

エルマン型ネットワークを用いたオフライン手書き文字認識

Offline Handwritten Character Recognition by Elman Network

篠沢 佳久†

Yoshihisa Shinozawa

1. まえがき

本研究においては、文字の構造に着目し、エルマン型ネットワーク[1]を利用したオフライン手書き文字の認識方法を提案する。日本で用いられている文字、特に漢字は偏や旁などの部首と非部首から構成されており、「休」と「体」、「明」と「暗」のように任意の部首と非部首の構成がその文字の特徴となっているものもあり、その構成要素の組み合わせの違いが文字を認識する上で、有効な場合もある。そこで本研究においては、文字画像全体から特徴を抽出し、抽出した特徴全てを認識に利用するのではなく、文字画像を数個に分割し、その分割したパターンを組み合わせを特徴として利用することを提案し、エルマン型ネットワークを利用することによってその実現を試みる。

2. 特徴の分割とネットワークの構造

文字は書き手によってはその形状がずれるものの、文字を構成する部首や非部首などの局所的ではあるが、その文字を認識する上で特徴となりうる部品（これを部分パターンと呼ぶ）はほぼ同じ位置に書かれ、その相対的な位置関係は大きくは変動しない。しかし文字画像から部首や非部首などの部分パターンを直接抽出することは困難であるため、本研究においては、文字画像を均等に分割し（図1）、これらを認識に利用するための部分パターンとして扱う。

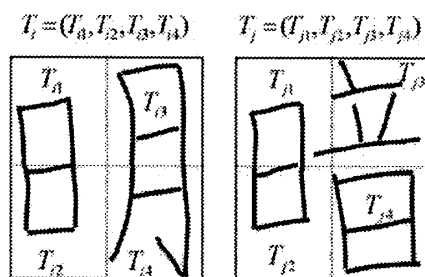


図1 文字画像の分割

本研究において、文字画像の学習および認識に利用するニューラルネットワークは入力層、中間層、出力層の3層から構成される（図2）。一般的にニューラルネットワークを用いて文字画像を学習する場合、一字種の文字に対して出力層の一ニューロンを対応させると、出力層のニューロンは総文字数分必要となるため、ニューラルネットワークの規模も大きくなり、学習が困難となる。

そこで出力層のニューロンを一つとし、特定のニューラルネットワークはそれと対応した一字種のみを認識対象とする手法が提案されている[2]。従って認識対象文字数を N 個とすると、同様に N 個のニューラルネットワークが必要となる。本研究においても同じく一つのニューラルネットワークの認識対象を一字種とする。

まず文字種 i から抽出した特徴 T_i を図1に示すように M 個（図1の場合 $M=4$ ）の部分パターンに分割する。

$$T_i = (T_{i1}, T_{i2}, \dots, T_{iM})$$

本研究においては、文字画像から抽出した特徴を全て一度にニューラルネットワークの入力層に入力し、学習及び認識に利用するのではなく、分割した部分パターンを組み合わせた局所的な特徴のみを入力層に入力する。すなわち文字種 i のみを認識するニューラルネットワーク N_i においては、文字種 i の M 個の部分パターンを用いて、 P 個の新たな特徴 $T'_{i1}, T'_{i2}, \dots, T'_{ip}$ を組み合わせる。そしてその組み合わせた特徴 $T'_{i1} = (T_{i1}, T_{i2})$ を入力した場合は出力層を発火させる。さらに別の特徴 $T'_{i2} = (T_{i1}, T_{i3})$ を入力した場合も同様に出力層を発火させる。一方で、他の文字種 ($j \neq i$) においても同様に、 M 個に分割した特徴 $T_j = (T_{j1}, T_{j2}, \dots, T_{jM})$ を用いて P 個の新たな特徴 $T'_{j1}, T'_{j2}, \dots, T'_{jp}$ を組み合わせる。そして特徴 $T'_{j1} = (T_{j1}, T_{j2})$ や $T'_{j2} = (T_{j1}, T_{j3})$ を入力した場合は、出力層は発火させない、といった挙動をするようにニューラルネットワークの学習を行なう（図2）。ニューラルネットワークの構造をこのようにすることによって、特徴数が増えても入力層の個数を制限することができる。

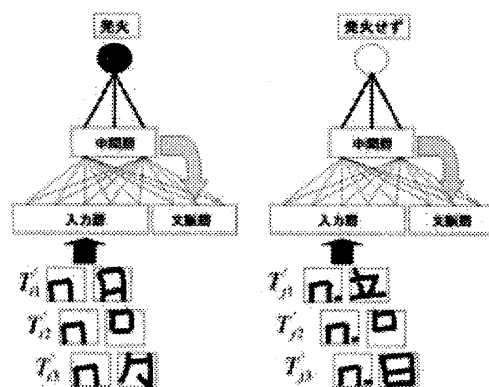


図2 「明」を認識するニューラルネットワーク

ここで問題となるのが、部分パターンの組み合わせを特徴として独立して入力し、正解となる教師信号を提示し、学習させても、以前入力した特徴の情報が残っていないと、学習できない場合もある（図2において、 T'_{i2} と T'_{j2} は類似したパターンであるが、出力層のニューロンの値は異なるように学習させなければならない）。そこで、以前までに入力した特徴の情報が残るように、ニューラルネットワー

† 慶應義塾大学理工学部管理工学科
Faculty of Science and Technology, Keio University

クの構造を中間層の出力値を文脈層にフィードバックさせたエルマン型(単純再帰型ネットワーク) [1]とする。

3. 学習方法

認識対象がN個の場合、教師信号となる文字種はそのニューラルネットワークと対応した1個であり、非教師信号の文字種は残りN-1個となる。しかし非教師信号として残りN-1個の文字種を利用するのは非効率である。そこで教師信号と類似している文字種のみを非教師信号として用いる。例えば「あ」「い」「う」「え」「お」の5文字を認識対象とした5個のニューラルネットワークを考える。ここで「あ」を認識させた場合、各ニューラルネットワークの出力値から「おうあえい」という認識結果になった場合、

- 「お」と「う」を認識対象としたニューラルネットワークは「あ」を入力すると発火しやすいことから、非教師信号の文字種として「あ」を使用する。
- 「え」と「い」を対象としたニューラルネットワークは「あ」を入力しても発火しづらいことから、非教師信号の文字種として「あ」は使用しない。

というように、学習文字パターンを利用して、非教師信号の文字種を決める。他のニューラルネットワークにおいても同様に、非教師信号の文字種を決める。個々のニューラルネットワークの学習アルゴリズムは下記の通りである。

Step1

学習文字パターンを用いて、大分類を行ない、文字種ごとに第K候補までに出現する文字候補の集合 C_i を調べる。文字種 i を対象としたニューラルネットワークにおいて、学習時に利用する非教師信号を C_i の中からランダムに選択し、その集合を R_i とする。また非教師信号の平均個数を r_{old} とする(全集合 R_i の要素数の平均)。

Step2

各ニューラルネットワークにおいて非教師信号 R_i を利用し、教師信号と非教師信号の文字種の判別ができるまでバックプロパゲーションアルゴリズムを利用して学習する。

Step3

文字種 i に対して、その文字候補の対象を集合 C_i として認識を行ない、上記で述べたようにその出力結果から、新たに非教師信号の集合 R_i を求める。なお新しい非教師信号の集合 R_i は元の集合 R_i の文字種も含める。その時の非教師信号の平均個数を r_{new} とする。

Step4

学習前後の非教師信号の平均個数(r_{old} と r_{new})との差が小さければ、学習を停止する。一方で学習後の非教師信号数 r_{new} が増加している場合は、 $R_i \rightarrow R_i$, $r_{new} \rightarrow r_{old}$ としてStep2に戻り再学習する。

4. 認識実験

認識実験のため文字パターンについては産業技術総合研究所提供のETL9Bの3036文字を利用した。学習文字パタ

ーンとして最初の一セット40パターンを利用し、認識用の未学習文字パターンとして二セット目の40パターンを用いた。文字パターンには非線形化処理を施し、 48×48 の大きさに正規化した後、196次元の輪郭線特徴を抽出した。抽出した特徴を4分割し、部分パターンの大きさを 8×8 の大きさとした。分割方法は図1に示した方法を含めて3通りの方法で行なった(図3)。その際、文字画像のずれを考慮し、部分パターンの各領域が重なるように分割した。各分割方法において、4分割した部分パターンから二つを選択し、それぞれ6種類の新たな特徴を組み合わせた。

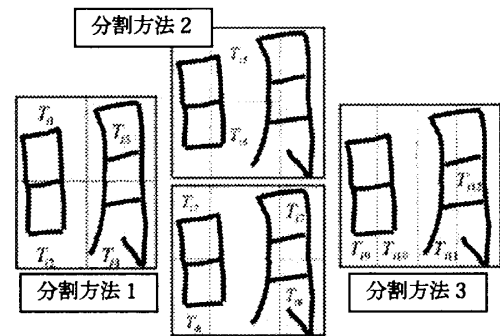


図3 文字パターンの分割方法

ニューラルネットワークの構造については、入力層のニューロンの個数は 8×8 の大きさの部分パターンを2個組み合わせているため128個、文脈層と中間層は32個、出力層は1個と固定した。Step1での大分類において、各文字候補の集合 C_i は第10候補までとした。Step2での学習回数は100回と固定し、これをStep2からStep4まで3回繰り返した。また入力層、文脈層への入力値はそれぞれ合計が1となるように正規化した。以上分割方法を変えて、入力する組み合わせ数を増やし、学習した結果を表1に示す。

表1 特徴の組み合わせ数を増やした時の認識率

組み合わせ方法	学習文字	未学習文字
ユークリッド距離	95.08%	92.47%
分割1(6通り)	99.17%	91.90%
分割1+2(12通り)	99.41%	92.75%
分割1+2+3(18通り)	99.39%	92.80%

実験結果より、12通りの特徴を用いた場合、ユークリッド距離の認識結果を上回ることができた。同一特徴を利用しているため、組み合わせ数をこれ以上増やしても認識率は向上しなかった。エルマン型ネットワークを利用することによって、特徴数が増加しても、入力層の個数を制限しつつ、認識率の向上を図ることができた。今後は異なる複数の特徴を併合して用い、認識率の向上を図る予定である。

参考文献

- [1] Elman, J. L., Finding Structure in Time, Cognitive Science, 14, 179-211(1990)
- [2] K.Saruta, N.Kato, M.Abe, and Y.Nemoto: High Accuracy Recognition on ETL9B Using Exclusive Learning Neural Network-II(ELNET-II), IEICE Trans., Vol.E79-D, No.5, pp.516-522(1996)