

回転文字認識のためのニューラルネットワークの構築

篠沢 佳久^{1,a)}

受付日 2013年9月30日, 採録日 2014年7月11日

概要: 本論文においては, 回転文字を対象としたオフライン手書き文字認識ニューラルネットワークの構築を試みる. 回転文字の認識は, 現在においても困難な問題の1つとされている. 本論文では, 数千種類のオフライン手書き文字を認識対象として, 回転文字の認識方法を構築することを目的とする. 回転文字の認識を実現するため, 一定の角度ごとに回転した文字を用意し, これらの文字の認識が可能な識別器を構築する手法が提案されている. 本論文においては, このような識別器を, ニューラルネットワークを用いて構築するため, その構造および学習方法を提案する. しかし, 回転文字を認識した場合, 回転していない文字と比較して, 認識精度の低下が予想される. そこで本論文においては, 回転した未知文字を認識するだけでなく, さらにその文字を任意の角度回転させて認識を行い, それぞれの結果を統合することによって最終的な認識結果を求める方法を提案する. そして産業技術総合研究所提供の文字画像データベース ETL9B, 3,036 字種を認識対象として, ニューラルネットワークを構築した結果, 文字種の多い場合においても, 回転文字の学習が可能であること, 文字を回転させて認識を行い, 認識結果を統合することによって, 認識率の向上が図れたことを示す.

キーワード: オフライン手書き文字認識, 回転, ニューラルネットワーク

Neural Networks for Rotated Character Recognition

YOSHIHISA SHINOZAWA^{1,a)}

Received: September 30, 2013, Accepted: July 11, 2014

Abstract: The recognition of rotated character patterns is thought to be one of the more difficult problems in the field of pattern recognition. It is expected that the recognition rate for rotated character patterns is low. We propose the construction of neural networks for offline handwritten character patterns that are rotated. Our method consists of two recognition processes. First, the proposed neural networks recognize an unknown character pattern. Then, the unknown character pattern is rotated by a certain number of degrees, and the neural networks recognize the rotated character pattern. After combining the candidates from each result, the final recognition results are obtained. We construct neural networks for the character image database ETL9B. The results obtained from our experiments show that the proposed neural networks can learn many types of rotated character patterns and improve the recognition rates.

Keywords: offline handwritten character recognition, rotation, neural networks

1. はじめに

手書き文字認識の研究は古くから行われており, 今日に至るまでさまざまな認識方法が考案されている. 字体が書き手によって異なり, 筆順情報を利用できないオフライン手書き文字認識の研究は, パターン認識の分野において困

難な課題の1つである. その要因として, 文字固有の問題(認識対象数や類似文字の多さ)だけでなく, 書き手によって異なる字体の問題(変形や回転)があげられる. 特に文字が多少傾くことによって, 誤認識となる場合もあり, 回転文字の認識は現在においても困難な問題である.

そのため回転文字の認識については, これまで多くの研究 [1], [2] がされてきた. 手法としては, 前処理において回転を補正した後, 認識を行う方法 [3], [4], 回転を補正せずに認識を行う方法に大別される. 前者は, 前処理について

¹ 慶應義塾大学
Keio University, Yokohama, Kanagawa 223-8522, Japan
^{a)} shino@ae.keio.ac.jp

の研究領域となる。後者については、さらに回転に頑強な特徴を抽出し、認識を行う手法 [5], [6], 回転文字の認識が可能な識別器を構築する手法 [7], [8], [9], [10], [11] に分けられる。

しかしながら、これらの研究は、数字、アルファベットや平仮名もしくは印刷文字など認識対象となる文字種が限定されている。そこで本論文においては、数千種類の手書き文字を対象として、回転文字認識の手法を構築することを目的とする。特に文字種が多い場合、前処理による傾きの補正や回転に不変な特徴の抽出が困難と考え、回転文字の認識が可能な識別器を構築するアプローチをとる。これについて先行研究 [11] においては、一定の角度ごとに回転した文字を学習サンプルとして用意し、これらの回転文字の認識が可能な識別器を構築することを試みている。

しかし、認識対象数が多く、一定の角度ごとに回転した文字を認識することを考えた場合、識別器は大規模化するため、高い識別精度を持ち、かつ構築が容易な識別器を用いる必要がある。そうした識別器の1つとして、ニューラルネットワークがあげられる。これまでニューラルネットワークによって文字認識を行う試みがされてきた [12], [13], [14], [15], [16]。ニューラルネットワークを用いる利点として、学習サンプルを適切に提示することによって、異なるカテゴリーの判別が可能な識別器を構築できることがあげられる。そのため数千種類の文字種を対象としたニューラルネットワークが提案され、高い識別精度が得られることが示されている [14], [15], [16]。特にこれらの先行研究においては、単一のニューラルネットワークを用いて、文字の認識を行っているのではなく、複数のニューラルネットワークを利用することによって対処している。しかしながら、文字種が多い場合、回転文字を対象としたニューラルネットワークの構築方法を提案した研究事例はない。

そこで本論文においては、文字種ごとに個別のニューラルネットワークを用いた先行研究 [15] の考えに基づき、回転文字認識への対応を試みる。これは、任意の角度 (θ 度) ずつ回転した文字の認識が可能なニューラルネットワークを文字種別に構築する。独立したニューラルネットワークを文字種分、構築することによって、文字種が多くなった場合においても、任意の角度回転した文字の認識が可能となる。しかし、回転文字を認識した場合、回転していない文字と比較して認識精度の低下が予想される。先行研究 [11] においては、文字の回転角を推定した場合、20度前後の推定誤りが生じる場合があることが報告されている。

そのため本論文においては、回転文字を認識する場合、さらにその文字を θ 度回転させて認識を行い、それぞれの認識結果を統合することによって最終的な結果を求める手法を提案する。すなわち、文字をさらに θ 度回転させ認識した場合においても、回転する前と同じ認識結果になるも

のと予想されるからである。また誤認識した場合、 θ 度回転し、再認識することによって、誤認識を訂正できる場合もあるものとする。

以上、本論文においては、一定の角度ごとに回転した文字の認識が可能なニューラルネットワークの構造および学習方法を提案する。認識時においては、複数回、文字を回転させて認識を行い、それぞれの認識結果を統合することによって最終的な認識結果を求める手法を提案し、認識精度の向上を図る。また認識対象の回転角が異なるニューラルネットワークを、それぞれ大分類、詳細認識に分けて適用することにより、計算量を抑制しながら認識精度の向上を試みる。そして産業技術総合研究所提供の文字画像データベース ETL9B, 3,036 字種を対象とした認識実験を通して提案手法の有効性を評価する。

2. 提案手法

2.1 ニューラルネットワークの構造と動作

まず提案するニューラルネットワークの構造について述べる。先行研究 [11] においては、一定の角度ごとに回転した文字を用意し、それぞれを対象とした識別器を構築することを試みている。本論文においては、これをニューラルネットワーク (以下、ネットワークと略す) に応用することを試みる。認識対象の文字種が P 個の場合は、 P 個のネットワークを用意する。各ネットワークと任意の1字種との対応づけを行い、特定のネットワークはあらかじめ定められた1字種のみ認識処理を行う。

各ネットワークの構造は入力層、中間層、出力層の3層から構成される階層型のフィードフォワード構造とする。入力層と中間層、中間層と出力層のニューロンとが結合している。図1に例として、「あ」を認識対象としたネットワークの構造を示す。出力層のニューロン数は N 個、中間

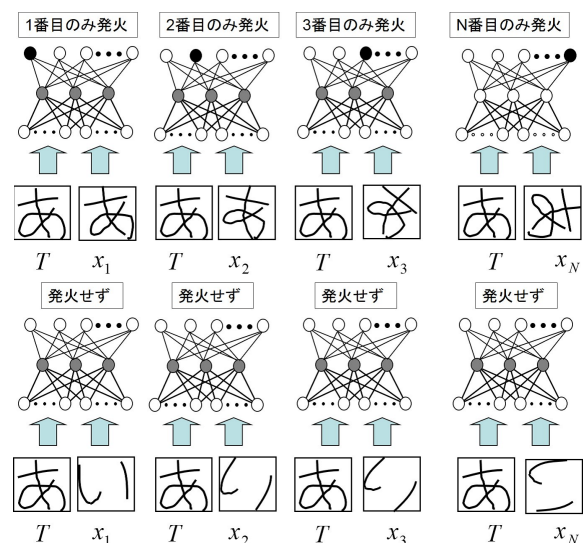


図1 提案するネットワークの構造と動作

Fig. 1 Structure of proposed neural networks.

層のニューロン数は H 個とする．入力層には，そのネットワークで処理する文字種のテンプレート（図 1 の場合は「あ」）から抽出した特徴と認識したい未知文字から抽出した特徴を同時に入力する．入力層において，テンプレートの特徴を入力する領域を比較対象部，未知文字の特徴を入力する領域を認識対象部と呼ぶ．したがって，入力層は 2 つの文字の特徴数分のニューロンが必要となる（入力層のニューロン数を J 個，文字の特徴数を $J/2$ 個とする）．比較対象部に入力するテンプレートは回転していない文字のみとする．一方で認識対象部に入力する文字は回転した場合も含める．出力層の N 個のニューロンについては，それぞれ，特定の角度回転した文字と対応づけをしておく．すなわち出力層の 1 番目のニューロンは，テンプレートと同じ傾きの文字，2 番目のニューロンは，テンプレートと比較して θ 度回転した文字，そして i ($i = 1, 2, \dots, N$) 番目のニューロンは $(i - 1) \times \theta$ 度回転した文字について処理するものと固定する．

次にネットワークの動作について述べる．文字種 p ($p = 1, 2, \dots, P$) を認識対象とするネットワークにおいて，入力層の j ($j = 1, 2, \dots, J$) 番目のニューロンと中間層の m ($m = 1, 2, \dots, H$) 番目のニューロンとの結合係数を w_{pjm} ，中間層の m 番目のニューロンと出力層の i 番目のニューロンとの結合係数を w_{pmi} とする．中間層の m 番目のニューロンの閾値を b_{hpm} ，出力値を h_{pm} ，出力層の i 番目のニューロンの閾値を b_{opi} ，出力値を o_{pi} とする．文字種 p のテンプレートの特徴 T_p と未知文字の特徴 x を入力層に入力した際，下記の式のようにネットワークを動作させる．活性化関数 f にはシグモイド関数を用いる．

$$h_{pm} = f \left(\sum_{j=1}^{\frac{J}{2}} w_{pjm} T_{pj} + \sum_{j=\frac{J}{2}+1}^J w_{pjm} x_j + b_{hpm} \right) \quad (1)$$

$$o_{pi} = f \left(\sum_{m=1}^H w_{pmi} h_{pm} + b_{opi} \right) \quad (2)$$

$$f(x) = \frac{1}{1 + \exp(-x)} \quad (3)$$

その際，未知文字 x が文字種 p で，かつテンプレートと比較して $(i - 1) \times \theta$ 度回転している場合，出力層の i 番目のニューロンのみが発火し，それ以外のニューロンは発火させないように教師信号 t_p を与える（式 (4)，図 1 上図）．

$$t_p = \begin{cases} \underbrace{(0, \dots, 0, 1, 0, \dots, 0)}_{i-1} \underbrace{(0, \dots, 0, 1, 0, \dots, 0)}_{N-i-1} & \text{文字 } p \text{ で } (i-1) \times \theta \\ \underbrace{(0, \dots, 0, 0, 0, \dots, 0)}_n & \text{文字 } p \text{ でない場合} \end{cases} \quad (4)$$

一方で文字種 p ではない場合，出力層のすべてのニューロンが発火しないように教師信号 t_p を与える（図 1 下図）．

この教師信号に従って動作するように（教師信号とネットワークの出力値の誤差二乗和 $\sum_{i=1}^N (t_{pi} - o_{pi})^2$ が最小になるように），ネットワーク内の結合係数を学習する．

このようなネットワークを構築することによって， $\theta, 2\theta, \dots, (N - 1) \times \theta$ 度回転した文字の認識が可能となる．1 つの文字種に対して，特定のネットワークによって対処させることで文字種が増加した場合においても，対応が容易となる．また回転していない文字をテンプレートとして入力すると同時に，回転文字を入力することによって，回転角を差分として学習できるものと考えた．したがって認識と同時に，回転角の推定も可能なため，傾き補正も可能となる．

2.2 未知文字パターンの認識

未知文字を認識する場合，学習後の P 個のネットワークに各テンプレート T_p と未知文字 x を入力し，各ネットワークを動作させる．そして出力層からの出力値の合計が最も大きいネットワークと対応する文字種を認識結果とする（図 2）．

しかし回転していない文字と比較して，回転文字を認識する場合，認識精度の低下が予想される．そこで未知文字をさらに θ 度回転（アフィン変換）し，この回転した文字に対しても同様に認識を行い，最終的な認識結果を求める（図 3）．これは，未知文字 x を認識した結果，ネットワーク p の出力層の第 i 番目のニューロンが最大発火したものとする（図 3 上図）．さらにこの未知文字 x を θ 度回転させた文字を x_θ とする．この回転させた文字 x_θ を認識する場合，ネットワーク p の出力層の第 $i + 1$ 番目のニューロンが最大発火するものと予想されるからである（図 3 下図）．また未知文字を誤認識した場合， θ 度回転したうえで，再認識することによって，誤認識を訂正できる場合もあるものと考えられる．

以上のように， θ 度ごとに回転した文字の認識が可能なネットワークを構築し，未知文字を認識する場合，その文字を認識させるだけでなく，さらに θ 度ごとに回転させた文字も認識させ，それぞれの認識結果を統合することによって，認識精度の向上を試みる．

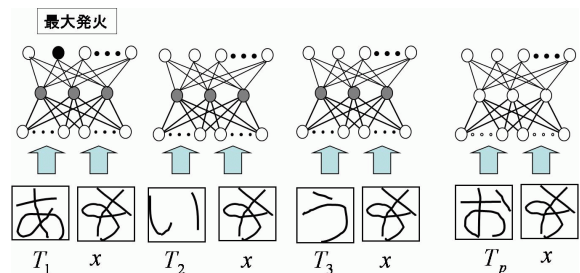


図 2 提案するネットワークでの文字パターンの認識

Fig. 2 Recognition of character patterns by proposed neural networks.

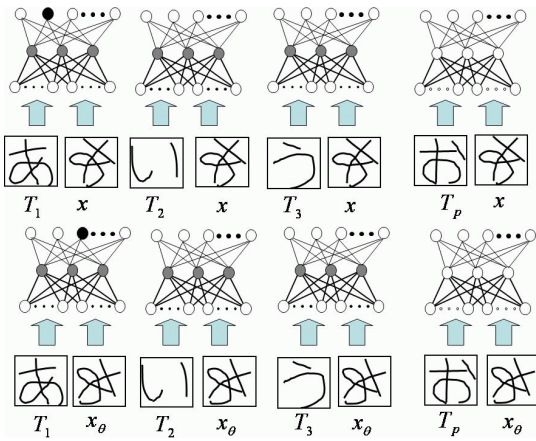


図3 回転させた文字パターンの認識

Fig. 3 Recognition of rotated character pattern.

3. 学習および認識方法

3.1 学習方法

文字種 p ($p = 1, 2, \dots, P$) を対象としたネットワーク p の学習方法を下記に示す。

- (1) 入力層の比較対象部に、そのネットワークで対象としているテンプレート T_p を入力する。
- (2) テンプレートと同字種の文字 T'_p (発火用パターンと呼ぶ) を用意する。
- (3) $i = 1$ ($i = 1, 2, \dots, N$) とする。
- (4) T'_p に対して $(i - 1) \times \theta$ 度のアフィン変換を行う。回転後の文字を T'_{pi} とする。
- (5) 認識対象部に T'_{pi} を入力する。
- (6) 出力層の i 番目のニューロンのみを発火させ、他のニューロンは発火しないように教師信号 t_p を与える。

$$t_p = (\underbrace{0, \dots, 0}_{i-1}, \underbrace{1, 0, \dots, 0}_{N-i-1})^t \quad (5)$$

- (7) ネットワークを動作させ、出力層の各ニューロン k の発火値 o_{pk} ($k = 1, 2, \dots, N$) を求める。教師信号と出力値の誤差二乗和 $\sum_{k=1}^N (t_{pk} - o_{pk})^2$ を計算し、教師信号と同じ出力をするようにネットワーク内の結合係数を学習する。学習にはバックプロパゲーションアルゴリズム (誤差逆伝播則) [17] を用いる。
- (8) $i = i + 1$ とし、 i が N 以下であれば (4) へ戻る。そうでなければ (9) の抑制用パターンの学習を行う。
- (9) テンプレートとは異なる字種の文字 T_d ($d \neq p$) (抑制用パターンと呼ぶ) を用意する。
- (10) $i = 1$ ($i = 1, 2, \dots, N$) とする。
- (11) T_d に対して $(i - 1) \times \theta$ 度のアフィン変換を行う。回転後の文字を T'_{di} とする。
- (12) 認識対象部に T'_{di} を入力する。
- (13) 出力層のすべてのニューロンを発火しないように教師信号 t_d を与える。

$$t_d = (0, 0, \dots, 0)^t \quad (6)$$

- (14) ネットワークを動作させ、出力層の各ニューロン k の発火値 o_{pk} を求める。(7) と同様に、教師信号と同じ出力をするようにネットワーク内の結合係数を学習する。
 - (15) $i = i + 1$ とし、 i が N 以下であれば (11) へ戻る。そうでなければ他の抑制用パターン T_d を用意し、(10) へ戻る。
- 以上、誤差二乗和の総和が一定値以下となるまで (2) から (15) を繰り返す。

3.2 認識方法

前節にて述べた学習済みのネットワークを P 個用意する。認識方法を下記に示す。

- (1) 各ネットワーク p ($p = 1, 2, \dots, P$) の入力層の比較対象部に、そのネットワークで対象とするテンプレート T_p を入力する。
- (2) $l = -r$ とする (r は整数値で $r \geq 0$)。
- (3) 未知文字 x に $(l \times \theta)$ 度のアフィン変換を行う。回転後の文字を x' とする。
- (4) 各ネットワークの認識対象部に x' を入力する。
- (5) 各ネットワークを動作させ、出力層の各ニューロン k の発火値 o_{plk} ($k = 1, 2, \dots, N$) を求める。
- (6) $l = l + 1$ とし、 l が r 以下であれば (3) へ戻る。
- (7) ネットワーク p ごとで出力層のニューロンの発火値の合計 O_p を求める。

$$O_p = \sum_{l=-r}^r \sum_{k=1}^N o_{plk} \quad (7)$$

- (8) 合計値 O_p が最大となる文字 p を認識結果とする。

以上のように認識時においては、 $-r \times \theta$ 度から $r \times \theta$ 度の範囲で、 θ 度ずつ未知文字を $2r$ 回、回転させて認識を行い、回転角ごとの認識結果を統合することによって最終的な結果を求める。たとえば α 度回転した文字を認識する際、 $r = 2$ とした場合は、 α 度回転した文字だけでなく、さらに $-2\theta + \alpha$ 度、 $-\theta + \alpha$ 度、 $\theta + \alpha$ 度および $2\theta + \alpha$ 度回転した文字も認識することによって、最終的な認識結果を求める。

4. 認識実験

4.1 回転文字の作成

提案手法の有効性を評価するために認識実験を行う。認識実験には産業技術総合研究所提供の ETL9B (3,036 文字) を利用した。ネットワーク構築のため学習用文字としては、1 文字あたり最初の 1 セット目の 40 個を利用した。認識実験で利用する未知文字 (評価用文字) としては、2 セット目の 40 個を利用した (大きさは 64×63)。文字の処理過程を図 4 に示す。

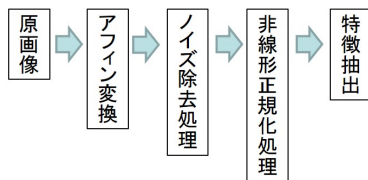


図 4 回転文字の作成

Fig. 4 Process of character patterns.

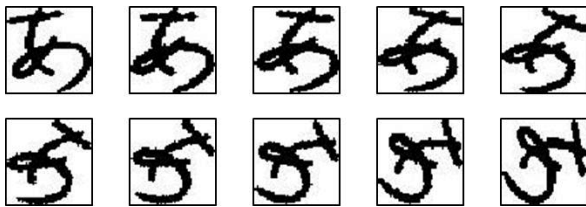


図 5 文字パターンの回転

Fig. 5 Rotation of a character pattern.

上記の原画像に対し、アフィン変換を行い、回転文字を作成する。文字の回転については時計方向に0度から350度を対象として、10度刻みとした。回転角については、-2度から2度の範囲の一樣乱数でばらつきを与える。原画像の中には、傾きのある文字も含まれていると考えたためである。たとえば10度回転させる場合は、8度から12度の範囲の値とした(0度の文字についても、-2度から2度の範囲のばらつきを与えている)。以上の過程によって0度から350度まで、10度ごとに回転した文字を作成した。図5に文字「あ」にアフィン変換を0度から90度の範囲で行った例を示す。

これらのアフィン変換後の文字に対して、フィルタによるノイズ除去処理[18]をした後、非線形正規化処理[19]を用いて48×48の大きさとした。そして正規化後の文字画像から特徴として、196次元の輪郭線特徴[20]を抽出した。学習用文字、評価用文字ともに、上記の過程によって、0度から350度まで10度刻みで回転文字を作成し、特徴を抽出する。ネットワークの学習については、学習用文字から抽出した特徴を用いて行う。一方で、ネットワークの評価については、評価用文字から抽出した特徴を用いて行う。

4.2 ニューラルネットワークの学習条件

各ネットワークについて、入力層のニューロン数は392個(特徴数の2倍必要)、中間層のニューロン数は64個、出力層のニューロン数は36個とした。ネットワークの総数は文字数と同じく3,036個である。

文字種 p を対象としたネットワークを学習する際、比較対象部に入力するテンプレートは40個の学習用文字の中からランダムに選択する。一方で、認識対象部への入力の場合、発火用パターンは文字種 p のみであるが、抑制用パターンは文字種 p 以外となる。この場合、抑制用パターンとして文字種 p 以外のすべての文字種を学習に用いるのは

効率が悪い。そこで、先行研究[15]の考えに基づき、発火用パターン(テンプレート)と類似している文字のみを抑制用パターンとして使用する。すなわち発火用パターンと類似している文字種のみを選ぶために、あらかじめ大分類を行っておき、大分類の結果からその上位候補に列挙された文字候補のみを抑制用パターンの対象として利用する。大分類のために用いる識別関数にはユークリッド距離を用いる。

たとえば未知文字として、「あ」を認識対象とする場合を考える。未知文字の特徴ベクトルを $\mathbf{a} = (a_1, a_2, \dots, a_{196})^t$ とする。一方、 θ 度回転した学習文字 p の特徴ベクトルを $\mathbf{x}_{\theta p} = (x_{\theta p1}, x_{\theta p2}, \dots, x_{\theta p196})^t$ とする($p = 1, 2, \dots, 3,036$)。これは、1字種につき40個ある学習用文字から抽出した特徴ベクトルの平均値である。2つの文字間のユークリッド距離 $d_{\theta p}$ は、式(8)から求めることができる。

$$d_{\theta p} = \sqrt{\sum_{j=1}^{196} (a_j - x_{\theta pj})^2} \quad (8)$$

そして未知文字と角度ごとにすべての学習文字とのユークリッド距離を求めた結果(最近傍法)、第1候補から順に『およなさ』というような順番の認識結果になったとする。この結果から上位候補の「お」「よ」「な」「さ」という文字は「あ」と類似しているということから、文字種「あ」を対象としたネットワークにおいて、これらを抑制用パターンとして用いる。本論文においては、大分類の結果、上位10位までに含まれる文字を抑制用パターンとして用いた。この場合、1字種あたりの抑制用パターンの個数は平均308.6個であった。したがって、抑制用パターンの学習時間については約1/10に削減することができた。各ネットワークの学習は誤差逆伝播則の改良法であるモーメント法[21]を使用した(モーメント係数は0.5)。パラメータの設定はすべての実験で同一とした。

4.3 ニューラルネットワークの学習結果

前節で述べた実験条件のもとでネットワークの学習を行った。その結果を表1、表2に示す。

表1には、各角度ごとでの認識率を学習用、評価用文字に分けて示す。この結果は、認識対象部に入力した文字を回転させずに認識した結果(3.2節の認識方法において、 $r = 0$ とした場合)を示している。そして提案手法との比較として、式(8)による最近傍法による認識結果も示す。また表2には、平均認識率、認識率のばらつき(分散)、すべての角度の中での最高認識率および最低認識率(括弧内の数字はその回転角)を示す。

表2より、学習用文字においては、提案手法の場合、平均認識率は99.33%、最近傍法の場合は91.48%となり、提案手法の方が高い結果となった。この結果より、文字種が多い場合においても、提案手法によって回転文字の学習

表 1 提案手法における認識結果

Table 1 Recognition results with proposed neural networks.

| 回転角 | 提案手法 | | 最近傍法 | |
|-----|--------|--------|--------|--------|
| | 学習用 | 評価用 | 学習用 | 評価用 |
| 0 | 99.57% | 89.24% | 94.26% | 89.61% |
| 10 | 99.49% | 87.05% | 92.76% | 87.37% |
| 20 | 99.33% | 85.67% | 91.91% | 85.91% |
| 30 | 99.20% | 84.48% | 90.49% | 83.57% |
| 40 | 99.14% | 84.04% | 89.55% | 82.43% |
| 50 | 99.10% | 84.29% | 89.83% | 82.87% |
| 60 | 99.24% | 85.19% | 90.82% | 84.32% |
| 70 | 99.38% | 86.26% | 91.95% | 85.97% |
| 80 | 99.48% | 87.02% | 92.83% | 87.14% |
| 90 | 99.57% | 88.36% | 93.16% | 88.23% |
| 100 | 99.43% | 86.80% | 92.62% | 87.05% |
| 110 | 99.31% | 85.40% | 91.67% | 85.58% |
| 120 | 99.14% | 84.05% | 90.30% | 83.15% |
| 130 | 99.07% | 83.67% | 89.30% | 81.89% |
| 140 | 99.14% | 84.30% | 89.66% | 82.48% |
| 150 | 99.26% | 85.05% | 90.65% | 83.99% |
| 160 | 99.43% | 86.02% | 91.84% | 85.72% |
| 170 | 99.47% | 86.83% | 92.87% | 87.11% |
| 180 | 99.56% | 88.53% | 93.28% | 88.36% |
| 190 | 99.47% | 87.07% | 92.56% | 87.17% |
| 200 | 99.37% | 85.72% | 91.71% | 85.63% |
| 210 | 99.19% | 84.33% | 90.33% | 83.23% |
| 220 | 99.11% | 83.80% | 89.53% | 82.14% |
| 230 | 99.15% | 84.21% | 89.87% | 82.79% |
| 240 | 99.27% | 85.04% | 91.04% | 84.44% |
| 250 | 99.41% | 86.01% | 92.29% | 86.20% |
| 260 | 99.54% | 86.99% | 93.06% | 87.38% |
| 270 | 99.57% | 88.42% | 93.51% | 88.49% |
| 280 | 99.53% | 87.02% | 92.65% | 87.34% |
| 290 | 99.34% | 85.60% | 91.85% | 85.89% |
| 300 | 99.24% | 84.44% | 90.50% | 83.35% |
| 310 | 99.16% | 83.93% | 89.44% | 82.12% |
| 320 | 99.13% | 84.16% | 89.70% | 82.65% |
| 330 | 99.25% | 84.87% | 90.77% | 84.08% |
| 340 | 99.39% | 85.80% | 92.08% | 85.92% |
| 350 | 99.51% | 86.88% | 92.94% | 87.36% |

表 2 提案手法における認識結果のまとめ

Table 2 Recognition results with proposed neural networks.

| | 提案手法 | | 最近傍法 | |
|-------|------------------|------------------|------------------|------------------|
| | 学習用 | 評価用 | 学習用 | 評価用 |
| 平均認識率 | 99.33% | 85.73% | 91.48% | 85.24% |
| 分散 | 0.025 | 2.285 | 1.950 | 4.779 |
| 最高認識率 | 99.57% (0度) | 89.24% (0度) | 94.26% (0度) | 89.61% (0度) |
| 最低認識率 | 99.07% (130度) | 83.67% (130度) | 89.30% (130度) | 81.89% (130度) |

が可能なことが分かる. また評価用文字においても, 提案手法の場合, 平均認識率は 85.73%, 最近傍法の場合は 85.24%となり, 提案手法の方が高い結果となった. 最近傍法の場合, 各文字種, 角度ごとに用意した識別関数を用いて認識しているが, それよりも高い結果が得られた. また最高認識率(提案手法の場合は 89.24%, 最近傍法の場合は 89.61%)と最低認識率(提案手法の場合は 83.67%, 最近傍法の場合は 81.89%)の差が小さく, 分散(提案手法の場合は 2.285, 最近傍法の場合は 4.779)も小さい結果となったことから, 最近傍法と比較して回転角の変動による影響も少ないことが分かる.

4.4 文字を回転させて認識を行った場合の結果

次に文字を回転させ認識を行った結果を示す. 以降は, 評価用文字に対する認識結果のみを示す. 文字の回転については, 下記に示す 6 パターンを行った.

回転方法 1 文字を回転させず (0度) に認識した結果, および時計方向に 10 度回転させ認識した結果を統合

回転方法 2 3.2 節の認識方法において, $r = 1$ とした場合 (-10度, 0度, 10度回転させて認識した結果を統合)

回転方法 3 $r = 2$ とした場合 (-20度, -10度, 0度, 10度, 20度回転)

回転方法 4 $r = 3$ とした場合 (-30度から 30度まで 10度ごとに回転)

回転方法 5 $r = 4$ とした場合 (-40度から 40度まで 10度ごとに回転)

回転方法 6 $r = 5$ とした場合 (-50度から 50度まで 10度ごとに回転)

ただし認識対象数が 3,036 個と多く, 計算量の抑制のため, まず文字を回転させずに認識し, その結果, 上位 20 個の候補に絞り (大分類), それら 20 個の候補を対象として, さらに文字を回転させて認識を行った (詳細認識). この場合, 上位 20 位までの分類率は 99.63%であった. 以上の回転方法によって認識した結果を表 3 に示す. 表 3 には, 表 2 と同様に, 回転方法ごとに平均認識率, 認識率の分散, すべての角度の中での最高認識率および最低認識率を示す.

表 3 より, 回転方法 1 では, 平均認識率が 90.18%となり, 回転しない場合 (85.73%) と比較して, 10 度回転させた結果と統合するだけで 4.45%の認識率の向上となった. 特に最低認識率が 83.67%から 88.66%と 5%向上した. さらには回転方法 6 では, 平均認識率が 96.41% (10.6%の認識率の向上) となり, 文字を回転させて認識する回数が多いほど, 認識率は向上する傾向にあることが分かる. また回転方法 5 および 6 の場合, 最低認識率が 96.14%と 96.22%であることから, 全角度において, 96%以上の認識率を得ることができた. このように, 回転させて認識する回数を増やすほど, 分散値が小さくなっていくことから, 文字の回転から受ける影響も少なくなることが分かる.

表 3 文字パターンを回転させた場合の認識結果

Table 3 Recognition results in rotating character patterns.

| | 平均認識率 | 分散 | 最高認識率 | 最低認識率 |
|----------------------|--------|-------|-----------------|------------------|
| 回転なし | 85.73% | 2.285 | 89.24% (0度) | 83.67% (130度) |
| 回転方法 1 (0度, 10度) | 90.18% | 1.435 | 92.35% (0度) | 88.66% (120度) |
| 回転方法 2 (-10度~10度) | 92.20% | 0.884 | 93.71% (0度) | 90.95% (130度) |
| 回転方法 3 (-20度~20度) | 94.47% | 0.258 | 95.25% (0度) | 93.76% (320度) |
| 回転方法 4 (-30度~30度) | 95.64% | 0.038 | 95.94% (0度) | 95.28% (230度) |
| 回転方法 5 (-40度~40度) | 96.22% | 0.002 | 96.32% (40度) | 96.14% (170度) |
| 回転方法 6 (-50度~50度) | 96.41% | 0.015 | 96.66% (40度) | 96.22% (170度) |

表 4 回転角を変えた場合の認識結果 (回転方法 2)

Table 4 Recognition results in rotating character patterns.

| | 平均認識率 | 分散 | 最高認識率 | 最低認識率 |
|-------------------|--------|-------|------------------|------------------|
| 回転方法 2 (10度) | 92.20% | 0.884 | 93.71% (0度) | 90.95% (130度) |
| 回転方法 2-a (20度) | 93.48% | 0.212 | 94.36% (0度) | 92.82% (140度) |
| 回転方法 2-b (30度) | 93.91% | 0.038 | 94.39% (30度) | 93.68% (80度) |
| 回転方法 2-c (40度) | 93.75% | 0.149 | 94.54% (320度) | 93.20% (170度) |

次に認識時において、回転角を変えて認識を行った。これまでの実験においては、回転角は10度であったが、回転方法2において、下記のように回転角を変えて認識を行った。

回転方法 2 $r = 1$ とし、回転角を10度とした場合 (-10度, 0度, 10度回転)

回転方法 2-a 回転角を20度とした場合 (-20度, 0度, 20度回転)

回転方法 2-b 回転角を30度とした場合 (-30度, 0度, 30度回転)

回転方法 2-c 回転角を40度とした場合 (-40度, 0度, 40度回転)

以上の結果を表4に示す。表4には、同じく平均認識率、認識率のばらつき、すべての角度の中での最高認識率および最低認識率を示す。

表4より、回転角を大きくすることによって、認識率が向上することが分かる。特に回転方法2-b (回転角30度) においては、平均認識率が93.91%となり、回転方法2 (92.20%) と比較して1.7%の認識率の向上を図ることができた。また、分散も小さくなっている (0.884 から 0.038) ことから、文字の回転から受ける影響が少なくなっている。

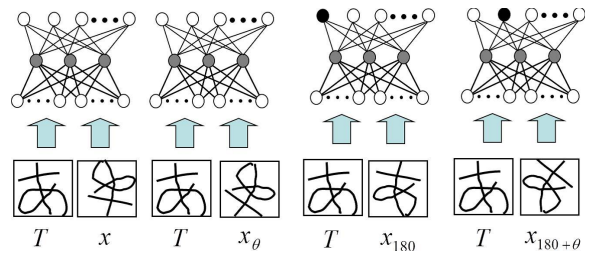


図 6 認識対象とする回転を制限したネットワーク

Fig. 6 Neural networks for limited character patterns.

ただし回転方法2-c (回転角40度) の場合、回転方法2-bと比較して、平均認識率 (回転方式2-cでは93.75%、2-bでは93.91%) および最低認識率 (回転方式2-cでは93.20%、2-bでは93.68%) が低下してしまっている。

そして回転方法2-aと回転方法3を比較すると、ほぼ同程度の認識率であることが分かる。したがって回転方法3においては、文字を10度回転させた認識結果の影響は少ないことが分かる。一方で、回転方法2-bと回転方法4を比較すると、回転方法4の方が認識結果が高く、10度および20度回転させて認識した影響が生じている。このように回転角を増やすと認識率が向上する一方で、認識しやすい角度、しにくい角度に違いがあることも分かる。

5. 考察

提案したネットワークの構造および計算量について考察する。4章で構築したネットワークは、0度から350度回転した文字を認識対象としていた。対象とする回転の範囲が狭まった場合、ネットワークの学習は容易になるものと考えられる。しかし、たとえば回転が0度から180度の範囲の文字を対象としたネットワークを構築した場合、180度を越えて回転した文字を認識することができない。そこでこのような文字については、さらに180度回転させて認識を行う必要がある。したがってこの場合、元の文字の回転を α 度とした場合、 α 度回転した文字を認識するだけでなく、 $\alpha + \theta$ 度、 $\alpha + 180$ 度、 $\alpha + 180 + \theta$ 度回転した文字も認識し、それぞれの認識結果を統合することになる (図6)。

そこで、前章で述べた回転の範囲が360度 (0度から350度) を対象としたネットワークのほか、180度 (0度から180度)、90度 (0度から90度) を対象としたネットワークの学習も行った。

- 回転の範囲が180度を認識対象とする場合
ネットワークの出力層のニューロン数は19個、他は360度の場合と同様である。認識時においては、 θ 度、180度、 $180 + \theta$ 度、文字を回転させる。
- 回転の範囲が90度を認識対象とする場合
ネットワークの出力層は10個、他は360度の場合と同じである。認識時においては、 θ 度、90度、 $90 + \theta$ 度、180度、 $180 + \theta$ 度、270度、 $270 + \theta$ 度、文字を

表 5 対象範囲の異なるネットワークでの認識結果

Table 5 Recognition results with neural networks for limited character patterns.

| | 範囲 | 平均認識率 | 分散 | 最高認識率 | 最低認識率 |
|--------|-------|--------|-------|----------------|----------------|
| 回転なし | 360 度 | 85.73% | 2.285 | 89.24% (0 度) | 83.67% (130 度) |
| | 180 度 | 87.25% | 3.375 | 92.28% (90 度) | 84.93% (310 度) |
| | 90 度 | 88.45% | 4.730 | 93.67% (270 度) | 85.97% (40 度) |
| 回転方法 1 | 360 度 | 90.18% | 1.435 | 92.35% (0 度) | 88.66% (120 度) |
| | 180 度 | 91.13% | 2.019 | 93.95% (260 度) | 89.36% (310 度) |
| | 90 度 | 91.82% | 2.857 | 94.64% (270 度) | 89.74% (40 度) |
| 回転方法 2 | 360 度 | 92.20% | 0.884 | 93.71% (0 度) | 90.95% (130 度) |
| | 180 度 | 92.90% | 1.296 | 94.74% (100 度) | 91.43% (310 度) |
| | 90 度 | 93.49% | 1.777 | 95.28% (270 度) | 91.81% (40 度) |
| 回転方法 3 | 360 度 | 94.47% | 0.258 | 95.25% (0 度) | 93.76% (320 度) |
| | 180 度 | 94.94% | 0.418 | 95.75% (100 度) | 94.02% (320 度) |
| | 90 度 | 95.41% | 0.626 | 96.14% (260 度) | 94.22% (50 度) |
| 回転方法 4 | 360 度 | 95.64% | 0.038 | 95.94% (0 度) | 95.28% (230 度) |
| | 180 度 | 96.01% | 0.079 | 96.34% (110 度) | 95.49% (320 度) |
| | 90 度 | 96.39% | 0.111 | 96.64% (250 度) | 95.68% (50 度) |
| 回転方法 5 | 360 度 | 96.22% | 0.002 | 96.32% (40 度) | 96.14% (170 度) |
| | 180 度 | 96.53% | 0.015 | 96.74% (130 度) | 96.34% (340 度) |
| | 90 度 | 96.83% | 0.001 | 96.90% (40 度) | 96.77% (80 度) |
| 回転方法 6 | 360 度 | 96.41% | 0.015 | 96.66% (40 度) | 96.22% (170 度) |
| | 180 度 | 96.73% | 0.030 | 97.0% (230 度) | 96.44% (360 度) |
| | 90 度 | 96.76% | 0.012 | 97.0% (230 度) | 96.63% (350 度) |

回転させる。

2つのネットワークの学習条件は、4.1節で述べたと同様に行った。また認識については、4.4節で述べた回転方法1から回転方法6による認識を同様に行った。回転の範囲が狭いネットワークの場合、認識時においてより計算量が必要となる。180度を対象としたネットワークについては、360度の場合と比較して2倍、90度を対象としたネットワークについては4倍必要となる。そこで計算量の抑制のため、4.4節と同様に、まず360度を対象としたネットワークを用いて、文字を回転させずに認識し、その上位候補20位に対して、180度もしくは90度を対象としたネットワークを用いて、認識を行った。以上の認識結果について表5に示す。

表5より、回転角の範囲が狭いネットワークの方が認識精度が高いことが分かる。回転せず認識した場合、360度を対象としたネットワークでは平均認識率が85.73%、180度では87.25%、90度では88.45%と認識率は向上していく傾向にある。これは、前述したように、認識対象の範囲が狭いネットワークの方が、学習対象数が少なく学習が容易なため、高い精度が得られたものと考えられる。ただし、回転する回数が多くなるにつれて、その差は小さくなっていく(回転方法6の場合、360度を対象としたネットワークでは平均認識率が96.41%、180度では96.73%、90度では96.76%)。このように、回転の範囲を制限して学習したネットワークを用いて、その範囲外の回転文字を認識する場合は、範囲内まで回転させることによって認識が可能な

ことが分かる。

以上のように、回転角の範囲が異なるネットワークを大分類部と詳細認識部に分けて認識することによって、計算量を抑制しながら認識率の向上が図れた。

6. むすび

本論文においては、回転したオフライン手書き文字を対象としてニューラルネットワークによる認識手法を提案した。これは、一定の角度ごとに回転した文字の認識が可能なニューラルネットワークの構造および学習方法について提案した。認識時においては、未知文字をさらに回転させて認識を行い、それぞれの結果を統合することによって最終的な認識結果を求める方法を提案した。そして文字画像データベース ETL9B, 3,036 字種を対象として、ニューラルネットワークの学習を行い、認識実験を行った結果、文字種の多い場合においても、回転文字の学習が可能であり、最近傍法と比較して高い認識率を得ることができた。そしてさらに文字を回転させて認識を行い、認識結果を統合することによって、認識率の向上を図ることができ、提案手法の有効性を示すことができた。また回転角の範囲が異なるニューラルネットワークを組み合わせ、大分類、詳細認識を行うことによって計算量を抑制しながら認識率の向上を図れることも示せた。

今後は、考察で述べた回転の範囲が異なるニューラルネットワークの組合せ方を検討することによって、認識率のさらなる向上を図ると同時に、回転文字の補正にも応用

していく予定である。

謝辞 本研究を行う際に、手書き文字データベースを提供していただきました産業技術総合研究所の皆様へ感謝いたします。

参考文献

- [1] 須崎健一, 荒屋眞二, 中村良三: 位置ずれ・回転パターンを認識するニューラルネットワーク, 情報処理学会論文誌, Vol.34, No.11, pp.2277-2281 (1993).
- [2] 須崎健一, 荒屋眞二, 中村良三: 任意に回転したパターンと回転角度を認識する複写学習モデル, 情報処理学会論文誌, Vol.35, No.7, pp.1311-1317 (1994).
- [3] 三小田聡, 井村 隆, 増山 博, 佐藤嘉伸, 田村進一: ニューラルネットワークによる図形の正規化, 映像情報メディア学会誌, Vol.52, No.11, pp.1713-1723 (1998).
- [4] 平 英二, 石田敏之, 内田誠一, 迫江博昭: 手書き文字列の非一様な傾き補正法, 電子情報通信学会技術研究報告, PRMU2000-207, pp.47-52 (2001).
- [5] 森田義則, 袋谷賢吉, 三日市政司: 回転不変量を用いた数字の認識, 電子情報通信学会論文誌, Vol.J73-D-II, No.11, pp.1906-1909 (1990).
- [6] 鳥潟文男, 木村重規, 鈴木章夫, 矢野雅文: Hough 変換を用いた回転と大きさに変な文字認識ニューラルネットワーク, 電子情報通信学会技術研究報告, PRMU99-88, pp.1-6 (1999).
- [7] 勘解由哲, 大西 昇, 杉江 昇: ニューラルネットによる回転と大きさに対して不変な手書き数字認識, 電子情報通信学会論文誌, Vol.J75-D-II, No.11, pp.1900-1908 (1992).
- [8] 佐藤俊治, 黒岩文介, 阿曾弘具, 三宅章吾: 回転対応型ネオコグニトロン, 電子情報通信学会論文誌, Vol.J81-D-II, No.6, pp.1365-1374 (1998).
- [9] Kulkarni, U.V., Sontakke, T.R. and Randale, G.D.: Fuzzy hyperline segment neural network for rotation invariant handwritten character recognition, *Proc. International Joint Conference on Neural Networks (IJCNN '01)*, Vol.4, pp.2918-2923 (2001).
- [10] 篠川敏行, 長谷博行, 角谷 浩, 米田政明: パラメトリック固有空間法による回転文字の認識, 画像電子学会誌, Vol.33, No.6, pp.1123-1131 (2004).
- [11] 山村昌史, 大山 航, 若林哲史, 木村文隆: 文字の回転に対して頑強な文字認識手法, 電子情報通信学会技術研究報告, PRMU2011-205, pp.111-112 (2012).
- [12] 羅 四維, 岩根典之, 山崎晴明: 人工神経網による手書き漢字の認識法, 情報処理学会論文誌, Vol.30, No.8, pp.962-969 (1989).
- [13] 朱 小燕, 山内康一郎, 神保孝志, 梅野正義: 階層的ニューラルネットワークを用いた手書き文字認識, 電子情報通信学会論文誌, Vol.J73-D-II, No.1, pp.54-61 (1990).
- [14] 堀田健一, 岩田 彰, 松尾啓志, 鈴木宣夫: 大規模ニューラルネット CombNET-II, 電子情報通信学会論文誌, Vol.J75-D-II, No.3, pp.545-553 (1992).
- [15] Saruta, K., Kato, N., Abe, M. and Nemoto, Y.: High Accuracy Recognition of ETL9B Using Exclusive Learning Neural Network-II (ELNET-II), *IEICE Trans.*, Vol.E79-D, No.5, pp.516-522 (1996).
- [16] 石井 力, 和泉勇治, 加藤 寧, 根元義章: ニューラルネットワークを用いた二者択一方式の手書き文字認識システム, 電子情報通信学会論文誌, Vol.J83-D-II, No.3, pp.988-995 (2000).
- [17] Rumelhart, D.E. and McClelland, J.L.: *Parallel Distributed Processing*, Vol.1, MIT Press (1986).
- [18] 大沢 裕, 山川健一: 図面の認識と理解, 昭晃堂 (1989).

- [19] 山田博三, 齊藤泰一, 山本和彦: 非線形正規化法の改良, 電子情報通信学会総合大会, D-439 (1988).
- [20] 孫 寧, 安倍正人, 根元義章: 改良型方向線素特徴量および部分空間法を用いた高精度な手書き文字認識システム, 電子情報通信学会論文誌, Vol.J78-D-II, No.6, pp.922-930 (1995).
- [21] 中野 馨: 入門と実習 ニューロコンピュータ, 技術評論社 (1988).



篠沢 佳久 (正会員)

1999年慶應義塾大学大学院理工学研究科後期博士課程修了。博士(工学)。視覚のモデル化に興味を持つ。電子情報通信学会, 情報システム学会, ヒューマンインタフェース学会各会員。