

単純再帰型ニューラルネットワークの心理学モデルとしての応用可能性

Psychological applicability of simple recurrent neural networks

浅川 伸一
Shin-ichi Asakawa

要旨

単純再帰型ニューラルネットワークを取り上げ、その心理モデルとしての可能性を検討した。SRN はその強力な処理能力から、複数の心理現象を説明するモデルとして有効である。説明できる心理現象としては、短期記憶、反応時間、選択的注意、プライミング、高次判別分析、連想記憶などである。本稿では、これらの心理モデルの実現方法を議論した。全てのモデルは文脈層から中間層への結合係数行列の入力信号によって定まる中間層の状態遷移として記述される。

キーワード: 単純再帰型ニューラルネットワーク, 短期記憶, 反応時間, 選択的注意, 間接プライミング, 連想記憶

abstract

The purpose in initiating the research in this article was to explore the mathematical properties in the hidden layer in simple recurrent network (SRN) proposed by Elman1990. By means of the powerful ability of this network, it is possible to describe several psychological phenomena within the framework of these kind of networks such as short term memory, reaction time, selective attention, indirect priming, and associative memory. In this article, the author discussed the implimentation of these psychological phenomena. All the psyochogolical models described here can be considered as the state transiton of the units in the hidden layer.

key words: simple recurrent network, short term memory, reaction time, selective attention, indirect priming, associative memory

一般に任意の有限状態機械 (finite state automata, 以下 FSA) はニューラルネットワークで実現可能なことが証明されている (吉則, 1993, を参照)。FSA が有限のテープ長をもつチューリングマシンと同等なことから、ニューラルネッ

トワークは現在のコンピュータの処理能力を完全に表現可能である。このことは、人間の hochi 認知機能をコンピュータ上に実現しようとする認知科学、認知心理学のモデル論的試みのすべては、ニューラルネットワークとして表現可能であることを意味する。本稿ではこのようなニューラルネットワークモデルのうち単純再帰型ネットワーク (simple recurrent networks, 以下 SRN) を取り上げ、SRN の結合係数行列を適切に設定することによって SRN の動作が理解可能なことを示し、さらにその心理学的応用の可能性について検討する。

Elman 型の SRN Elman (1990) は、入力信号を処理する入力ユニットと、直前の中間層の状態を保持する文脈ユニットを持つ 3 層のネットワークである。図 1 に Elman1990 の考案したニューラルネットワークの概念図を示した。

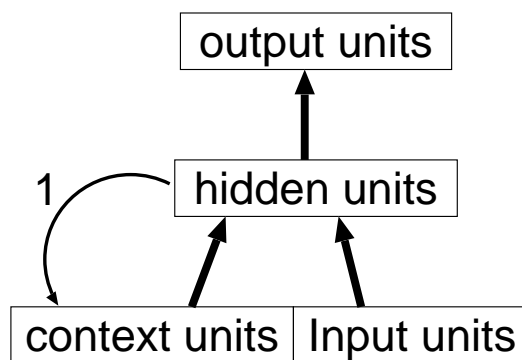


図 1 Elman の単純再帰型ニューラルネットワークの概略図

SRN とは 3 層のフィードフォワード型ニューラルネットワークに再帰結合を加えたニューラルネットワークである。逆方向への信号を再帰信号 (recurrent feedback) と呼ぶ。再帰信号によって、Elman 型の SRN では一時刻前の中間層の状態を保存することができる。このような再帰信号を受け取るユニットを文脈層あるいは関連層と呼ぶ。文脈層ユニットは一時刻前の中間層の状態をコピーするだけなので、中間層ユニットから文脈層ユニットへの結合強度は常に 1.0 に固定される。結合強度の変化学習は順方向の結合についてだけ行われるので、通常のフィードフォー

ド型ニューラルネットワークで使われる任意の誤差修正則が適用可能である。

SRN の処理能力は極めて高い。SRN の記号処理能力については Adamson と Damper によるレビューがある Adamson and Damper (1999)。良く知られているとおり 3 層のパーセプトロンは 2 値論理関数 (AND, OR, XOR 等) を実現可能である Minsky and Papert (1988)。ところが SRN はこのような単純な 2 値論理関数を越えた汎用的なニューラルネットワークなのである。SRN における文脈ユニットをチューリングマシンにおけるメモリと同一視すれば SRN は FSA と同じような構造を持っていることが想像できる。実際 SRN は FSA 以上の能力を持っていることが示唆されている Cleeremans, Servan-Schreiber, and McClelland (1995)。もし無限精度の計算を許すなら、SRN はチューリングマシンの計算能力を有する Siegelmann (1999)。しかし、計算誤差や雑音がある場合には FSA またはそれ以下の計算能力になることが明らかとなっている Maass and Orponen (1998)。

一般に任意の文字列から背後の文法を類推する問題を文法推論と呼ぶがこれは NP hard 問題として知られている。Cleeremans, Servan-Schreiber, and McClelland 1995 の研究では SRN が文法推論を行なうことができることが示された。さらに、SRN では push, pop 関数を実装可能であり、従ってスタック構造を獲得できることが数値実験から明らかにされている Rodrigues, Wiles, and Elman (1999)。

なお、本稿で考察する幾つかの心理モデルは SRN で記述を試みているが、全く同様の議論が一般的な再帰型ニューラルネットワーク (recurrent neural networks; RNN) にもそのまま当てはまる議論である。SRN の学習では単純な勾配降下法がそのまま適用できるのに対し、RNN における学習では、Back Propagation Through Time (BPTT) や Real Time Recurrent Learning (RTRL) などの誤差逆伝播法を時間展開させた学習方法が採用されることになる Williams and Zipser (1989, 1995)。

SRN の応用としては Elman らが示した言語理解、単語予測、および Cleeremans らの示した文法推論などが知られている。しかし上述のとおり SRN は強力な汎用能力を有しているので、文章理解のモデルにとどまらず、さまざまな心理学モデルを説明できる可能性がある。すなわち SRN には記号処理系としてのモデル (たとえば記号的コネクショニストモデル Holyoak and Hummel (2000) など) をその一部として含む、より一般的な能力があることを主張するものである。本稿では今までに提案されていなかった言語理解以外の心理学モデルの応用可能性を議論する。

以下では、まず限定的な条件化で厳密解が存在すること

を示し、次に、SRN の性質から短期記憶のモデル、反応時間のモデル、選択的注意のモデル、間接プライミングのモデル、連想記憶のモデル、と見なすことができることを示す。

1 定式化と厳密解

1.1 離散状態

中間層の出力信号をベクトル \mathbf{y} で表し、文脈層から中間層へ結合係数行列を \mathbf{W} とする。前述のとおり文脈層は一時刻前の中間層の状態をそのままコピーするものであるため、文脈層と中間層のユニット数は同数である。そこで、文脈層から中間層への結合係数に全結合を仮定すれば \mathbf{W} は正方行列となる。時刻 t における入力層からの信号をベクトル $\mathbf{x}(t)$ と表すことにすれば

$$\mathbf{y}(t) = f(\mathbf{W}\mathbf{y}(t-1) + \mathbf{x}(t)), \quad (1)$$

と表現できる。ここで $f(\cdot)$ は行列の各要素に非線形変換を作用させる関数であるとする。一般には $f(\cdot)$ は $f(x) = 1/(1 + \exp\{-x\})$ などのシグモイド関数が用いられる。

シグモイド関数は非線形ではあるが、0 近傍で考えれば線形近似できるので、SRN の中間層の状態は $\mathbf{W}\mathbf{y}(t-1) + \mathbf{x}(t)$ によって定まると考えることができる。ある時刻 t までの入力 $\mathbf{x}(t = 0, 1, \dots, t)$ が定常であった場合、ある時刻 t での SRN の中間層の状態は現在の入力と過去の入力履歴の集合によって決まることになる。時刻 t における中間層の状態 $\mathbf{y}(t)$ は以下のように表現できる (図 2)。

$$\mathbf{y}(t) = \mathbf{W}^t \mathbf{y}(0) + \sum_{\tau=0}^{t-1} \mathbf{W}^\tau \mathbf{x}(\tau) \quad (2)$$

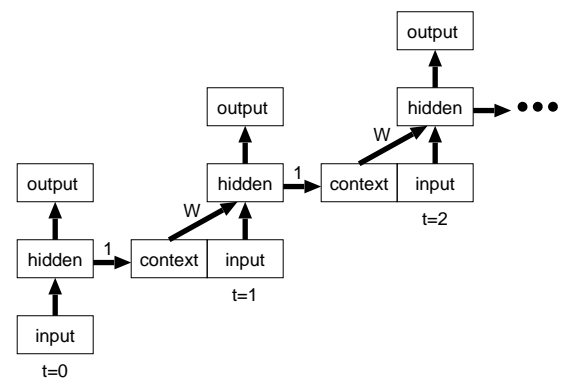


図2 Elman 型 SRN の時間発展

すなわち SRN の中間層の状態は過去の入力履歴 $\mathbf{x}(t)$ と初期値 $\mathbf{y}(0)$ に依存する。このことは、SRN が一時刻前の状態を保存しておくという単純な構造にもかかわらず、過去の全履歴に依存した出力を生じることを示している。後述するように、SRN のこの性質が連想記憶や間接プライミングのモデルとして有効なのである。

SRN の挙動は初期値 $y(0)$ と結合係数行列 W の構造によって定まる。行列の構造とは固有値のことである。一般に $Wx = \lambda x$ となる λ が存在する場合 λ のことを行列 W の固有値と言い、 λ に対応する x のことを固有ベクトルと言う。固有値とその固有値に対応する固有ベクトルを求めることを固有値問題を解くという。多変量解析の手法のほとんどはこの固有値問題を解くことに帰着する (例えば主成分分析, 古典的多次元尺度構成法, 判別分析など)。行列 W が正定値行列であれば行列 W のランク数だけ実数の範囲で固有値 λ が存在する。行動科学の分野や多変量解析の場合には一般に行列 W が正定値行列であることが保証されている。例えば W を相関係数行列と見なしたとき, 固有値のことを主成分負荷量とよび, 対応する固有ベクトルのことを主成分得点と呼ぶ。ところが, SRN の場合には文脈層から中間層への帰還信号の結合係数を定める W ではどんな値でもとりうるので, 正定値行列なとは限らない。すなわち固有値 λ は複素数をとることがあり得る。 W をベクトル x^t を x^{t+1} へ写像する行列であると見なした場合 x を変換する行列 W の固有値の正負や, 複素数になった場合の挙動の仕方は極めて複雑な軌跡を描くことになる。SRN における結合係数行列 W におけるこのような性質のため, 単なる線形変換のような単純な解と係数の関係を越えて, 複雑な解の挙動が可能となり, SRN の記述力, すなわち次節以降で説明する心理学的モデルの応用可能性を高めていると考えられる。

W の固有値を調べることによって SRN がどのように振舞うかを以下のとおりに分類することができる皓司 (1995)。

1. 全ての固有値が実根の場合

(a) すべての固有値が互いに異なる場合

- i. 全ての固有値が負の場合, 安定な平衡点
- ii. 全ての固有値が正の場合, 不安定な平衡点
- iii. 一つの固有値が 0 の場合, 0 でない固有値に対応する固有ベクトルをとおり平行な解曲線群
- iv. 固有値の符号が異なる場合, 鞍形点 (馬の背のように, あるいは峠のように, ある方向から見れば極小点だが, 別の方向から見れば極大点であるような形をいう。)

(b) 重根で対角化可能な場合

- i. 全ての固有値が負の場合, 安定な平衡点
- ii. 全ての固有値が正の場合, 不安定な平衡点

(c) 重根で対角化不可能な場合

2. 複素根の場合

- (a) 実部が正の場合, 不安定な渦状点
- (b) 実部が負の場合, 安定な渦状点
- (c) 実部が 0 の場合, 渦心点

文脈層からの影響を表す W のすべての固有値が 1 より小さければ過去の入力からの影響が指数関数に従って小さくなることを表している。すなわち $w_{ij} < 1$ ならば過去の影響は指数関数に従って急速に小さくなる。

上記の理由から SRN が安定な平衡点 (たとえば attractor など) を持つための必要十分条件は W のすべての固有値が 1 以下であること, のように思えるが, そうではない。一般にはシグモイド関数などによる非線形変換が行なわれるので, 上記の線形性が成り立たない。なぜならばシグモイド関数の値域は $[0, 1]$ に限定されるため W の全ての固有値が 1 以上でも発散することはないからである。一例を示せば,

$$W = \begin{bmatrix} 0 & -0.25 \\ -0.25 & 0 \end{bmatrix} \quad (3)$$

であれば W の 2 つの固有値は $\lambda = 0.25, -0.25$ となるので鞍点となり正の固有値に対応する固有ベクトルの方向に発散する。このときの中間層の状態のベクトル場を図 3 左に示した。一方, 同じ結合係数行列でもシグモイド関数の出力を仮定すれば図 3 右のように $(1, 1)$ が安定な平衡点となる。なお, シグモイド関数の傾きを小さくし原点近傍での挙動を考えれば任意の区間で線形近似が成り立つと考えても良い。従って本稿では線形の中間層を用いた場合に説明が容易になる場合には, 線形出力ユニットを仮定する。

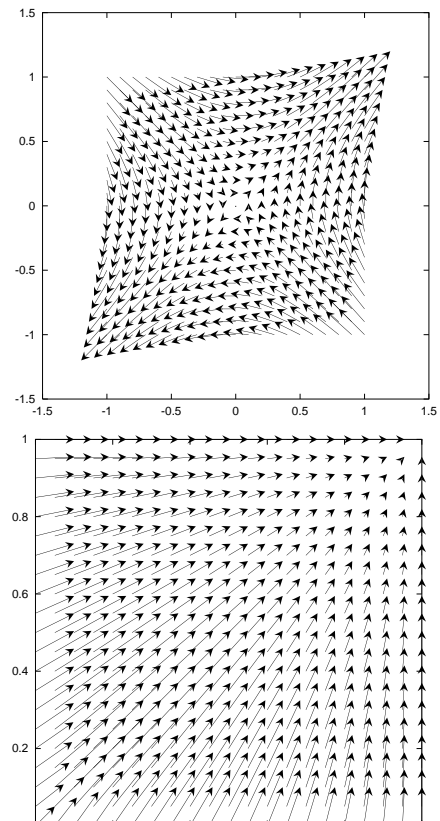


図 3 中間層の遷移状態を表すベクトル場。線形ユニットの場合 (左) と左と同じ結合係数行列を持つシグモイド関数による非線形変換の違い

各時刻における入力 $\mathbf{x}(t)$ が変化した場合,

$$\frac{d\mathbf{y}(t)}{dt} = \mathbf{W}\mathbf{y}(t) \quad (4)$$

$$\mathbf{y}(t) = \mathbf{y}_0 e^{-\mathbf{W}(t)} + \int_{t_0}^t e^{\mathbf{W}(t-u)} \mathbf{x}(u) du \quad (5)$$

1.2 時系列

ある時刻 t_0 から t までの入力 $\mathbf{x}(t)$ が定常であった場合, SRN の状態は現在の入力と過去の入力履歴の集合によって決まることになるので, 時刻 t における出力 $\mathbf{y}(t)$ は以下のように表現できる。

$$\mathbf{y}(t) = \mathbf{y}_0 \exp(\mathbf{W}(t)) \quad (6)$$

ここで \mathbf{y}_0 は初期値である。6 より SRN が安定な平衡点をもつための必要十分条件は固有方程式 $|\mathbf{W}(t) - \lambda \mathbf{I}| = 0$ のすべての根が負の実数部を持つことである。一般にはシグモイド関数などによる非線形変換が行なわれるので, 上記の線形性が成り立たず, かならずしも $|\mathbf{W}(t) - \lambda \mathbf{I}| = 0$ という条件を満たす必要がない。なぜならばシグモイド関数の値域は $[0, 1]$ に限定されるため 6 式が発散することはないからである。

1.3 連続時間近似と線形近似による形式的な厳密解

外部入力 $\mathbf{x}(t)$ が任意の時間間隔で一定であると仮定し, 時間間隔 $t = 1, 2, \dots$ が時定数 τ に比べて十分に長いと考えて連続近似可能であると考えれば, 微分方程式

$$\tau \frac{d\mathbf{y}(t)}{dt} = f(\mathbf{W}\mathbf{y}(t) + \mathbf{x}(t)) \quad (7)$$

の挙動を調べればよい。この式は単純ではないが, $f(\cdot)$ が原点近傍で線形近似できると仮定する。すると

$$\tau \frac{d\mathbf{y}(t)}{dt} = \mathbf{W}\mathbf{y}(t) + \mathbf{x}(t) \quad (8)$$

となって解析が可能になる。この式の斉次方程式

$$\tau \frac{d\mathbf{y}(t)}{dt} = \mathbf{W}\mathbf{y}(t) \quad (9)$$

の解

$$\mathbf{y}(t) = \mathbf{y}_0 e^{\mathbf{W}(t-t_0)} \quad (10)$$

を元に定数変化法を使って求めることができる。 \mathbf{y}_0 の部分を t の関数 $\mathbf{z}(t)$ と考えて

$$\mathbf{y}(t) = \mathbf{z}(t) e^{\mathbf{W}(t-t_0)} \quad (11)$$

とおけば

$$\tau \frac{d\mathbf{y}(t)}{dt} = \frac{d\mathbf{z}(t)}{dt} e^{\mathbf{W}(t-t_0)} + \mathbf{z}(t) \mathbf{W} e^{\mathbf{W}(t-t_0)} \quad (12)$$

$$= e^{\mathbf{W}(t-t_0)} \frac{d\mathbf{z}(t)}{dt} + \mathbf{W}\mathbf{y}(t) \quad (13)$$

これに非斉次方程式に代入して

$$\mathbf{x}(t) = e^{\mathbf{W}(t-t_0)} \frac{d\mathbf{z}(t)}{dt} \quad (14)$$

をえる。この式を $\mathbf{z}(t)$ について解くと

$$\frac{d\mathbf{z}(t)}{dt} = e^{-\mathbf{W}(t-t_0)} \mathbf{x}(t) \quad (15)$$

$$\mathbf{z}(t) = \int_{t_0}^t e^{-\mathbf{W}(u-t_0)} \mathbf{x}(u) du + \mathbf{a} \quad (16)$$

非斉次方程式の初期値を $\mathbf{y}(t_0) = \mathbf{y}_0$ とすると斉次方程式の解を用いた定数変化法の定義式 $\mathbf{y}(t) = \mathbf{z}(t) e^{\mathbf{W}(t-t_0)}$ から $\mathbf{y}_0 = \mathbf{y}(t_0) = \mathbf{z}(t_0) e^{\mathbf{W}(t_0-t_0)} = \mathbf{z}(t_0)$, すなわち $\mathbf{z}(t_0) = \mathbf{a}$. 従って

$$\mathbf{z}(t) = \int_{t_0}^t e^{-\mathbf{W}(u-t_0)} \mathbf{x}(u) du + \mathbf{y}_0 \quad (17)$$

ゆえに解は

$$\begin{aligned} \mathbf{y}(t) &= \mathbf{y}_0 e^{-\mathbf{W}(t-t_0)} + \int_{t_0}^t e^{\mathbf{W}(t-t_0)} e^{-\mathbf{W}(u-t_0)} \mathbf{x}(u) du \\ &= \mathbf{y}_0 e^{-\mathbf{W}(t-t_0)} + \int_{t_0}^t e^{\mathbf{W}(t-u)} \mathbf{x}(u) du \end{aligned} \quad (19)$$

となる。この式は SRN が初期値 \mathbf{y}_0 と結合係数行列の構造によって決定論的に定まることを意味している。

上記では原点近傍で議論を展開したが, 数学的には任意の領域に分割した領域内で線形性が満たされ, かつ, リプシッツの条件

$$|f(\mathbf{y}_1) - f(\mathbf{y}_2)| \leq N w_{\max} |\mathbf{y}_1 - \mathbf{y}_2| \quad (20)$$

が成立する場合に当てはまる議論である。ここで w_{\max} は結合係数行列 \mathbf{W} の要素 w_{ij} の中で最大のものとする。すなわち (20) が成り立つ条件とは, 関数 f によって写像された 2 つのベクトルの距離が変換前のベクトルの距離に結合係数の最大値に次元数をかけた値よりも小さいか等しい, という条件である。

この場合ネットワークが安定な平衡点を持つための必要十分条件は固有方程式

$$|\mathbf{W} - \lambda \mathbf{x}| = 0 \quad (21)$$

のすべての根が負の実数部を持つことである。すなわち (21) 式に対応する固有方程式の解, 行列 \mathbf{W} のすべての固有値が負の実数部を持つ必要がある。全ての固有値が負の実数部を持つことが SRN において安定した解が得られる必要十分条件である。ここでも実際の SRN においては, シグモイド関数の非線形性によって上式は必要十分条件とはならないことに注意が必要である。

2 心理学モデルとしての SRN

上記の形式解のみでは明確でないが、SRN の心理学的応用として以下のことが考えられる。以下では SRN の中間層の状態の遷移によって心理現象が説明可能であることを示す。簡単のため中間層の状態空間を 2 次元として図示しているが、中間層の数は任意であるため、一般性を損なうことなく多次元状態空間への一般化は可能である。

2.1 短期記憶のモデルとしての SRN

文脈層からの信号 $\mathbf{W}(t)$ と入力層の信号とで表される結合係数行列の固有方程式が複素解を持つならばリミットサイクル等の解が得られることになる。特に、中間層ユニットの出力を線形とし、かつ結合係数行列 \mathbf{W} の固有値の実部が 0 であれば、中間層の遷移状態は完全な渦心点になる。簡単な例として図 4 のような中間層の素子数が 2 個の場合を考える。

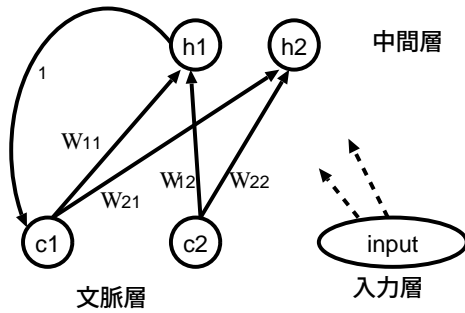


図 4 2 ユニットの単純な SRN

文脈層から中間層への結合係数 $\mathbf{W}(t)$ が

$$\begin{pmatrix} w_{11} + x_1(t) & w_{12} \\ w_{21} & w_{22} + x_2(t) \end{pmatrix} = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \quad (22)$$

のようになっていたとする。このとき $\mathbf{W}(t)$ は回転行列であるから、例えば $\theta = \pi/2$ なら、単位円上の 4 点を循環する。この 4 点に対応する状態と特定の出力が結び付いているのなら、入力層からの信号と文脈層からの信号の和 $\cos \theta = w_{11} + x_1(t) = w_{22} + x_2(t)$ が定常であるかぎり 4 状態を永遠に循環する出力が得られる。 $n\theta = 2\pi$ を満たすかぎり n 個の状態を循環する出力を得ることができるが、 θ の取り方は任意であるので、SRN には任意の長さの系列を記憶する能力 Rodrigues et al. (1999) があることがわかる。図 5 に係数行列が回転行列になる場合の中間層の状態を示した。図 5 左図は完全な回転行列である。一方、図 5 右図は回転行列に外部入力加わったときの状態を示している。

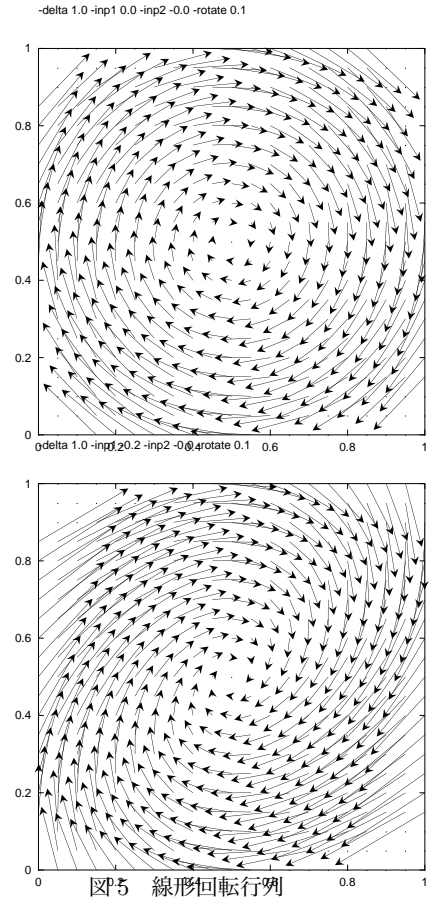


図 5 線形回転行列

このように SRN の中間層ユニットの活性値で定義される中間層の状態遷移が円周軌道上の軌跡となると想定する。この円周軌道上の任意の一点が特定の記憶項目と対応していると考え。1つの出力層ユニットによって中間層の円周軌道を横切る一本の直線が定義される。(図 6)

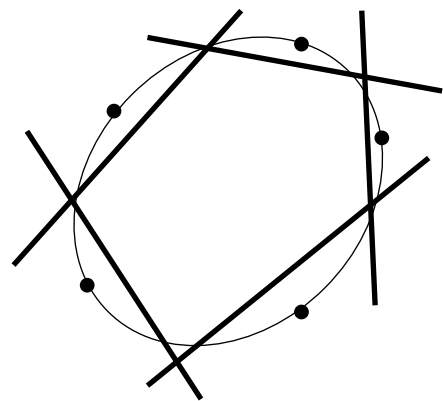


図 6 中間層から出力層への写像は線形分離可能なニューロンの動作方程式で記述できる

中間層の内部状態を表す軌道が円または楕円となるので、円周上の一点を含み他の点を含まない領域に分割することが可能である。この分割によって任意の記憶項目と中間層との円周軌道上の一点とを同一視することが可能

になる。

すなわち SRN を短期記憶と見なすモデルとは次のようになる。まず、中間層の意味付けは、入力信号と文脈層から中間層への帰還信号とによって活性値が定まる中間層ユニットの状態を多次元空間上の 1 点と見なすことに相当する。次に、出力層の意味付けは、中間層ユニットの活性値で表現された多次元空間を線形分離する機構と見なすこと相当する伸一 (2001)。換言すれば、各時刻における中間層ユニットの状態が 1 つの記憶項目を表象し、出力層のユニットは中間層の活性値で定義された多次元空間の領域を線形に分割することによって記銘内容を取り出す機構であると解釈することが短期記憶のモデルとしての SRN の要点である。2 層のパーセプトロンが入力空間を線形に分離すること機構であることは Minsky と Papert1988 以来明らかであり、中間層と出力層との関係は、このような多次元空間の線形分割と考えることが出来る。保持している内容を言語報告するような実験課題においては中間層の特定の状態と報告すべき言語表象の結びつきの強弱によって表現される。出力層ユニットがシグモイド関数だと仮定すれば、各出力層の活性値を想起すべき項目の正解率と解釈できるかも知れない。

中間層の状態を線形分離する能力が $n = 7 \pm 2$ 程度 (すなわち $2\pi/7$ だけ回転する行列 W を用いる) の状態しか保持できないと仮定すれば、 7 ± 2 個の長さからなる系列を記憶できる。すなわち SRN は短期記憶のモデルとなりうる。

短期記憶が 30 秒から 60 秒程度といわれる短期記憶の保持期間については、図 5 の左右の図の様子から次のように解釈できる。 7 ± 2 程度の短期記憶を保持してするためには SRN に意図的に外部入力を加える必要がある。ところが外部入力が入ると時間経過にしたがって楕円軌道が小さくなる (すなわち渦心点) になる (図 5 の右)。したがって、時間の経過にしたがって渦巻き状に中間層の状態は定常点に向かって縮小するので、次第に記銘項目を保持するのがむずかしくなってくると解釈できる。ゆえに反復リハーサルなどの短期記憶を保持するため内的努力は、この中間層の楕円軌道を保持するために用いられると考えることが出来る。記銘すべき項目が多ければ多い程、すなわち中間層の楕円軌道を乱すような外部入力が多い程中間層ユニットの状態である楕円軌道を保持することが困難になってくると考えればよい。

また、語想起や喚語困難など記憶障害の場合などは SRN によって次のように説明される。

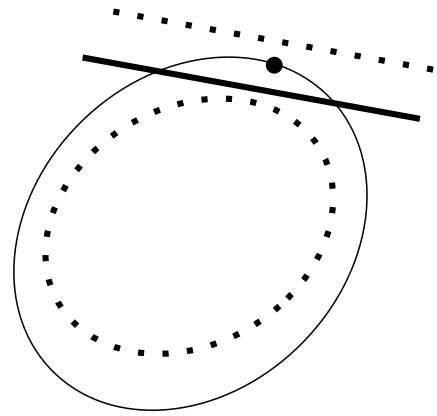


図 7 語想起、記憶の誤り、を示す模式図

中間層から出力層への写像に劣化が起これば線形判別直線 (図 7 中の点線で描かれた直線) が移動し任意の状態空間を分割することができなくなる。一方、文脈層から中間層への結合係数に劣化が起きれば回転する円の形状が変化し、任意の状態を表現することができなくなる (図 7 中の点線で描いた円)。SRN による短期記憶モデルでは、中間層の円環状態による記憶の保持の問題と、出力層ユニットが中間層の状態空間を分割することによる記銘内容の検索の問題とを明確に区分して捉えることができる。

2.2 反応時間のモデルとしての SRN

中間層の状態がある一定の領域に入ったとき反応が起きると仮定すれば、SRN は反応時間のモデルとなり得る。例えば

$$W = \begin{bmatrix} 5 & -6 \\ -6 & 5 \end{bmatrix} \quad (23)$$

であるとし、かつ、入力信号は $x_1 = x_2 = 1.0$ とした場合の中間層の遷移を図 8 に示した。

図 8 3 つの平衡点をもつ SRN のベクトル場。2 つの平衡点は安定だが、他の 1 つは不安定である。

図 8 から平衡点が 3 点存在することがわかる。(0.5, 0.5) は不安定な平衡点である。他の 2 つ (1, 0) 及び (0, 1) は安定な平衡点である。このとき、出力層のユニットは中間層の状態を線形分離可能な領域に区分すると考えれば、中間層から出力層への結合係数によって任意の領域に分割された A または B の領域に入ったときに反応が生起すると考える。入力が定常であると仮定したとき、A, B どちらかの領域に入るまでの繰返し数を反応時間と同一視する。特定の反応基準に達するまでの繰返し数を反応時間と同一視する考え方は、ニューラルネットワークモデルによる反応時間の説明である GRAIN (Graded Random, and Interactive Networks) モデル McClelland (1993) と同じ発想である。本稿で提案する SRN による反応時間と GRAIN モデルに

よるそれとの違いは、GRAIN モデルが選択肢に対応する個々ユニットに反応基準を仮定しているのに対して、SRN の反応時間は全ての中間層ユニットの状態空間を仮定し、この状態空間の線形分離可能な領域に入ることによって反応が生起すると考える点である。図 9 に SRN を反応時間のモデルと見なしたときの概念図を示した。

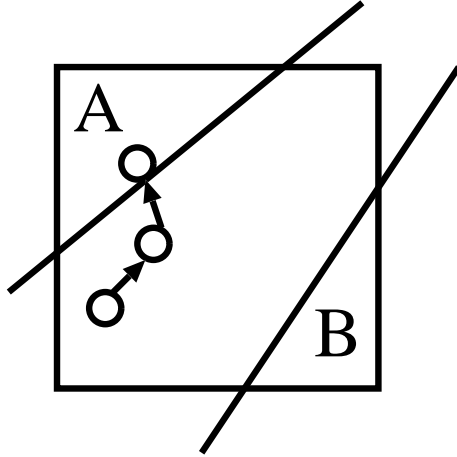


図 9 SRN を反応時間のモデルとして捉えた場合の概念図。2 肢選択反応時間の場合 A, B いずれかの領域に入るまでの繰り返し数が反応時間とみなされる。

図 9 では A, B という 2 肢選択反応時間の場合である。A, B いずれかの領域に入るまでの繰り返し数は結合係数行列と入力信号によって定まる W の固有値の大きさに依存する。固有値が大きければ速やかに A, B いずれかの領域に引き込まれるが、固有値が小さければ各領域に達するまでの繰り返し数が多くなり、反応時間が長くなると予想される。反応時間における条件間の差異や被験者効果、学習効果は W の固有値の大きさによって調整されていると考えることができる。

なお、 W の値の定め方によって、SRN は複数の反応時間モデルと等価な振る舞いをするモデルと見なすことができる。例えば $w_{ij} = 0$ であれば、単純な競争モデル Townsend and Ashby (1983) であり、 $w_{ij} < 0$ ならば Ratcliff らの酔歩モデルと見なすことができる Ratcliff (1978, 1981, 1988)。

競争モデルとは各反応に 1 つずつ対応する内的な情報蓄積機を仮定するモデルである。ある一定以上の情報が蓄積されたとき該当する反応が観測されることを仮定するモデルであり、各情報蓄積機間での相互作用は仮定されない ($w_{ij} = 0$)。

一方、酔歩モデルでは情報蓄積機間に負の結合 ($w_{ij} < 0$) を仮定する。一方の反応に対応する情報が与えられることは、他方の反応を抑制することに相当し、反応はどちらか一方の情報蓄積機が一定の閾値に達したときに生起するものとする。

より一般的に、文脈層から中間層への結合係数を $w_{ii} < 0$ なら減衰あるいは崩壊係数と捉え、 $w_{ij} < 0$ なら互いに競合すると考えることができる。すなわち SRN は LCAM (Leaky Competing Accumulator Model) モデル Usher and McClelland (2001) と同じ枠組と捉えることができる。LCAM モデルは選択肢が 2 個以上の場合を許し、かつ、パラメータの与え方によっては、競争モデルや酔歩モデル (拡散モデル) をその特別な場合として含むモデルである。

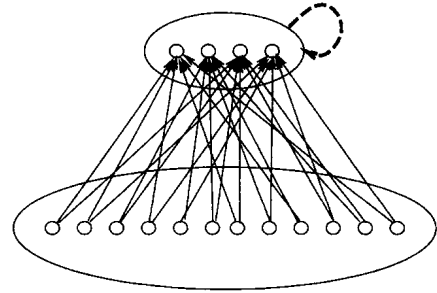


図 10 LCAM モデルの概念図

LCAM モデルは、側抑制と再帰興奮、および線形閾値関数 ($f_i = x_i$ for $x_i \geq 0$ and $f_i = 0$ for $x_i < 0$) を特徴とするモデルである。LCAM モデルでは制限された時間範囲内に、もっとも高い活性値を持ったユニットが選択され、通常の反応時間課題では、もっとも速く基準に達したユニットによって反応が決定されるモデルである。LCAM モデルの特徴である側抑制と再帰興奮は SRN の W の中に埋め込むことが可能である。すなわち側抑制は $w_{ij} < 0$ を意味し、再帰興奮は $w_{ii} > 1$ とすればよい。このことから SRN は記述力の高い一般的な反応時間モデルと見なすことができる。

反応時間のモデルを考える場合、反応時間の分布を予測し、エラー率も予測できなければならない。LCAM モデルでは動作方程式として確率微分方程式を採用し、動作方程式中にガウシアンノイズを導入し反応時間の分布やエラー率の説明を試みている。

SRN を反応時間モデルとみなして、反応時間の分布やエラー率を考える場合、LCAM モデルと同じように何らかの確率論的要素、たとえば攪乱項 (ガウシアンノイズのような) を導入しなければならないだろう。SRN の結合係数が定まっていれば、その動作は決定論的に定まるからである。一定時間内に反応が生じるように制約をかけた場合、その時点でもっとも高い活性値をもったユニットによって反応が決定され则认为ることにより、反応時間の分布やエラー率の予測を行なうことになる。LCAM モデルはその動作を確率過程の一つであるオルンシュタイン-ウーレンベック過程として記述でき、反応時間の分布やエラー率をシミュレー

トしている。SRN は LCAM モデルを特殊な場合として含む一般モデルであるので同じような分析が可能になる。

モデルの動作方程式は次式 (24) のとおりである。

$$dx_i = [I_i - \lambda x_i] \frac{dt}{\tau} + \xi_i \sqrt{\frac{dt}{\tau}} \quad (24)$$

ここで、 ξ_i は平均 0、分散 σ^2 のガウシアンノイズであり、 ξ の時定数の平方根をとるのはガウシアンノイズを標準偏差を変数とみなすためである。外部入力を ρ , leak 係数を λ , 興奮性の再帰結合を α , 側抑制の影響を β とすれば

$$dx_i = \left[\rho_i - \lambda x_i + \alpha f_i - \beta \sum_{i' \neq i} f_{i'} \right] \frac{dt}{\tau} + \xi_i \sqrt{\frac{dt}{\tau}} \quad (25)$$

となる。ここで $\lambda - \alpha$ を κ とおいて、選択肢が二つだけのシンプルな場合を書き下せば、

$$\begin{aligned} dx_1 &= [\rho_1 - \kappa x_1 - \beta x_2] \frac{dt}{\tau} + \xi_1 \sqrt{\frac{dt}{\tau}} \\ dx_2 &= [\rho_1 - \kappa x_2 - \beta x_1] \frac{dt}{\tau} + \xi_2 \sqrt{\frac{dt}{\tau}}, \end{aligned} \quad (26)$$

さらに、新たな変数 $x = x_1 - x_2$ を考え、外部入力に $\rho_1 + \rho_2 = 1$ という制約を置いて、

$$dx = [(2\rho_1 - 1) - (\kappa - \beta)x] \frac{dt}{\tau} + \xi \sqrt{\frac{dt}{\tau}} \quad (27)$$

$\nu = 2\rho_1 - 1$ とすれば (27) は Ornstein-Uhlenbeck (OU) 過程となる。 $\kappa = \beta = 0$ なら古典的アキュムレータモデルに一致する。 κ は自己結合、 β は側抑制である。Busmeyer and Townsend1993 でも κ が使われている。Smith1995 では $\kappa < 0$ が単純反応時間のモデルで使われている。すなわち、意志決定モデルから提案された Roe, Busmeyer, Townsend2001 の MDFT モデルも入力表現の違いを除けばニューラルネットワークモデルの構成上からは、LCAM モデルと等価であるとみなすことができる。

2.3 選択的注意のモデルとしての SRN

文脈層ユニットから中間層ユニットへの結合係数行列 \mathbf{W} において $w_{ij} = -1$ すなわち各ユニットの活性値が競合して相互抑制する場合を考える。 $w_{ii} = 1$ とすれば、中間層の状態遷移は、少しでも入力信号が大きい刺激に対する反応を選択するニューラルネットワークとなる。すなわち文脈層から中間層への結合係数が $w_{ij} = -1$ の SRN は、勝者占有方式となるので選択的注意のモデルとしてとらえることができる。図 11 はこのことを示している。

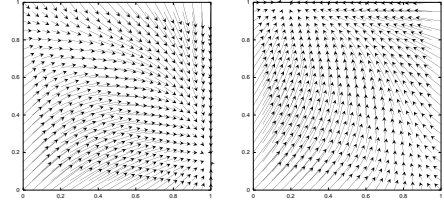


図 11 選択的注意モデルとしての SRN。入力信号のわずかな違いによって平衡点が異なる

すなわち、2 種類の入力信号のうちいずれか一方の信号強度が他方の信号強度より大きければ、大きい方のユニットの活性値が優勢となり、他方のユニットの活性を抑制してしまう。図 11 左は 1 番目のユニット (横軸方向のユニット) に対する信号強度の活性値付近に平衡点が存在し、時間の経過に伴って (1, 0) 付近に収束することを示している。他方 2 番目のユニットの信号強度が 1 番目のユニットの信号強度よりも強ければ図 11 右のように (0, 1) 付近に平衡点が存在する。このことは、相互抑制 $w_{ij} = -1$ であれば、入力信号が n 個の場合でも成り立つ性質である。すなわち複数の信号のうち最も大きな信号強度の入力だけが生き残り、他の入力信号に対応する活性値を抑制してしまうことを意味している。このことは、複数の視覚対象が視野内に存在した場合、もっとも目立つ視覚対象に注意が捕獲されることのアナロジーとして考えることができるだろう。

図 11 ではお互いユニットの抑制のすべてが $w_{ij} = -1 (i \neq j)$ と対称な場合を示したが、SRN は $w_{ij} < 0$ であれば相互抑制として働くので、特定の視覚対象に選択的に注意が向けられる選好の程度を w_{ij} の絶対値として表現可能である。また、入力信号が左右の視野を表す 2 つであるとし、いずれか一方の入力信号強度が劣化していると考えれば、反側空間無視の脳損傷患者のモデルとしても用いることができる。

さらに、注意の対象をあるものから別のものにシフトさせるには、結合係数行列 \mathbf{W} の対角要素 w_{ii} を $0 < w_{ii} < 1$ として徐々に減衰する項を含めれば良い。このとき w_{ii} の値が小さければ小さい程急速に活性値が減衰するので、この減衰によって注意を向けられた対象に対する活性値が相対的に小さくなり、別の対象に注意を向けるメカニズムとなり得るだろう。

注意のモデルとしては選択的注意のモデルだけでなく注意分割の側面についても言及しておく必要があるだろう。注意分割の場合、同時に複数個 (同時に提示された全刺激より少ないが 2 個以上) の対象に注意を分割することを考えなければならない。これまでの説明では勝者占有 (winner-take-all) 方式によって 1 つの対象に注意が捕獲される過程を考えてきたが、注意を向ける対象を複数にする

ニューラルネットワークを用いれば注意の分割問題も説明できると考える。すなわち勝者占有方式にあたる結合係数行列を k 者占有方式 ($k \Rightarrow 2$) に換えればよい。例えば kWTA (k -winners-take-all) 回路 Majani, Erlanson, and Abu-Mostafa (1989) を用いればよい。ただし Majani らの kWTA 回路を実現するためには文脈層のユニット数を中間層 +1 だけ用意しこの 1 つ余分なユニットによって全体の活性値が制限されるようにしなければならない。

2.4 間接プライミングのモデルとしての SRN

Cleeremans, Servan-Schreiber, and McClelland 1995 によれば SRN は系列予測を行なうことができる。彼らの研究によれば SRN は $AxxxxA'$, $BxxxxB'$ のような 2 つの異なる系列を区別して識別する能力があり、中間の文字 x を任意の個数埋め込むことができる。すなわち SRN は、任意の時間の後に、最初に与えた刺激の違いに応じて異なる出力を生じる能力を有する。

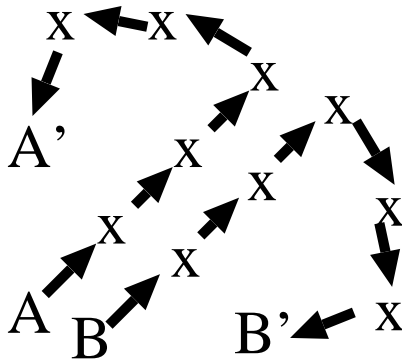


図 12 間接プライミングモデルとしての SRN。任意の長さの系列を判別することが可能である

間接プライミングモデルとしての SRN は反応時間モデルとしての SRN とほぼ同様の枠組で説明可能である。最初の記号 A, B をプライム刺激としてとらえ、最後の記号 A', B' を反応ととらえれば SRN はプライミングのモデルとして表現できる。最初に与える刺激をプライム刺激と考えておけば、SRN で系列想起をしたときの中間層ユニットの状態を、プライム刺激を与えなかったときの中間層ユニットの状態と比較することができる。すなわち最も反応時間が早くなる (従ってプライミング効果の最も大きい) 中間層のユニットの状態と、その他の中間層の状態との距離とがプライミング量 (すなわち反応時間の差異) に比例すると考えれば良い。中間層ユニットの一つ一つの活性値が 0 から 1 の間の値を取るとすれば、 n 個の中間層によって n 次元の距離が定義できる。中間層ユニットの活性値で定義されたこの距離に、反応時間の促進、あるいは抑制が比例すると考える。また中間層から出力層への変換には、シグモイド関数

など S 字状の非線形の変換がなされることで、異なるプライム刺激の効果が大きく変化することを説明できるだろう。

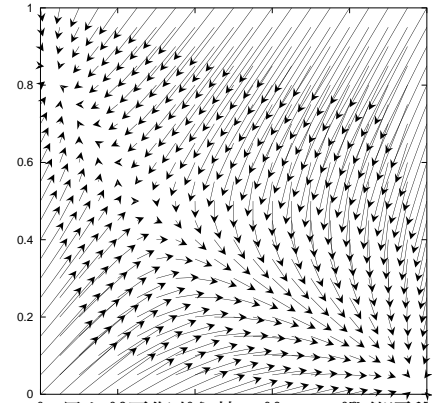


図 13 2 つの異なる平衡点を持つ SRN の状態遷移図

例えば図 13 のように異なる 2 つの平衡点を持つ SRN を構成することが可能である。類似度の高いプライム刺激を与えても、上の説明で用いたように中間層の状態 (図 13 の場合 2 次元平面上の距離) とプライミング効果 (反応時間の差異) が異なる SRN を作ることが可能である。

2.5 連想記憶のモデルとしての SRN

今まで見てきた心理学モデルの場合と同じく、連想記憶においても SRN の解の挙動は文脈層ユニットから中間層ユニットへの結合係数と入力信号とを合わせた $\mathbf{W}(t)$ の固有値によって定まる。 $\mathbf{W}(t)$ の固有値が 1 より大きければ、対応する固有ベクトルの方向へ大きくなる。シグモイド関数の影響により、固有ベクトルの方向に向かった出力は平衡状態になる。この平衡状態が任意の記憶に対応していると考えれば SRN は連想記憶、または自己想起記憶装置として機能すると言える。実際に $\mathbf{W}(t)$ の値を変化させることによって非線形で複雑な状態遷移を起こさせることができる。すなわちある状態から任意の状態へと遷移する FSA が形成され得る。

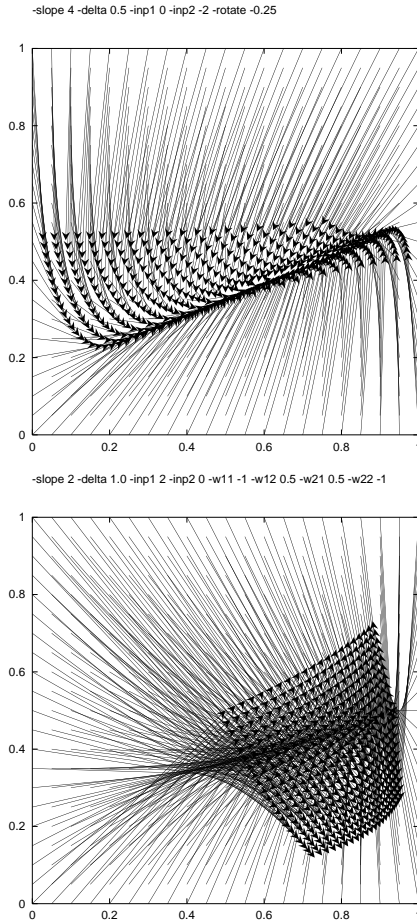


図 14 複雑な状態遷移を示す例

図 14 は複雑な状態遷移の例である。任意の状態空間から別の状態空間へ遷移するとき、その状態は複雑で遷移がクロスしたり折れ曲がったりしていることが見て取れる。このような状態遷移が SRN の特徴であり、中間層のユニットの状態の分割された空間を特定の連想項目と見なすことで連想記憶として振る舞うことが可能になる。反応時間やプライミングの節で述べたとおり、中間層ユニットから出力層ユニットへの結合係数によって n 個の中間層ユニットで表現された空間 $[0, 1]^n$ を線形判別することができるからである。

上記では 3 時刻で入力定常であることを仮定したが、より一般的に n 時刻まで入力定常であると考えれば

$$y = \sum_{\tau=1}^n x^t W^{\tau} x + x \quad (28)$$

となる。この式はテーラー展開によって任意の関数 y を近似している式であると解釈できる。このことから任意の関数を任意の精度で近似可能であるという、SRN の強力な能力を示していると解釈可能である。

2.6 分類学習のモデルとして —比較分類回路—

すべての結合係数が $-1 < w_{ij} < 0$ の場合入力の大きさを比較し x_1 が x_2/w より大きいときは x_1 のみ出力する。

x_2 が x_1/w より大きいときは x_2 のみ non-zero。 x_1 と x_2 との値が接近しているときには両方とも 0 でない出力をする。すなわち

$$wx_1 < x_2 < \frac{x_1}{w} \text{ または } wx_2 < x_1 < \frac{x_2}{w} \quad (29)$$

の範囲では両方出力される。 w の値が 0 に近ければ 2 つのユニットが同時に発火する領域が広がり、 $w = -1$ のとき同時に発火しなくなる。

2.7 系列予測あるいは文法推論のモデルとして

一般に任意の文字列から背後の文法を類推する問題を文法推論と呼ぶがこれは NP hard 問題として知られている。Elman1990 はこの問題に対して SRN が FSA 認識機として働くことを示した。

文法構造の発見を意図して行なわれた実験によって、SRN の学習は次の 3 つのフェーズで進行することが示されている Cleeremans et al. (1995)。

1. 最初のフェーズでは、文脈情報を無視する傾向にある。SRN は各入力文字と可能な全ての後続文字との連合を安定して抽出できる。文脈情報とは無関係に後続する文字の統計情報を反映するようになる。
2. 第 2 フェーズでは、文脈層上にコピーされたパターンは現在の文字に先行した文字によって指定される独自のコードを表象するようになり、同一文字が分注の異なる場所で起きることを区別し始め、SRN は、この文脈情報を利用するようになる。
3. 最後に、先行する要素の生起を反映する文脈情報の微小な差異によって、単語の文中における位置の相違を区別するようになる、

2 層の回路を考えて、下層には $w_{ij} = -1$ すなわち Winner-take-all 回路、上層は $w_{ij} = -2$ すなわち短期記憶回路を用いることでオートマトンとして動作する回路を作ることができる

中間層ユニットを $1-0$ の状態をとるとすれば、 n 個のユニットによって 2^n 個の状態を表現できる。各状態に対して個別の出力を考えれば SRN は完全に FSA に一致する。具体的には以下のように作る。

また、Cleeremans ら 1995 は、中間層の状態が 1 か 0 以外の段階的な変化で微妙な差異を表現できるという意味で Gated state machine と呼んだ。

2.8 ワーキングメモリのモデルとして —スタック操作関数—

入力 $x(t)$ によって $W(t)$ の固有値—従って SRN の振る舞い—は突然変化する。これによって平衡状態から突然抜け出して、出力 $y(t)$ はまったく異なる軌跡を描くことになる。このような振る舞いは再帰関数に例えることができ

るので、この作用によって埋め込み文の処理 Weckerly and Elman (1992) などが可能になると考えられる。

SRN には以上のような強力な性質があるので、言語情報処理などへの応用が可能なであろう。

先の時系列を接近効果、回転行列を用いた記憶容量は初頭効果を表すと考えれば、両者の組合せによって SRN は記憶実験の結果 U 字型の忘却曲線を表現するモデルになりうる。

2.9 リズムジェネレータ

層内のユニットが隣接するユニットとのみ抑制性の結合を有している場合、すなわち結合係数行列が

$$\mathbf{W} = \begin{pmatrix} 0 & w_a & 0 & \cdots & 0 & w_b \\ w_b & 0 & w_a & \cdots & 0 & 0 \\ 0 & w_b & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ w_a & 0 & 0 & \cdots & w_b & 0 \end{pmatrix} \quad (30)$$

ただし

$$1 > w_b \geq \frac{1}{w_a} > 0 \quad (31)$$

の場合、リズムジェネレータとして振る舞う。

2.10 結合振動子

2 つの素子 x, y を考えて $w_{yx} = -1, w_{xy} = 1, w_{xx} = w_{yy} = c$ とすると

$$\begin{pmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{pmatrix} = \begin{pmatrix} c & 1 \\ -1 & c \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \quad (32)$$

$$\frac{d}{dt} \mathbf{x} = \mathbf{M} \mathbf{x} \quad (33)$$

このときの振る舞いは行列 \mathbf{M} によってきまり、 \mathbf{M} の 2 つの固有値は $c \pm i$ で与えられる。 $c = 0$ なら同心円状の解となる。このことは次のようにしても確かめることができる。 x, y を極座標に変換して

$$\begin{aligned} x &= r \cos \theta \\ y &= r \sin \theta \end{aligned}$$

で表される r, θ を用いて書き直せば

$$\begin{aligned} \begin{pmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{pmatrix} &= \begin{pmatrix} \frac{dr}{dt} \cos \theta - r \sin \theta \frac{d\theta}{dt} \\ \frac{dr}{dt} \sin \theta + r \cos \theta \frac{d\theta}{dt} \end{pmatrix} \\ &= \begin{pmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{pmatrix} \begin{pmatrix} \frac{dr}{dt} \\ \frac{d\theta}{dt} \end{pmatrix} \\ &= \begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{pmatrix} \begin{pmatrix} \frac{dr}{dt} \\ \frac{d\theta}{dt} \end{pmatrix} \\ &= \mathbf{J} \begin{pmatrix} \frac{dr}{dt} \\ \frac{d\theta}{dt} \end{pmatrix} \end{aligned} \quad (34)$$

なので、行列 \mathbf{M} の x, y を r, θ に変換して

$$\begin{aligned} \left\{ \begin{pmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{pmatrix} = \right\} \mathbf{J} \begin{pmatrix} \frac{dr}{dt} \\ \frac{d\theta}{dt} \end{pmatrix} &= \mathbf{M} \begin{pmatrix} x \\ y \end{pmatrix} \\ \mathbf{J} \begin{pmatrix} \frac{dr}{dt} \\ \frac{d\theta}{dt} \end{pmatrix} &= \mathbf{M} \begin{pmatrix} r \cos \theta \\ r \sin \theta \end{pmatrix} \\ \begin{pmatrix} \frac{dr}{dt} \\ \frac{d\theta}{dt} \end{pmatrix} &= \mathbf{J}^{-1} \mathbf{M} \begin{pmatrix} r \cos \theta \\ r \sin \theta \end{pmatrix} \\ \begin{pmatrix} \frac{dr}{dt} \\ \frac{d\theta}{dt} \end{pmatrix} &= \begin{pmatrix} cr \\ -1 \end{pmatrix} \end{aligned} \quad (35)$$

従って解曲線

$$r(t) = r(0)e^{ct} \quad (36)$$

$$\theta(t) = \theta_0 - t \quad (37)$$

を得る。この曲線群は c の値によって、 $c > 0$ なら渦巻き状に発散、 $c < 0$ なら原点に収束し、 $c = 0$ のとき渦心点となる。いずれの場合も時計回り ($-t$) 方向に進行する。

c が x, y の関数であれば、より複雑な関数となる。例えば $c = 1 - x^2 - y^2$ のような

$$\begin{aligned} \frac{dx}{dt} &= \{1 - x^2 - y^2\}x + y \\ \frac{dy}{dt} &= -x + \{1 - x^2 - y^2\}y \end{aligned}$$

であればリミットサイクルとなる。

3 考察

本稿では SRN がさまざまなモデルを説明する能力があることを示してきた。これら以外にも、SRN には時系列予測、文法推論が可能である Cleeremans et al. (1995)。このように一つのニューラルネットワークで複数の心理現象を説明できることから以下のような点が指摘できる。

シナプス結合の種類以上に多様な機能が脳内で実現されている理由

SRN の文脈層から中間層ユニットへの結合係数 w_{ij} をパラメータと考えれば、パラメータを様々なに設定することによって、異なる認知機能を表現できる。このことは、単一ユニットの動作が同一であるという証拠 (外部信号によって惹起される、脱分極、過分極、不応期の存在等) から一歩進んで、神経結合の様子が一見同じように見える領域でも異なる認知機能を実現し得ることを示している。

高次脳機能マップが簡単に描けない理由

SRN の機能を決めるのは文脈層から中間層への結合係数と入力信号とで表される \mathbf{W} である。文脈層から中間層への結合係数 w_{ij} はほとんどの神経細胞で出生直後には可塑性を持った状態になっているのだろう。脳の機能が多様であるのに対し、ニューロンの結合方式は興奮か抑制かなどの単純な結合様式しか発見されていない。このことの説明として SRN の能力の多様性が挙げられる。

例えば Damasio and Damasio(1992) は、言語や他の高次認知機能は個人によって大きく異なる可能性を示唆している (Elman et al., 1996, も参照)。ある脳内の領域がどのような機能を果たすかは、その後の経験によって如何様にでも変化するのであろう。ある人にとって言語機能に関係する領域でも別の人にとっては異なる認知機能を実現するために使われる領域であっても不思議はない。

機能的脳画像研究の限界 fMRI に代表される非侵襲的機能的脳画像測定法では、ある実験課題を解決するために必要な部位の活性化の有無が確認されるだけである。換言すれば、その部位がどのような機能を果たしているのかを確かめる方法ではない。脳内の一部位に対して、その部位の計算論的な意味付けをするためには、その部位が他の部位とどのような神経結合をしているのかを知らなければならない。本論文で示したように、ある部位と他の部位とのシナプス結合の強さが微妙に変わることによって、様々な認知機能が実現できる。すなわちこのことは、脳内の一部の部位の活性化だけを測定しているだけでは脳の計算論的な意味付けができないのだ、ということを示していると思われる。同時に活性化する部位の神経結合の程度 (結合系数) が分かれば、その部位の計算論的役割が明らかにできる可能性がある。任意の二つの部位が SRN のような関係にあるとすれば、少なくとも本論文で挙げたすべてのモデルの中の 1 つである可能性がありうる。そうだとすれば、SRN の結合系数行列を測定できる生理学的手段なり指標なりが開発されれば、この部位間の結合系数によって表される神経回路の計算論的な意味を確認することができるかも知れない。かつて Marr が小脳はパーセプトロンであることを予言し、伊藤によって生理学的に証明されたように、大脳皮質の一部が SRN のように振舞うことが生理学的に確認されるようになるかも知れない。

laterality の起源 なぜ言語野は右半球に存在するのか。左右の半球間で相互抑制回路が構成されている。articulatory output はほとんどが両側性支配なので出生直後は左右いずれの半球にも言語野が形成される可能性がある。しかし利手の使用によって周囲の両野にも活性が伝播すると考えれば audio, visual input と articulatory output との間に Hebbian rule が存在して結合強度が変化し、結果として利手の contralateral hemisphere に言語野が形成されやすい

2 段階認識モデルへの批判 入力刺激の同定 → 認識過程のような古典的な認識モデルでは説明することが難し

いモデルを表現できる。V1 は $-1 < w_{ij} < 0$ のような大雑把な分類のみを行なうことができるが V1 からのフィードバックに支えられて $w_{ij} = -1$ のように振る舞うことができるようになる。Farah も指摘しているとおり。Jordan net もこの変形。Zeki の再帰帰仮説や ART も同様の発想によるものである。従って priming の説明もこのモデルで可能であろう。

4 結語

本稿では SRN の応用可能性を議論し、複数の心理現象が一つの SRN というモデルの枠組みで記述できることを示した。モデルの持つ高い記述能力は FSA として記述できるすべてを含んでいる。ここではその心理現象への適用可能性のみを示した。

本稿では実際の心理実験データへの当てはめは行っていない。SRN の複雑な振舞いは中間層のユニット数と文脈層からの帰還信号の結合係数によって定まるが、本論文で示したとおり原則として中間層数のユニットを 2 とした場合でも十分に複雑な心理現象を説明できる可能性がある。ただし実際のデータへの当てはめについては、より多くの中間層ユニットが必要なのかも知れない。ニューラルネットワークの複雑さは中間層ユニット数の数で表現される場合とモデルの階層数やアーキテクチャで表現される場合とがあるが、SRN の能力の高さを考えた場合、後者のアーキテクチャの複雑さを増す必要はほとんど不要なのではないかと思われる。

References

- Adamson, M. J., & Damper, R. I. (1999). B-raam: A connectionist model which develops holistic internal representations of symbolic structures. *Connection Science*, 11, 41-71. (21-Oct.2001)
- Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory. *Psychological Review*, 100, 432-459.
- Cleeremans, A., Servan-Schreiber, D., & McClelland, J. L. (1995). Graded state machines: The representation of temporal contingencies in feedback networks. In Y. Chauvin & D. E. Rumelhart (Eds.), *Backpropagation: Theory, architectures, and applications* (p. 273-312). New Jersey, USA: Lawrence Erlbaum Associate.
- Damasio, A. R., & Damasio, H. (1992). Brain and language. *Scientific American*, 267, 88-95.
- Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14, 179-211.

- Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Re-thinking innateness: A connectionist perspective on development*. MIT Press.
- Holyoak, K. J., & Hummel, J. E. (2000). The proper treatment of symbols in a connectionist architecture. In E. Dietrich & A. B. Markman (Eds.), *Cognitive dynamics: Conceptual and representational change in humans and machines* (p. 229-263). Mahwah, NJ: LEA.
- Maass, W., & Orponen, P. (1998). On the effect of analog noise in discrete-time analog computations. *Neural Computation*, 10, 1071-1095.
- Majani, E., Erlanson, R., & Abu-Mostafa, Y. (1989). On the k-winners-take-all network. In D. S. Touretzky (Ed.), *Advances in neural information processing systems* (Vol. 1, pp. 634-642). San Mateo, CA: Morgan Kaufmann.
- McClelland, J. L. (1993). Towards a theory of information processes in graded, random, interactive networks. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence and cognitive neuroscience* (pp. 655-688). Cambridge, MA, USA: MIT Press.
- Minsky, M., & Papert, S. (1988). *Perceptrons* (Expanded Edition ed.; 中. 馨 & 坂. 豊, Trans.). Cambridge, MA: MIT Press. (パーセプトロン, 1993, パーソナルメディア)
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85, 59-108.
- Ratcliff, R. (1981). A theory of order relations in perceptual matching. *Psychological Review*, 88, 552-572.
- Ratcliff, R. (1988). Continuous versus discrete information processing: Modeling the accumulation of partial information. *Psychological Review*, 95, 238-255.
- Rodrigues, P., Wiles, J., & Elman, J. L. (1999). A recurrent network that learns to count. *Connection Science*, 11, 5-40.
- Roe, R. M., Busemeyer, J. R., & Townsend, J. T. (2001). Multialternative decision field theory: A dynamic connectionist model of decision making. *Psychological Review*, 108, 370-392.
- Siegelmann, H. T. (1999). *Neural networks and analog computation: Beyond the turing limit*. Boston, MA, USA: Birkhauser. (21-Oct.2001)
- Smith, P. L. (1995). Psychophysically principled models of visual simple reaction time. *Psychological Review*, 102, 567-593.
- Townsend, T. T., & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes*. Cambridge: Cambridge University Press.
- Usher, M., & McClelland, J. L. (2001). On the time course of perceptual choice: The leaky competing accumulator model. *Psychological Review*, 108, 550-592.
- Weckerly, J., & Elman, J. L. (1992). A PDP approach to processing center-embedded sentences. *Proceedings of the 14th Annual Conference of the Cognitive Science Society*.
- Williams, R. J., & Zipser, D. (1989). A learning algorithm for continually running fully recurrent neural networks. *Neural Computation*, 1, 270-280.
- Williams, R. J., & Zipser, D. (1995). Gradient-based learning algorithms for recurrent networks and their computational complexity. In Y. Chauvin & D. E. Rumelhart (Eds.), *Backpropagation: Theory, architectures, and applications* (pp. 434-486). New Jersey: Lawrence Erlbaum Associate.
- 吉則, 上. (1993). **ニューロコンピューティングの数学的基礎**. 東京: 近代数学社.
- 伸一, 浅. (2001). ニューラルネットワークの数理的基礎. In 守. 一雄, 都. 誉史, & 楠. 孝 (Eds.), **コネクションストモデルと心理学** (pp. 166-203). 京都: 北大路書房.
- 皓司, 笠. (1995). **新微分方程式対話 [新版]**. 東京: 日本評論社.