

ニューラルネットを用いた画像認識

苫小牧工業高等専門学校 ○西村尚也, 三上 剛, 吉村 斎

要 旨

本研究では、ニューラルネットを用いた画像認識の一環として、文字認識を行った。ニューラルネットは RBFN を用いた。RBFN の入力として画像を扱うが、主成分分析を行うことで冗長な情報を削減して学習認識をより効率的にした。

1 はじめに

近年、生産ラインにおける検査や医療分野、指紋照合などのセキュリティ分野など、幅広い分野で画像認識が用いられるようになってきた。

本研究では、画像認識の一環として文字認識を扱う。主成分分析カーネル主成分分析により冗長なデータを削減し、認識を行う。認識の方法としてニューラルネットの一種である RBF ネットワークを用いる。

2 主成分分析

データ解析を行うためには、高次元の膨大なデータを扱わなければならない。この高次元のデータの中には、特徴の変化を顕著に表すものと、冗長なデータとが存在する。主成分分析は、冗長なデータを削除してより効率的にデータ解析を行うことができるようにするための手法である。主成分分析では、最もよく特徴を表している、言い換えれば最も分散の大きい方向から順に新たな次元を定義していく。このようにして複数の特徴を統合して新たな次元を定義していくことで、小さい次元で多くの特徴の変化を表すことができ、冗長な次元のデータを削除して効率的にデータ解析を行うことができる。

主成分分析を拡張したものに、カーネル主成分分析がある。カーネル主成分分析は、データを極めて高次元の空間に写像し、その空間上で主成分分析を行うため、主成分分析とは違い、非線形の特徴を扱うことができる。このことから、カーネル主成分分析を行うと、データが線形分離可能となるような単純な特徴分布に変換されるため、ニューラルネットでの特徴分離が容易になる。また、データを高次元の空間に写像しながら、実際には高次元空間上での複雑な計算をせずに、カーネルの計算のみで複雑な識別関数を構成することができる。

カーネル主成分分析は (1) 式の固有値問題となる。

$$m\lambda\alpha = \alpha K \quad (1)$$

ここで、 K は (2) 式の行列である。

$$K_{ij} = \psi(\mathbf{x}_i) \cdot \psi(\mathbf{x}_j) \quad (2)$$

固有ベクトル α を求めたら、 $\psi(\mathbf{x}) \cdot \psi(\mathbf{y})$ に (3) 式を用いて、(4) 式により、入力ベクトル \mathbf{x} を \mathbf{V} に写像することができる。

$$k(\mathbf{x}, \mathbf{y}) = (\psi(\mathbf{x}) \cdot \psi(\mathbf{y})) \quad (3)$$

$$\mathbf{V} \cdot \psi(\mathbf{x}) = \frac{1}{\lambda} \sum_{i=1}^N \alpha_i k(\mathbf{x}_i, \mathbf{x}) \quad (4)$$

$k(\mathbf{x}_i, \mathbf{x})$ は、(5) 式、(6) 式で表すことができ、(5) 式を用いると主成分分析と同じ結果が得られる。

$$k(\mathbf{x}, \mathbf{y}) = \mathbf{x} \cdot \mathbf{y} \quad (5)$$

$$k(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{2\sigma^2}\right) \quad (6)$$

3 RBF ネットワーク

RBF ネットワークは階層型ニューラルネットの一種である。3 層構造で、1 層目が入力層、2 層目が中間層、3 層目が出力層である。入力層には、主成分分析であらかじめ次元変換したデータを入力する。

データが入力されると、最初に k 平均法を用いてクラスタリングを行い、クラスを中心 μ_j を求める。中間層では、(7) 式を用いて入力ベクトル \mathbf{x} とクラスを中心 μ_j との距離の正規関数を出力する。

$$k(\mathbf{x}, \mu_j) = \exp\left(-\frac{\|\mathbf{x} - \mu_j\|^2}{2\sigma^2}\right) \quad (7)$$

出力層では、(8) 式のような出力が得られる。

$$y(\mathbf{x}) = g\left(\sum_{j=1}^l w_{jk} h_j(\mathbf{x}) + w_0\right) \quad (8)$$

$$g(a) = \frac{1}{1 + \exp(-a)} \quad (9)$$

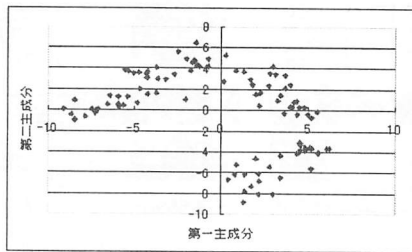
学習を行う際には、教師信号 d_k を入力し、(10) 式の評価関数を用いて、逆誤差伝播法を用いる。学習によって

値を定めるのは、重み w_{jk} とクラスを中心 μ_j である。

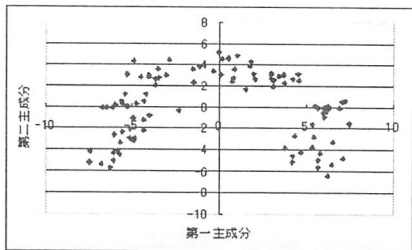
$$E = \frac{1}{2} \sum_{k=0}^n (y_k - d_k)^2 \quad (10)$$

4 実験結果

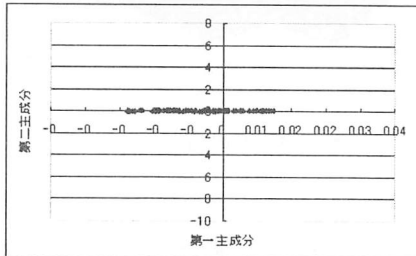
最初に、図 1(a) に示す 2 次元の人工データに対して主成分分析、(6) 式を用いたカーネル主成分分析を行った結果を図 1(b)、図 1(c) にそれぞれ示す。カーネル主成分分析の結果、複雑なデータを第一主成分のみで表せることがわかる。



(a) 人工データ



(b) 主成分分析の結果



(c) カーネル主成分分析の結果

図 1. 人工データに対する主成分分析の結果

主成分分析を行い、RBF ネットワークを構成して学習認識を行うプログラムを作成し、実験を行った。入力データには、手書き数字データベースの USPS (U.S. Postal Service digit database) を用いた。USPS は、"0"~"9" の 10 種類の手書き数字からなり、合計で学習用として 7291 文字、試験用として 2007 文字のデータがある。図 2(a) に使用したデータの一部を示す。

"0", "1", "2", "3", "4" の 5 パターンに対して、主成分分析を行い、学習認識を行ってみたところ、表 1 のような結果が得られた。

また、認識できたパターンと認識できなかったパターンの一部をそれぞれ図 2(b)、図 2(c) に示す。図 2(b) を見てみると、認識できたパターンは比較的形の整った数字であることが分かる。それに対し、図 2(c) を見てみると、認識できなかったパターンは本来の数字と比べてかなり形が崩れている。このことから、あまり形の崩れたパターンは認識できなかったと考える。



(a) 訓練データ



(b) 認識されたパターン



(c) 誤認識されたパターン

図 2. 手書き数字データ

表 1. PCA を用いたときの認識率

数字	訓練データ	テストデータ	認識データ	認識率
0	1194	359	345	96.1%
1	1005	264	253	95.8%
2	731	198	166	83.8%
3	658	166	130	78.3%
4	652	200	179	89.5%
計	4240	1187	1073	90.4%

5 おわりに

画像認識の一環として、主成分分析と RBF ネットワークを用いて手書きの数字認識の実験を行った。その結果、約 90% の認識率が得られたが、中間層の数などのパラメータの設定など、改善する余地があると考えられる。また、今回は 5 パターンに対して実験を行ったが、数字は全部で 10 パターンあるので、10 パターンの場合に対しても実験する必要がある。

6 参考文献

- [1] 坂野鋭, 武川直樹, 中村太一: "核非線形相互部分空間法による物体認識", 電子情報通信学会論文誌, D-II, Vol.J84-D-II, No.8, pp.1549-1556, 2001.
- [2] L.Huang, A.Shimizu, Y.Hagihara, H.Kobatake: "Robust Face Detection Using a Modified Radial Basis Function Network", IEICE on Inf. & Sys., Vol.E85-D, pp.1154-1162, 2002.