

ニューラルネットワークを用いた古文書個別文字認識に関する 一検討

和泉 勇治[†] 加藤 寧[†] 根元 義章[†] 山田 奨治[‡] 柴山 守^{††} 川口 洋^{††}
([†]東北大学大学院情報科学研究科) ([‡]国際日本文化研究センター)
(^{††}大阪市立大学術情報総合センター) (^{††}帝塚山大学経営情報学部)
〒 980-0845 仙台市青葉区荒巻字青葉 05
Tel.022-217-7140
E-mail:wai@nemoto.ecei.tohoku.ac.jp

あらまし

古文書中の文字には、くすし字や変形文字が多く、高精度な認識が困難である。本稿では、柔軟な情報処理が可能なニューラルネットワークの適用を試み、その有効性について考える。また、統計的手法の一つである改良型マハラノビス距離を用いた認識システムの認識結果とも比較を行い、その認識精度と問題点について考察する。

キーワード：古文書個別文字認識、ニューラルネットワーク、改良型マハラノビス距離

A Study for Character Recognitoin of Ancient Documents using Neural Network

Yuji Waizumi[†] Nei Kato[†] Yoshiaki Nemoto[†] Shoji Yamada[‡] Mamoto Shibayama^{††}
Hiroshi Kawaguchi^{††}
([†]Graduate School of Information Sciences, Tohoku University.)
([‡]International research Center for Japanese Studies.)
(^{††}Osaka City University Media Center.)
(^{††}Faculty of Business Administration, Tezukayama University.)
Aza05, Aoba, Aramaki Aoba-ku, Sendai, 980-0845 Japan
Tel.022-217-7140
E-mail:wai@nemoto.ecei.tohoku.ac.jp

abstract

For character recognition problem including deformation characters, such as ancient documents, the recognition system using neural network is considered to work effectively. We demonstrate the recognition accuracy of the system comparing with statistical method.

keywords : Ancient Document Recognition, Neural Network, Modified Mahalanobis Distance

1 まえがき

近年、パターンの統計的性質を用いた文字認識技術の研究が盛んに行われている。その中で、ベイズ識別やマハラノビス距離を用いた認識システムの有効性が確認されている。一般にそれらの手法は、パターンの分布が正規分布をしていると仮定し、学習サンプルからその分布の推定を行っている。パターン分布の推定を行う際、十分な学習サンプル数を確保できる場合は、非常に高い認識精度を得ることができる。しかし、学習サンプル数が少ない場合や字種間で異なる場合は、共分散行列などの認識に必要なパラメータを高い精度で推定することができず、認識精度の低下が生じてしまう問題点がある。図1に示されるような古文書文字の認識問題の場合、同一字種であってもくずしや書風により文字の形状が異なり、それらの分布が複数のクラスによって構成されていると考えられる。そのため、上記した統計的手法のように学習対象の分布形状を仮定しなければならない場合、その仮定が認識対象に対し妥当なものでなければ、高い認識精度は期待できず、より柔軟な認識手法の確立が必要であると言える。

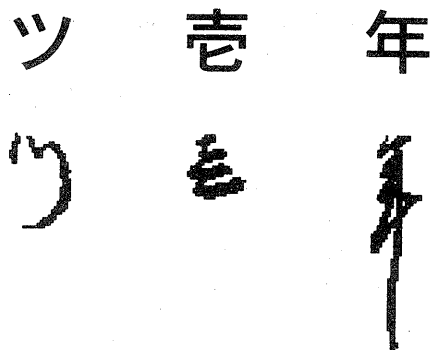


図1: 古文書文字の例

そのような認識問題に有効であると考えられる手法に、ニューラルネットワークを用いた文字認識手法がある。ニューラルネットワークは、その柔軟な情報処理と高い汎化能力により、高い認識精度が期待できる。しかし、その学習は、学習サンプルを繰り返し投入しながら学習する必要があり、統計的な手法と比較して学習の計算量が膨大になる欠点があり、十分な学習を行うことが不可能であった。しかし、最近の計算機の

性能向上に伴い、従来、学習が困難であった認識問題に対しても十分な学習を行うことが可能になり、統計的手法に迫る認識精度を確保できるようになって来ている。

主に、文字認識に用いられるニューラルネットワークモデルとして、MLP(Multi Layerd Perceptron)[1]、LVQ(Learning Vector Quantization)[2]、RBF(Radial Basis Function)[3]などが挙げられる。一般に、最も高い認識精度が得られているモデルはMLPである。MLPは教師あり学習であるBack Propagationによって学習を行う。教師あり学習を行うことにより、一つのネットワークが複数の字種を学習の対象とすることが可能となる。このことは、各字種の分布を推定するだけでなく、識別に必要な字種間の差異を学習することが可能であると考えられ、特に、形状の似た類似字種に対し高い認識精度が期待できる。そこで本稿では、ニューラルネットワークを古文書個別文字認識に適用し、その認識性能を統計的手法と比較して報告する。

2 ニューラルネットワークのモデルと動作

2.1 ネットワークモデル

認識実験に用いるネットワークモデルを図2に示す。ネットワークは入力層を含めて3層構造で、通常のMLPとは異なり、隠れ層と出力層の間に伝達する信号を自乗する機能を持つ自乗結合を導入している[4]。隠れ層ニューロンは線形の活性化関数を用い、出力層ニューロンはガウス型の活性化関数を用いる。

2.2 ネットワークの出力と学習

2.2.1 順伝搬

入力ベクトルを \mathbf{x} 、隠れ層ニューロン i の重みベクトルを \mathbf{w}_i 、出力層ニューロンの重みベクトルを \mathbf{W} とすると、ネットワークの出力 O は以下の式で定義される。

$$h_i = \mathbf{x} \cdot \mathbf{w}_i + \theta_i \quad (1)$$

$$H_i = h_i^2 \quad (2)$$

$$O = \exp(-\mathbf{W} \cdot \mathbf{H} + \theta) \quad (3)$$

ここで、 h_i は隠れ層ニューロン i の出力、 \mathbf{H} は h_i を自乗した H_i を成分に持つベクトル、 θ_i 、 θ はbiasニューロンとそれに対する重みにより決定される値である。

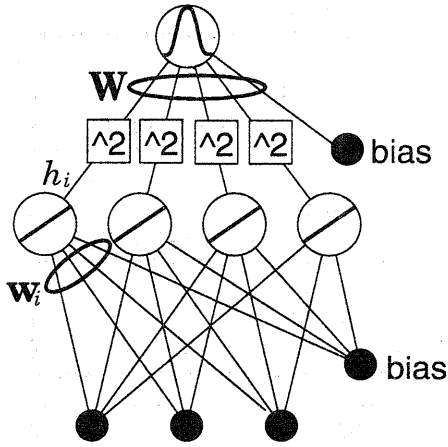


図 2: ネットワークモデル

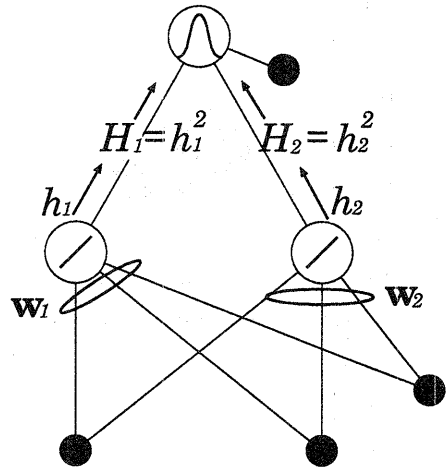


図 3: 2次元の楕円の学習:ネットワーク構成

2.2.2 逆伝搬

ネットワークの学習は、学習ベクトルとそれに対する教師信号の対を $(x^m, T^m), m = 1, \dots, M$ (M は定数) とした場合、式 (4) で定義される誤差に対し Back Propagation を適用して行う。

$$E = \frac{1}{2} \sum_{m=1}^M (T^m - O^m)^2 \quad (4)$$

自乗結合を導入したことにより学習則は次式のようになる。

$$\begin{aligned} \frac{dW_i}{dt} &= -\alpha \frac{\partial E}{\partial W_i} \\ &= -\alpha \sum_{m=1}^M (T^m - O^m) O^m h_i^{m2} \end{aligned} \quad (5)$$

$$\begin{aligned} \frac{dw_{ij}}{dt} &= -\alpha \frac{\partial E}{\partial w_{ij}} \\ &= -2\alpha \sum_{m=1}^M (T^m - O^m) O^m W_i h_i^m x_i^m \end{aligned} \quad (6)$$

ここで、 W_i と w_{ij} は、それぞれ、 \mathbf{W} と \mathbf{w}_i の各成分である。また、式 (1)、(3) の θ_i 、 θ は重み w_{ij} 、 W_i の一成分として表現した。

2.3 Weight Decay

ニューラルネットワークの汎化能力を向上させる手法として、式 (7) に示される Weight Decay がある。Weight Decay は、各ニューロンの持つ重み \mathbf{w} に式 (7)

を適用するアルゴリズムである。これは、重みベクトルが長くなると、ニューロンの活性化関数の傾きが急峻になり、未知入力の変動に過敏反応してしまうことを防ぐ効果がある。

$$\mathbf{w}_{t+1} = (1 - \beta)\mathbf{w}_t \quad (7)$$

β は崩壊のパラメータで $\beta \ll 1$ である。

2.3.1 2次元の楕円の学習

このモデルがどのような分布を表現し得るかを、2次元空間での楕円の学習を行うことで確認する。学習に用いるネットワーク構成は図 3 に示されるように 2 入力 1 出力で隠れ層ニューロン数は 2 とした。学習データは一様な分布の中からランダムに選出し、

$$(x - 3)^2 + (x - 3)y + y^2 < 0.2 \quad (8)$$

を満たす場合は教師信号に 1.0 を、それ以外は 0.0 を与え学習に用いた。

図 4 に学習データとネットワークの出力の等高線を示す。図から分かるように、学習対象としている 2 次元空間を構成する軸に対し、角度を持った楕円の長軸と短軸を学習出来ていることが分かる。

3 認識システムの概要

実験に用いる認識システムの概要を図 5 に示す。

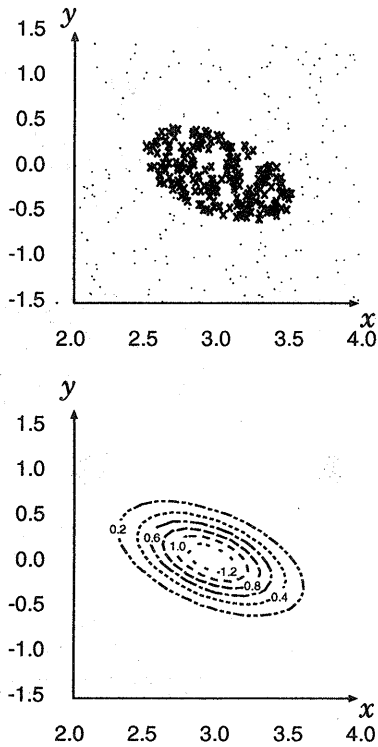


図 4: 2次元の楕円の学習:学習データとネットワークの出力の等高線

3.1 前処理

特徴抽出の前処理は、ノイズ除去としての孤立点除去と、文字の大きさの正規化を行う。特徴抽出で用いるイメージサイズが64ドット×64ドットであるため、入力イメージの幅と高さの大きい方を64ドットになるような倍率で、入力イメージの縦と横の比率を保つように正規化を行う。

3.2 特徴抽出

特徴量として改良型方向線素特徴量 [5] を用いる。改良型方向線素特徴量は196次元で構成される。抽出アルゴリズムは、前処理を施されたイメージに対し、輪郭線抽出・線素処理化を行う。線素としては縦(|), 横(—) 右上斜め(/), 左上斜め(\) の4種類を割り当てる。次に、これを8ドット×8ドットの正方形領域に分割し、その隣り合う4個ずつを一つの小領域

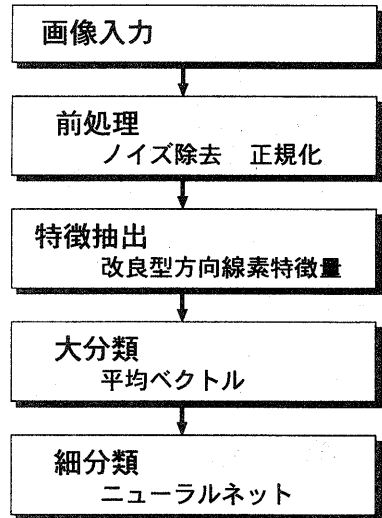


図 5: 認識システムの概要

とする。全部で49個の小領域となる。各小領域毎に線素の数を重み付きで数えることで特徴量とする。小領域49個×4種類の線素のため、196次元のベクトルが得られる(図6)。

3.3 大分類部

大分類部として、各字種の平均ベクトルを用いたパターンマッチングを用いる。距離尺度としては、ユークリッド距離を用いる。大分類部により、ある程度細分類部に入力する字種を削減することにより、誤認識

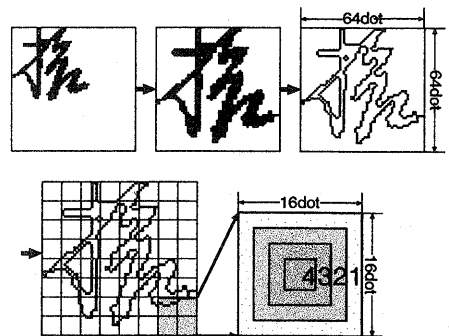


図 6: 方向線素特徴量

が生じる確率を抑えることができる。

3.4 細分類部

認識システムの細分類部として 2. に示したニューラルネットワークを用いる。細分類部の構成は、一つのネットワークが特定の一字種だけに発火するように割り当てられたモジュラー型のネットワーク構成とする。個々のモジュールは、割り当てられた字種に対しては発火し、それ以外の字種に対しては発火を抑制するように学習を行う。

4 古文書文字認識

古文書文字の認識実験を行う。3. で述べた認識システムの細分類部に、2. のニューラルネットワークを用いた場合と、統計的手法の一つである改良型マハラノビス距離を用いた場合とを比較する。

4.1 使用データ

実験に用いるデータは、「宗門改帳」古文書画像データベースに登録されている古文書画像から、川口ら [6] によって収集された 16 字種(ツ、一、二、三、四、五、六、七、八、九、十、卷、式、年、拾、廿)とする。各字種のサンプル数は、「廿」が 66 個で、その他の字種は 200 個である。学習には、「廿」以外は 80 個、「廿」は 33 個のサンプルを用いる。

4.2 ニューラルネットワークの構成

入力層、隠れ層、出力層の 3 層構成で、ニューロン数は、それぞれ、196、30、1 とした。図 7 に示されるように、各字種に一つの MLP モジュールが割り当て、最大の出力を得たモジュールに割り当てられた字種を認識結果とする。各モジュールは、割り当てられた字種に対し教師信号を 1.0、それ以外の字種に対し教師信号を 0.0 として学習を行う。学習率が 0.00001、教師信号との誤差絶対値の平均が 0.02 以下、または学習回数が 100 回を越えるまで学習を行う。Weight Decay の崩壊パラメータ β は、0.0 (Weight Decay 無し) と 0.000001 とした。

4.3 改良型マハラノビス距離

改良型マハラノビス距離 [7] は式 (9) で定義される。共分散行列の固有値にバイアスを加えることによって、

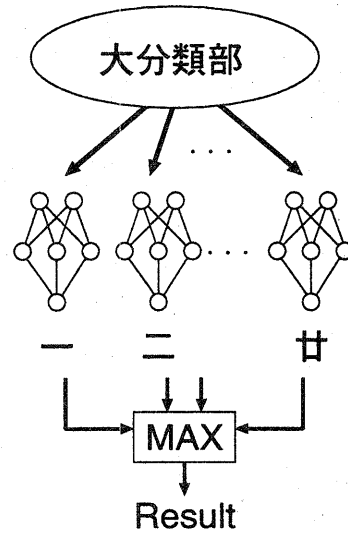


図 7: ニューラルネットワークの構成

小さい固有値の方距離に大きな影響を与えてしまうことを防いでいる。

$$D_m(x, u^i) = \sum_{j=1}^k \left(\frac{1}{\lambda_j + b} \right) \left((x - u^i)^t e_j \right)^2 \quad (9)$$

ここで、 b , x , u は、それぞれバイアスと入力ベクトル、標準パターンベクトルを表し、 e_j は、固有値 λ_j に対する固有ベクトルで、 $\lambda_j \leq \lambda_{j+1}$ である。

共分散行列から算出できる固有ベクトル数は、学習サンプル数によって決定され、サンプル数が 80 個の場合 79 個、33 個の場合 32 個となる。認識時には、「廿」以外の字種の標準パターンからの距離を求める時は $k = 79$ 、「廿」の標準パターンからの距離を求める時は $k = 32$ とした。計算に用いる次元数が小さい方が距離が小さくなるため、実験では、 D_m/k のように正規化された距離を用いて認識を行う。

4.4 認識結果

図 8、9 に、ニューラルネットワークを用いた認識率と改良型マハラノビス距離を用いた認識率をそれぞれ示す。図から分かるように、ニューラルネットワークの Weight Decay を適用したものが最も高い認識率の 96.67% が得られている。参考までに、「廿」を除いた

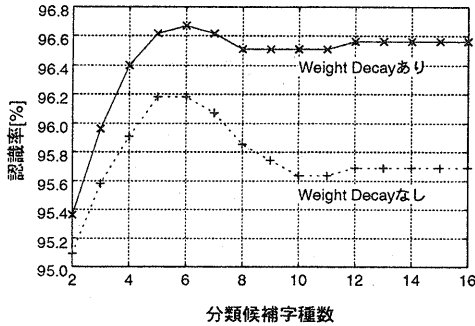


図 8: ニューラルネットワークを用いた認識率

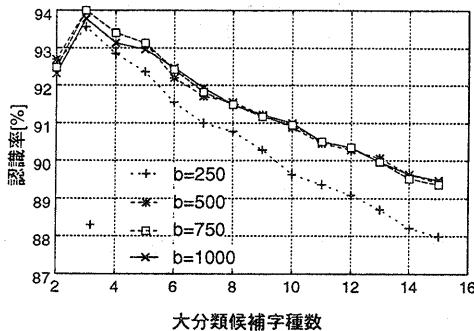


図 9: 改良型マハラノビス距離を用いた認識率

ニューラルネットワークと改良型マハラノビス距離の認識率はそれぞれ、97.05%と94.22%である。

表 1、2に、それぞれ、Weight Decay を適用したニューラルネットワークと改良型マハラノビス距離の誤認識を起こした字種の内訳を示す。これらは、それぞれの手法で最も高い認識率を得られた大分類候補字種数を用いた場合である。ニューラルネットワークを用いた場合では、サンプル数の少い「廿」を除いて全て90%以上の認識率を得られている。しかし、「廿」が他の比で20%程低い認識率となってしまった。「廿」の学習サンプル数は、他の字種の半分以下であったため、そのことが認識精度低下の原因になったと考えられる。ニューラルネットワークの認識精度とカテゴリ間の学習サンプル数の違いがどのような関係にあるか、調査が必要である。

表 1: 各字種の認識率 (ニューラルネットワークを用いた場合)

字種	正読数	誤読数	認識率 [%]
ツ	117	3	97.50
一	120	0	100.00
二	116	4	96.67
三	120	0	100.00
四	117	3	97.50
五	114	6	95.00
六	111	9	92.50
七	115	5	95.83
八	116	4	96.67
九	117	3	97.50
十	117	3	97.50
耂	118	2	98.33
弌	114	6	95.00
年	116	4	96.67
拾	119	1	99.17
廿	25	8	75.76
合計	1772	61	96.67

表 2: 各字種の認識率 (改良型マハラノビス距離を用いた場合)

字種	正読数	誤読数	認識率 [%]
ツ	117	3	97.50
一	116	4	96.67
二	110	10	91.67
三	116	4	96.67
四	118	2	98.33
五	111	9	92.50
六	109	11	90.83
七	106	14	83.33
八	106	14	83.33
九	113	7	94.17
十	117	3	97.50
耂	115	5	95.83
弌	112	8	93.33
年	113	7	95.17
拾	117	3	97.50
廿	27	6	81.82
合計	1723	110	94.00

5 まとめ

本稿では、古文書文字のようなくずしなどの変形の多い認識問題に対し、柔軟な情報処理が可能なニューラルネットワークが有効であると考え、統計的手法と比較してその認識精度を実験的に求めた。統計的手法の一つである改良型マハラノビス距離と比較して2.6%程高い認識率が得られたが、字種間の学習サンプル数に差があり、学習サンプル数が少ない字種の認識精度が極端に低くなってしまうことが明らかになった。古文書文字の場合、認識対象とする字種の十分な数のサンプルを収集することは困難であると考えられ、少ない学習サンプル数や字種間に偏りがある場合に対しても、高い認識精度を実現し得るネットワークアーキテクチャや少ない学習サンプルから認識精度の向上を可能にする学習サンプルの生成手法の検討が今後の課題であると言える。

参考文献

- [1] Rumelhart D.E., Maclelland J. E. and PDP research group: "Parallel Distributed Processing", 1,2 MIT Press, Cambridge, MA(1986).
- [2] T.Kohonen: "Self-organization and Associate Memory (2nd Edition)", Springer-verlag, pp.199-202 (1989).
- [3] M.J.D.Powell,"Radial basis function for multi-variable interpolation: a review", IMA Conference on Algorithms for the Approximation of Functions and Data, pp 143-167, RMCS, Shrivenham, 1985.
- [4] 和泉 勇治、加藤 寧、根元 義章、"自乗結合を持つニューラルネットワークによる手書き文字の高精度認識"、PRMU99-54、pp.37-44.
- [5] N.Sun,M.Abe and Y.Nemoto,"A Handwritten Character Recognition System by Using Improved Directional Element Feature and Subspace Method", Trans. IEICE J78-D-II, No. 6, pp. 922-930, 1995.
- [6] 日置 慎治、上原 邦彦、川口 洋、"「宗門改帳」に記録された年齢表記の認識"、挑戦 古文書 OCR、人文学と情報処理 No.18、pp.64-70.
- [7] 加藤 寧、安倍 正人、根元 義章:"改良型マハラノビス距離を用いた高精度な手書き文字認識", 信学論 (D-II), J79-D-II, No.1, pp45-52, 1996.