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

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，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 Elman(1990). 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, priming, higher–order descriminant analysis, 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, priming, associative memory

一般に任意の有限状態機械 (finete 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 (上坂, 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，人間の高次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 Elman(1990) の考案したニューラルネットワークの概念図を示した。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

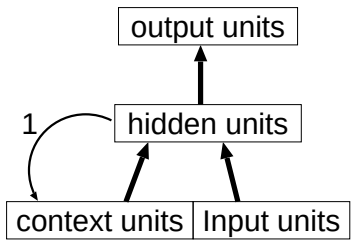

図 1: Elman の単純再帰型ニューラルネットワークの概略図

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ワ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 & Damper, 1999)。良く知られているとおり 3 層のパーセプトロンが 2 値論理関数 (AND, OR, XOR 等) を実現可能である (Minsky & 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 & 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 & 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-Schreiver, 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 & 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) などの学習方法が採用されることになる (Williams & Zipser, 1989; Williams, & Zipser, 1995)。

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，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 $\boldsymbol{y}$ で表し，文脈層から中間層へ結合係数行列を $\boldsymbol{W}$ 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 $\boldsymbol{W}$ 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 $\boldsymbol{x}(t)$ と表すことにすれば

$$\boldsymbol{y}(t) = f\left(\boldsymbol{W}\boldsymbol{y}(t-1) + \boldsymbol{x}(t)\right), \tag{1}$$

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 $\boldsymbol{W}\boldsymbol{y}(t-1) + \boldsymbol{x}(t)$ 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 $\boldsymbol{x}(t = 0, 1, \cdots, t)$ 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 $y(t)$ は以下のように表現できる (図 2)。

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

すなわち SRN の中間層の状態は過去の入力履歴 $\boldsymbol{x}(t)$ と初期値 $\boldsymbol{y}(0)$ に依存

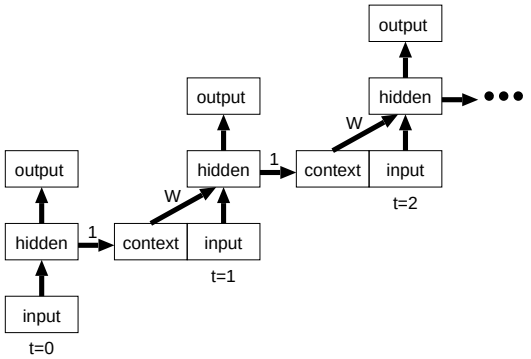

図 2: Elman 型 SRN の時間発展

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 $\boldsymbol{y}(0)$ と結合係数行列 $\boldsymbol{W}$ の構造によって定まる。文脈層からの影響を表す $\boldsymbol{W}$ 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。

しかし，上記の理由から SRN が安定な平衡点を持つための必要十分条件は $\boldsymbol{W}$ 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 $\boldsymbol{W}$ 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，

$$\boldsymbol{W} = \begin{bmatrix} 0 & -0.25 \\ -0.25 & 0 \end{bmatrix} \tag{3}$$

であれば $\boldsymbol{W}$ 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。

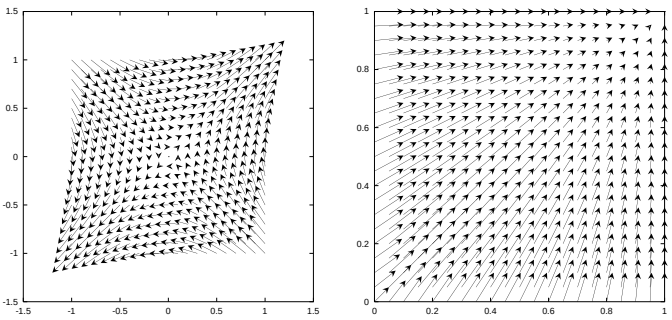

図 3: 中間層の遷移状態を表すベクトル場。線形ユニットの場合 (左) と左と同じ結合係数行列を持つシグモイド関数による非線形変換の違い

1.2 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 $x(t)$ が任意の時間感覚で一定であると仮定し，時間間隔 $t = 1, 2, \cdots$ 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

$$\tau \frac{d\,\boldsymbol{y}(t)}{dt} = f\left(\boldsymbol{W}\boldsymbol{y}(t) + \boldsymbol{x}(t)\right) \tag{4}$$

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

$$\tau \frac{d\,\boldsymbol{y}(t)}{dt} = \boldsymbol{W}\boldsymbol{y}(t) + \boldsymbol{x}(t) \tag{5}$$

となって解析が可能になる。解は

$$\boldsymbol{y}(t) = \boldsymbol{y}_0 e^{-\boldsymbol{W}(t-t_0)} + \int_{t_0}^{t} e^{\boldsymbol{W}(t-t_0)} e^{-\boldsymbol{W}(u-t_0)} \boldsymbol{x}(u) du \tag{6}$$

$$= \boldsymbol{y}_0 e^{-\boldsymbol{W}(t-t_0)} + \int_{t_0}^{t} e^{\boldsymbol{W}(t-u)} \boldsymbol{x}(u) du \tag{7}$$

となる。この式は SRN が初期値 $\boldsymbol{y}_0$ 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

$$|f(\boldsymbol{y}_1) - f(\boldsymbol{y}_2)| \leq N w_{\max} |\boldsymbol{y}_1 - \boldsymbol{y}_2| \tag{8}$$

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 $\boldsymbol{W}$ の要素 w_{ij} の中で最大のものとする。

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

$$|\boldsymbol{W} - \lambda \boldsymbol{x}| = \boldsymbol{0} \tag{9}$$

のすべての根が負の実数部を持つことである。ここでもやはり，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 $\boldsymbol{W}(t)$ 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 $\boldsymbol{W}$ の固有値の実部が 0 であれば，中間層の繊維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。文脈層から中間層へ

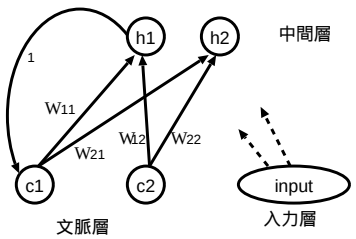

図 4: 2 ユニットだけの単純な SRN

の結合係数 $\boldsymbol{W}(t)$ が

$$\begin{pmatrix} w_{11} + x_1(t) & w_{12} \\ w_{21} & w_{22} + x_2(t) \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \tag{10}$$

のようになっていたとする。このとき $\boldsymbol{W}(t)$ 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が加わったときの状態を示している。入力系列を再生する能力が $n = 7 \pm 2$ 程度 (すなわち $2\pi/7$ だ

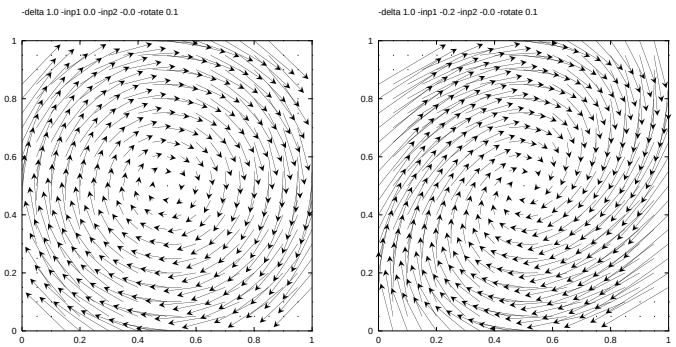

図 5: 線形回転行列

け回転する行列 $\boldsymbol{W}$ 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。

このような円周軌道上の一点として内部状態を定義した場合，特定の領域が特定の記憶項目と結び付いていると考えるのならば，SRN の中間層から出力層への変換は，中間層の円周軌道上を横切る直線として与えられる (図 6)。中間層の動作を示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かならず可能なので，任意の記憶項目と中間層との円周軌道とを同一視することが可能になるからである。

記憶における初頭効果や接近効果は，入力信号によって軌道が楕円になることによって想起しやすい項目と想起しにくい項目とができることによると考えられる。

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。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

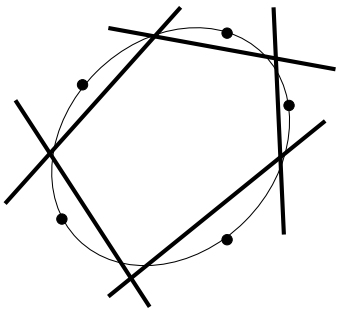

図 6: 中間層から出力層への写像は線形分離可能なニューロンの動作方程式で記述できる

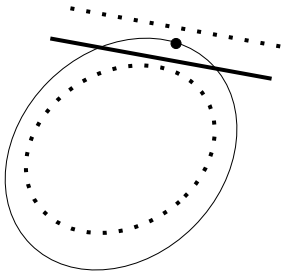

図 7: 語想起，記憶の誤り，を示す模式図

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名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生起すると仮定すれば，SRN は反応時間のモデルとなり得る。例えば

$$\boldsymbol{W} = \begin{bmatrix} 5 & -6 \\ -6 & 5 \end{bmatrix} \tag{11}$$

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 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

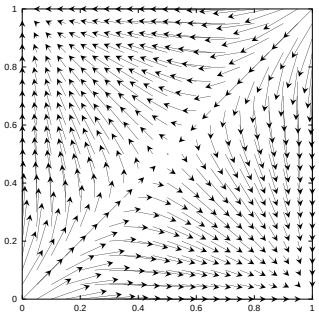

図 8: 3 つの平衡点をもつ SRN のベクトル場。2 つの平衡点は安定だが，他の 1 つは不安定である。

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り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り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 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で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 で

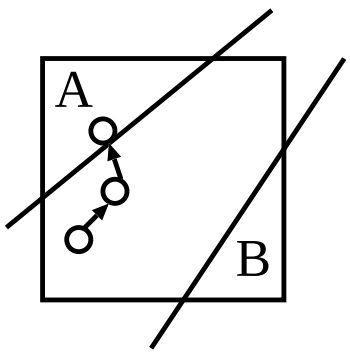

図 9: SRN を反応時間のモデルとして捉えた場合の概念図。2 肢選択反応時間の場合 A, B いずれかの領域に入るまでの繰り返し数が反応時間とみなされる。

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 $\boldsymbol{W}$ 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 $\boldsymbol{W}$ 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，$\boldsymbol{W}$ 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 & 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)。

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 モデル (Usher & McClelland, 2001) と同じ枠組と捉えることができる。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，もっとも速く規準に達したユニットが overt な反応を決定するモデルである。LCAM モデルの特徴である側抑制と再帰興奮は SRN の $\boldsymbol{W}$ 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。

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 $\boldsymbol{W}$ 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 10 はこのことを示している。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 10 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 10 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

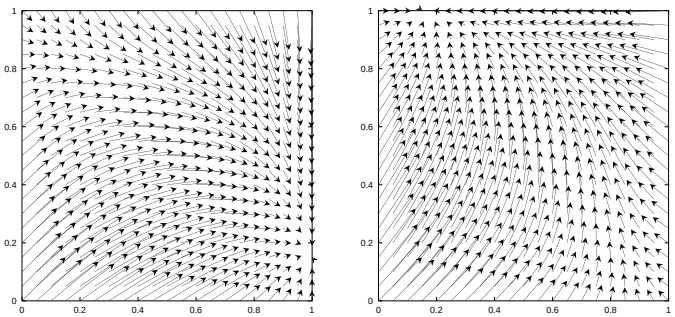

図 10: 選択的注意モデルとしての SRN。入力信号のわずかな違いによって平衡点が異なる

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 10 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 $0 < w_{ii} < 1$ として徐々に減衰する項を含めれば良い。このとき w_{ii} の値が大きければ大きい程急速に減衰するので，最初に注意を向けた対象から注意をそらし，別の対象に注意を向けるメカニズムとなり得るだろう。

2.4 プライミングのモデルとしての SRN

Cleeremans, Servan-Schreiver, and McClelland (1995) によれば SRN が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。

プライミングのモデルとしての 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活性値の強さをプライミング量と直接関連すると捉えればよい。例えば図 12 のように異なる 2 つの平衡点を持つ SRN を構成す

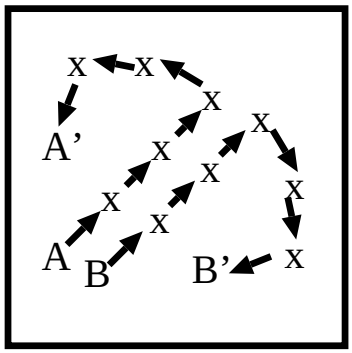

図 11: プライミングモデルとしての SRN。任意の長さの系列を判別することが可能である

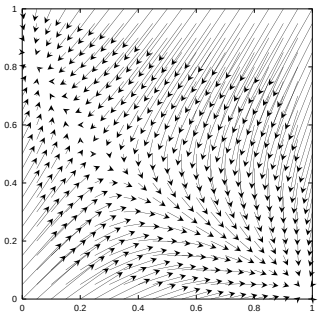

図 12: 2 つの異なる平衡点を持つ SRN の状態遷移図

ることが可能なので、類似度の高いプライム刺激を与えても、プライムミング量の異なる SRN を作ることが可能である。

また、出力層ユニットによる中間層の状態の領域分割に滑らかなシグモイド関数を仮定すれば、プライミングの程度を $[0, 1]$ の区間の実数として表現可能であろう。

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 $\boldsymbol{W}(t)$ の固有値によって定まる。$\boldsymbol{W}(t)$ 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 $\boldsymbol{W}(t)$ 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 13 は複雑な状態遷移の

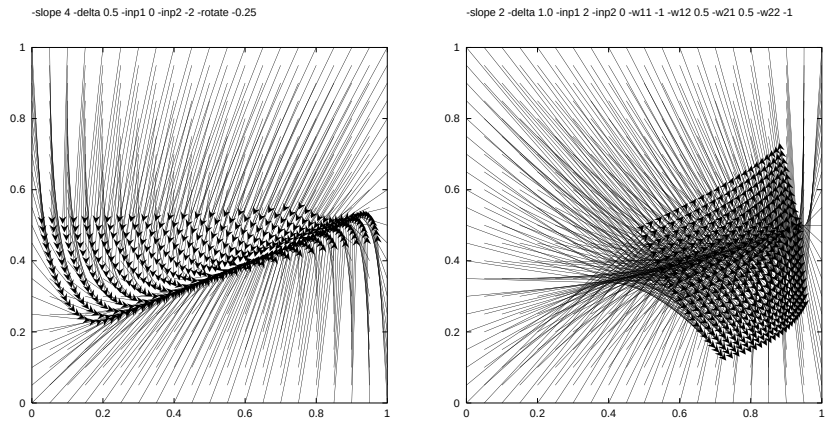

図 13: 複雑な状態遷移を示す例

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。

2.6 高次判別分析のモデルとしての SRN

文脈層から中間層への結合係数 w_{ij} が対称であるとき，すなわち $w_{ji} = w_{ij}$ という条件を満たし，時刻 $t = 0, 1, 2$ の 3 時刻で入力 $\boldsymbol{x}$ が定常だと仮定する。このとき出力 $\boldsymbol{y}$ の線形近似として

$$\boldsymbol{y} = \boldsymbol{x}^t \boldsymbol{W}^2 \boldsymbol{x} + \boldsymbol{x} \tag{12}$$

となって，2 次の判別分析と同じ能力を有する。(12) 式の右辺第 1 項は行列の 2 次形式に一致するため 2 次の判別関数とみなすことができる。

このことの重要性は，Minsky と Papert(1988) が指摘したパーセプトロンの限界を SRN は突破しているということにある。パーセプトロンをパターン分類機とみなした場合、4 層が必要であるとされている (Bishop, 1995)。SRN をパターン分類機とみなした場合，2 次以上の高次判別分析と同等の能力を有するため，フィードフォワード型パーセプトロン以上の。柔軟性をもつ判別機械であるということができる．

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, Servan-Schreiber & McClelland, 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 $\boldsymbol{W}$ 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　すくなくとも現在の時間的，空間的解像度を用いた機能的脳画像研究 (PET, fMRI, MEG) では，ある領野がある認知課題に関連することを示すことができても，その領野の機能を特定することは難しい。本稿で列挙したとおり w_{ij} の大きさによって異なる認知機能が実現できることから w_{ij} の大きさを細かく測定できる非侵襲的方法が開発されない限り，その領野の機能を調べることは，現在の機能的脳画像研究では不可能である。逆説的な言い方をすれば，機能的脳画像研究では機能が測れないことを意味している。

4　結語

本稿ではSRNの高い能力によって複数の心理現象が一つのこのニューラルネットワークモデルの枠組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。実際のデータへのあてはめなど未解決の問題も存在するが、多様な心理現象を統一した枠組で記述できることは大きな魅力であろう。

文献

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.

Bishop, C. (1995). *Neural Networks for Pattern Recongnition*. Oxford University Press.

Cleeremans, A., Servan-Schreiber, D. & McClelland, J. L. (1995). Graded state machines: The representaion of temoral contingencies in feedback networks. In Y. Chauvin & D. E. Rumelhart (Eds.), *Backpropagation: Theory, Architectures, and Applications* chapter 9, (pp. 273–312). New Jersey: 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-Simth, A., Parisi, D. & Plunkett, K. (1996). *Rethinking Innateness: A connectionist perspective on development*. MIT Press.

Maass, W. & Orponen, P. (1998). On the effect of analog noise in discrete–time analog computations. *Neural Computation*, *10*, 1071–1095.

McClelland, J. L. (1993). Towards a theory of information processes in graded, random, interactive networks. In D. E. Meyer & S. Kornblum (Eds.), *Attention & Performance XIV: Synergies in experimental psychology, artificial intelligence and cognitive neuroscience* chapter 27, (pp. 655–688). Cambridge, MA: MIT Press.

Minsky, M. & Papert, S. (1988). *Perceptrons, Expanded Edition* (2 Ed.). Cambridge, MA: MIT Press.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59–108.

Ratcliff, R. (1981). A theory of order relations inperceptual 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(1)*, 5–40.

Siegelmann, H. T. (1999). *Neural Netwroks and Analog Computation: Beyond the Turing Limit*. Birkhauser.

Townsend, T. T. & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes*. Cambridge: Cambridge University Press.

上坂, 吉則. (1993). ニューロコンピューティングの数学的基礎. 近代数学社.

Usher, M. & McClelland, J. L. (2001). On the time course of perceptual choice: The leaky competing accumulator model. *Psychological Review*, *108(3)*, 550–592.

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* chapter 13, (pp. 434–486). New Jersey: Lawrence Erlbaum Associate.

Williams, R. J. & Zipser, D. (1989). A learning algorithm for continually running fully recurrent neural networks. *Neural Computation*, *1(2)*, 270.