1で表現される神経細胞の集団である連想記憶の状態は、各成分が±1のベクトル(状態ベクトル)で表現できます。すると、この状態ベクトルの長さは、どんな場合でも一定値であることに気づきます。つまり、連想記憶の状態は球面上の一点として表現できるということです。

さて、連想記憶の興奮パターンの時間的な変化—(連想記憶が行う情報処理プロセス)—は、状態ベクトルの成分が1から-1(もしくはその逆)に変化するというので、状態ベクトルの長さを変えるような変化ではありません。ということは、連想記憶の興奮パターンの変化というものは、この球面上を点から点へと移り変わ

って行く変化だということになります。これをもう少し詳しく見ると、この変化は球面上を流れる川のように描くことができることが分かります(図4参照)。

さらに、■→◆→▲という外部からの刺激をコバリアンス学習則で記憶させると、図4に示したように、いずれも川の流れの同じような場所に記憶されることが分かるのです。これは、時間について局所的な記憶方法であるコバリアンス学習則では、■→◆→▲という時系列であっても、順序を入れ替えた◆→■→▲という時系列であっても全く同じに記憶されるということです。

この結果から「時間遅れを入れないと、時間について局所的な記憶方法では時間的な変化を記憶することは全くできない」ということになるのかと言えば、そうでもありません。この幾何的描像は、もし刺激■→◆→▲を川の流れ—(プロセスの流れ)—に沿って配置できれば、■の次は◆(■→◆)という時間的にずれた2つの刺激の組を記憶しなくても、時系列を記憶することも、思い出すこともできることを暗示しています。

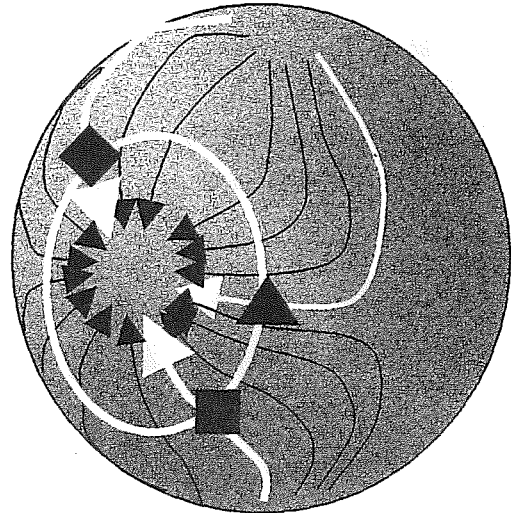
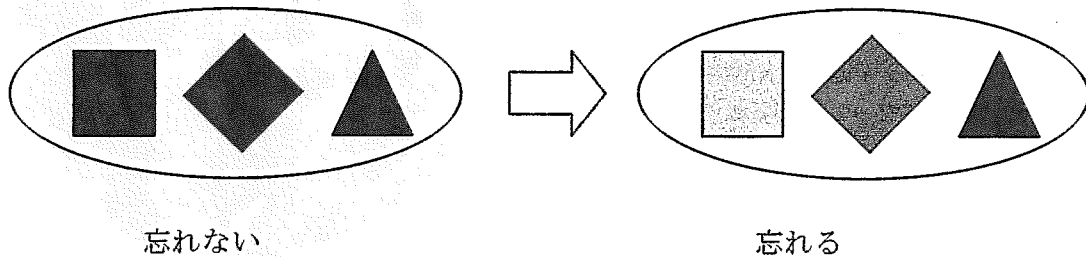


図4. 時間的にも場所的にも局所的なシナプス結合の変化規則で刺激を記憶した連想記憶の記憶の構造と興奮パターンの変化の様子を示す概念図

3.3. 「忘れる」こと

このような幾何的な視点に立てば、「時間遅れメカニズムを使うことなく、時間的にも場所的にも局所的なシナプス結合の変化の規則で、時間的に変化する刺激を記憶することはできるか」という一見無理難題とも思える問題は、「時間的にも場所的にも局所的なシナプス結合の変化の規則で、刺激の記憶の流れに沿って入ってきた順に配置できるか」という簡単な問題に置き換えることができるわけです。これに対する答

えは、「先に入っていた刺激から忘れるようにすればよい」という極めて単純なものです。シナプスレベルで「忘れる」ことは、神経細胞の興奮などに関係なく「シナプスの結合の強さは時間とともにだんだん弱くなる」という局所的な変化に当たります。このように「忘れる」ことを考えると、今までは同じ強さで覚えていたためにどの順番で入ってきたかが分からなかった3つの刺激（左下図）も、古い記憶ほど記憶が薄れていけばどの順番で入ってきたが一目瞭然になります（右下図）。



これを先ほどの球面を使って描くと、3つの刺激は連想記憶の情報処理プロセスを表現した流れの中に、図5のように並ぶのです。

この結果、最初の刺激■が入ってくると、「忘れる」ことを取り入れた連想記憶は、まず忘れかかっていた刺激■を思い出し、あとは流れに沿って、■→◆→▲という時系列を思い出すことができます。

この「忘れる」の効果を観るために行った実験の一つが、結果を図6に示した、次のようなものです：2つの無関係の時間的に変化する刺激

①：A→B→C→D→E→…、(図中の番号 0～19)

②：■→◆→▲→★→●→…、(図中の番号20～39)

を考える。そして、この2つの刺激がA→■→B→◆→C→▲→D→★→…という具合に交互にやってくる場合に、「忘れる」ことを取り入れた連想記憶がどのように記憶を形成するかを調べる。具体的には、まず、連想記憶に、混在した系列を、時間的にも場所的にも局所的なシナプス結合の変化則、しかし「忘れる」メカニズムを付け加え

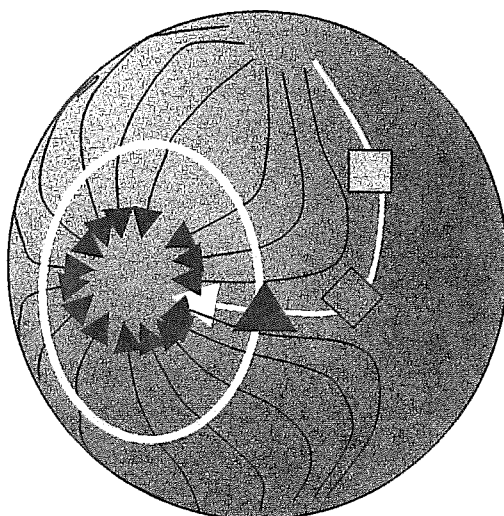


図5. 「忘れる」を取り入れた連想記憶の記憶の構造と興奮パターンの変化の様子を示す概念図

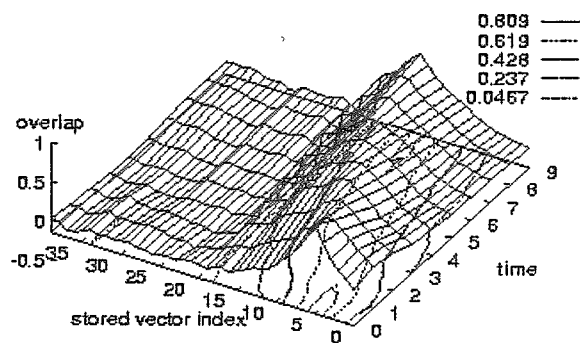


図6. 連想記憶がどのように時系列を思い出すかを示す図。Overlap軸の値は、時刻timeに何番めの刺激がどの位強く思い出されているかを示している

た方法で記憶させる。次に系列①の刺激の一つ（例えば、C）を入力した場合に、連想記憶がどういう動作をするかを検証する。

図6における overlap 軸は、時刻 time に何番めの刺激がどの位強く思い出されているかを示しています。この図から、系列①の刺激の一つを入力した場合、もともとは2つの系列が混在していた系列から、A→B→C→D→E→…という系列①だけが選ばれて思い出されていることが読み取れます。

この結果は、

- ① 忘却を取り入れることで、場所局所性と時間局所性を満たしているような記憶法であっても時間的に変化する刺激を記憶できる
- ② しかもこの記憶法は、時系列刺激の中から一つの系列を選択的に取出したり、ノイズ刺激を選り分けることができる

ことを示しています。これら性質は、時間遅れ学習則と比較して現実の脳の性質に近いと、私は考えています。

さて、ここで最初の疑問に戻りましょう：脳にとって、「忘れる」ことは、よくないことなのでしょうか。それとも、何かもっと別の大切な役割があるのでしょうか。

上の結果を得た私の答えは、「忘れる」ことは大切であるというものです。「忘れる」ことは、私たちの脳が外から入ってくる刺激（情報）を正しく関係づけて整理する上で、大切な役割を果たしているのです。

■今後の展望

脳の数理モデルを解析する幾何的手法は、よく用いられる統計的手法と比較して、定量的な評価を行とする場合精度に劣る面あるものの、脳の数理モデルの隠された可能性を明らかにすることができます。この性質を利用して、今後は、記憶した内容が段々変わって行く、間違っ覚えてしまうといった現象のメカニズムとプロセスを解き明かしたいと思っています。多分、そこには、「忘れる」ことのように何か大切な役割が隠されているのではないかと、私には思えてなりません。

■成果リスト

論文

- ・ 目木義久、金道敏樹、黒川弘章、笹瀬巖（1998）：教師なしクラスタリングによって未知のデータの階層的なクラスタ構造を推定する競合モデル 電子情報通信学会論文誌 D-II、J81 巻、2200-2210.
- ・ T.Kindo & H.akeya: (1998) A geometrical analysis of associative memory, Neural Networks, 11,39-51, (1998).
- ・ M.Hirahara, N.Oka, & T.Kindo (1997) Associative memory with a sparse encoding Mechanism for Storing Correlated patterns, NeuralNetworks,10, 1627-1636.
- ・ T.Kindo (1997) A geometrical analysis of associative memory dynamics, Computers and Systems in Japan, 28, 68-81.
- ・ H.akeya & T.Kindo (1997) Eigenspace separation of autocorrelation memory matrices for capacity expansion, Neural Networks, 10, 833-843.

- H.Kekeya & T.Kindo (1996) Hierarchical Concept Formation in Associative Memory Composed of Neuro-Window Elements, Neural Networks, 9.1095-1098

口頭発表

- T.Kindo & Y.Meki (1998) Color palette organized from image with competitive neural network, Proceeding of EANN-98, 366-373.
- N.Oka & T.Kindo (1997) Inserting, Refining and Extracting Symbolic Information into, by and from a Neural Network of Normalized Local Basis Functions, Collection of Poster Abstracts of the Fifteenth International Joint Conference on Artificial Intelligence (IJCAI-97), p. 80.
- T.Kindo, Y.Meki, R.Ito & T.Shida (1997) An Artificial Neural Network that Organizes Prototypes and Hierarchical Representation of External Stimuli, ICONIP-97
- Y.Meki, T.Kindo, H.Kurokawa, & I.Sasase (1997) Competitive model to classify unknown data into Hierarchical clusters through unsupervised learning, Proceedings of PACRIM-97,815-818.
- M. Hirahara, N. Oka & T.Kindo (1996) A self-organizing information server, Proceedings of 6st International Conference on Human-Computer Interaction.
- 掛谷英紀、金道敏樹、岡田真人、岡部洋一 (1996) 連続時間型ニューロウィンドウ法の解析、電子通信学会ニューロコンピューティング研究会、3月
- 金道敏樹、掛谷英紀(1996) 忘却を行う自己相関連想記憶の幾何学的性質、電子通信学会ニューロコンピューティング研究会、3月
- 金道敏樹、岡夏樹(1995) 連想記憶の幾何学的理論による心理現象の説明の試み、認知科学会