

4M-6

リカレントネットを用いたオンライン文字認識システム

糟谷 勇児<sup>†</sup> 山名 早人<sup>†</sup>

早稲田大学理工学部情報学科<sup>†</sup>

1. はじめに

近年、リカレントネットに関する様々な研究が行われている[3]。しかし多くの研究ではリカレントネットを用いて行うタスクが単純すぎたり、逆に専門分野に特化しすぎていたりして、リカレントネットの応用に関する一般的な知識が得られているとは言いがたい。また多くのリカレントネットの応用研究では使用しているリカレントネットは層状のものであり、全結合リカレントネットに関しては特に情報が得られていない。全結合リカレントネットは層状リカレントネットより一般的な概念であるので、より実際の脳に則した応用が可能となる。そこで本研究では全結合を用いたリカレントネットを筆跡からの文字認識（オンライン文字認識）に用いることで、全結合リカレントネットの応用上の注意点や応用可能性を探ることを目的とする。

2. リカレントネットとは

リカレントネットとはフィードバック結合を持つニューラルネットである。リカレントネットではこのフィードバック結合の効果により、フィードバックされた過去の情報と現在の情報を統合した処理、すなわち時系列処理を行うことが可能である。リカレントネットには層状のものと同結合を持つものがある。ただし本研究では全結合のリカレントネットを用いることとする。全結合リカレントネットの動作と学習は以下の式で表される[2]。

入力層ニューロン

$$y_i(t) = input_i(t) \tag{1}$$

隠れ層・出力層ニューロン

$$y_i(t) = f(s_i(t)) \tag{2}$$

$$s_i(t+1) = \sum_{k=1}^n w_{ik} y_k(t) \tag{3}$$

$$f(x) = (1 + e^{-x})^{-1} \tag{4}$$

ここで  $y_i(t)$  は  $i$  番目ニューロンの出力、 $s_i(t)$  は時刻  $t$  で  $i$  番目のニューロンの内部状態、 $input_i(t)$  は時刻  $t$  で  $i$  番目のニューロンへの外部入力、 $w_{ik}$  は  $k$  番目のニューロンから  $i$  番目のニューロンへの結合重み、 $f$  は出力関数とする。また図 1 に全結合リカレントネットの概念図を示す。図 1 で、入力層のニューロンは外部入力を受けるニューロンであり、出力層のニューロンは出力を観測するニューロンである。隠れ層のニューロンには外部との直接の接続を持たない。

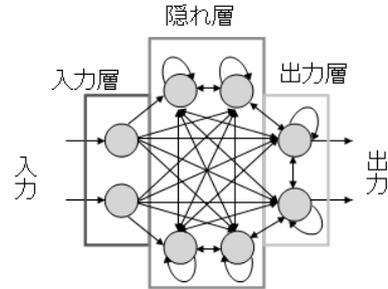


図 1 全結合リカレントネット

全結合リカレントネットはバックプロパゲーションを拡張した学習法である BPTT (Back Propagation Through Time) を用いて学習を行うことが出来る。BPTT はバックプロパゲーション同様、サンプルとそのサンプルを入力したときの理想の出力（教師信号）から学習を行う。BPTT では(5)式で表される二乗誤差関数  $E$  を勾配法で小さくすることにより学習を行う[2]。

$$E = \frac{1}{2} \sum_{t=1}^T \sum_{k=1}^n \mu_k(t) [y_k(t) - d_k(t)]^2 \tag{5}$$

ここで  $T$  は時刻の最大値であり、 $d_k(t)$  は時刻  $t$  における  $k$  番目のニューロンの教師信号、 $\mu_k(t)$  はマスク関数であり、時刻  $t$  における  $k$  番目のニューロンの出力を誤差関数にどのように反映させるかを決定する関数である。例えば  $\mu_{10}(5) = 0$  ならば 10 番目のニューロンの  $t=5$  での出力は誤差に影響を与えない。BPTT では以下の式により結合重みの更新を行う[2]。

$$\Delta w_{ij} = -\alpha \sum_{t=1}^T \frac{\partial E}{\partial s_i(t)} y_j(t-1) \tag{6}$$

ここで  $\Delta w_{ij}$  は結合重み  $w_{ij}$  の更新量、 $\alpha$  は学習係数である。BPTT の詳しいアルゴリズムについては紙面の関係から説明を省略する。詳しくは[1][2]等を参照されたい。

3. 文字認識システムの概要

リカレントネットを用いた文字認識システムでは以下の①～⑥の手順で筆跡から文字を認識する。①～⑥の記号は図 2 に示すシステムのプロトタイプ中の記号と一致している。

- ①ユーザーがウインドウに文字を書く
- ②ユーザーの文字の筆跡（ペンの位置座標）を短い時間間隔でサンプリングする。このサンプリングされた時刻  $t$  における座標ベクトルを  $\mathbf{x}(t)$ 、また筆記を開始した時刻を 0、終了した時刻を  $T$  とする。認識にはこの  $\mathbf{x}(t)$  から時間的に等間隔に 10 点の座標  $\{\mathbf{x}(0), \mathbf{x}(T/9), \mathbf{x}(2T/9), \dots, \mathbf{x}(T)\}$  を用いる。
- ③10 点の座標データを以下の式で絶対値を 1 に正規化した速度ベクトル  $\bar{\mathbf{v}}(k)$  に変形する

On-line handwriting character recognition with fully connected recurrent neural networks

<sup>†</sup>Yuji KASUYA, <sup>†</sup>Hayato YAMANA

<sup>†</sup>Department of Information and Computer Science, Science and Engineering, Waseda University

$$\mathbf{v}(k) = \mathbf{x}(Tk/9) - \mathbf{x}(T(k-1)/9) \quad (k=1, 2, \dots, 9) \quad (7)$$

$$\bar{\mathbf{v}}(k) = \mathbf{v}(k) / |\mathbf{v}(k)| \quad (8)$$

ただし  $|\mathbf{v}(k)| = 0$  のときは  $\bar{\mathbf{v}}(k) = \mathbf{0}$  である。

- ④リカレントネットには 2 個の入力層ニューロンを用意し、それぞれには  $\bar{\mathbf{v}}(k)$  の x 成分、y 成分を入力する。
- ⑤出力層には n 個のニューロンを用意する。n 個のニューロンの内、二つのニューロンの出力値を 1 にし、出力が 1 となる二つのニューロンの組み合わせで文字を判別する（例えば (1, 1, 0, 0, 0) は文字 1、(0, 1, 1, 0, 0) は文字 2 などとする）。
- ⑥システムは出力が最も 1 に近い 2 つのニューロンの組み合わせから、どの文字が入力されたかを判断し、ユーザーに通知する

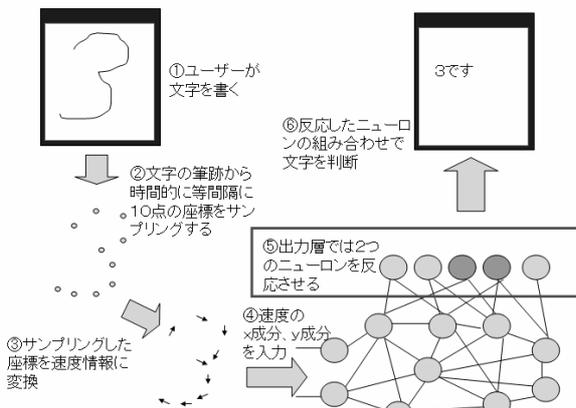


図2 文字認識システムの概要

座標データを④で速度データに変換することで、文字の位置や大きさに依存しない認識が可能となる。また⑤の出力方式により  ${}_n C_2$  通りの文字表現が可能である。

#### 4. 実験

本章では 3 章で提案した文字認識システムに関して行った実験について述べる。実験では文字認識システムに 0~9 までの数字を学習させ、認識率を調べた。学習には筆者および山名研の学生 5 人から採取した文字サンプル各文字 20 個ずつと、東京農工大学中川研究室オンライン手書きデータベース「TUAT Nakagawa Lab. HANDS-kuchibue\_d-97-06-10」から 4 人分各文字 10 個ずつを使用した。テストサンプルとしては、学習サンプルを採取した学生とは異なる学生 5 人から採取したデータ 50 文字（以下研究室データと呼ぶ）と、中川研究室オンライン手書きデータベースから学習に用いていない 6 人分 300 文字（以下データベースと呼ぶ）を使用した。

リカレントネットの学習法はとして BPTT の他に RTRL[2] (Real Time Recurrent learning) があるが、学習時間が BPTT と比較して数百倍かかり、今回の適用には不向きであることが分かった。そこでリカレントネットの学習には BPTT を用いたが、BPTT に慣性項を用いることで学習を高速化できることが分かった（最大 3 倍）。バックプロパゲーションに慣性項を用いる方法については[3]が詳しい。これは BPTT にも応用可能である。またマスク関数  $\mu_k(t)$  はニューロンによらず

$$\mu(t) = \begin{cases} 0 & 1 \leq t < T \\ 1 & t = T \end{cases} \quad (9)$$

と設定するとき最も認識率がよいことがわかった。よって本研究ではこれらの結果を使用した。

以下の表 1 に本システムの実験結果を示す。ただしこの実験では学習係数  $\alpha=0.01$ 、学習回数を 1000 回とした。また出力層のニューロンは 5 個とした。プログラムは Java で作成し、Celeron1.3GHz の環境で時間を測定した。

表 1 から研究室データで最大 94%、データベースで最大 91% の認識率であることが分かる。また学習には、隠れ層ニューロンの数が 40 個で約 4 分半、50 個で約 6 分の時間を要した。

この結果と比較するためにパーセプトロンを用いて同じタスクを行った。実験ではパーセプトロンのニューロン数を 20, 30, 40, 50 と変えて認識率を測定した。この結果最大の認識率はニューロン数 20 個のときのデータベースの認識率 61% であった。またリカレントネットを認識する文字の数だけ用意し、複数並列に用いて同様なタスクを行った結果、認識率は研究室データで最大 98%、データベースで 94% と向上した。ただし学習は 1 台の PC で行ったため、学習時間は約 10 分と増加した。

表 1 隠れ層ニューロンの数に対する認識率

隠れ層ニューロンの数	研究室データの認識率[%]		データベースの認識率[%]	
	平均	最大	平均	最大
30	83	86	85	88
35	87	90	85	88
40	88	94	84	88
45	88	94	85	90
50	86	88	89	91

#### 5. まとめ

本稿では全結合リカレントネットを用いたオンライン文字認識システムについて報告した。認識率は最大 91% を示し、全結合リカレントネットの応用可能性を示した。この結果はパーセプトロンの認識率を大きく上回り、時系列認識において全結合リカレントネットは有効である。今回の実験の結果、慣性項を用いた BPTT の有用性や、マスク関数の設定法、その他種々のパラメータの設定例や学習時間などが分かった。またリカレントネットを複数並列に用いることで認識性能を向上させることが出来ることが分かった。これらの情報に注意して全結合リカレントネットを用いることにより、より高度な応用が行いやすくなると思われる。

#### 参考文献

- [1] Barak A Pearlmuter : Gradient Calculations for Dynamic Recurrent Neural Networks : A Survey, IEEE Trans on Neural Networks ,Vol 6 , Issue 5 , pp1212-1228(1995)
- [2] M. A. Arbib : The Handbook of Brain Theory and Neural Networks. MIT Press(1995)
- [3] 電気学会 GA ニューロを用いた学習法とその応用調査専門委員会 : 学習とそのアルゴリズム—ニューラルネットワーク・遺伝アルゴリズム・強化学習, 森北出版(2002)