論　文

主成分の順位を保持する非線形主成分分析

三枝　亮[†]　坂野　鋭[††]　橋本　周司[†††]

Nonlinear Principal Component Analysis to Preserve the Order of Principal Components

Ryo SAEGUSA[†], Hitoshi SAKANO[††], and Shuji HASHIMOTO[†††]

あらまし　主成分分析（Principal Component Analysis: PCA）は，線形変換によりデータの次元数を削減する多変量解析法である．PCA は，その理論と実装の簡潔さから様々なデータの解析に用いられる．しかしながら，PCA によるデータの表現は，その線形性のために冗長となる場合がある．この問題を解決するために，いくつかの非線形 PCA の手法が提案されている．しかしながら，それらの手法には，生成される主成分が陽に順位付けされないことと，その主成分数をパラメータの調整前に決定する必要がある，という 2 つの欠点がある．本論文では，主成分の順位を保持した状態でデータを主成分に非線形変換するアルゴリズムを導出し，そのアルゴリズムを実現するニューラルネットワークを提案する．提案手法は，主成分の順位に対応した階層的な構造をもつため，その主成分の順位を保持することが可能である．また，パラメータの調整前に主成分数を決定する必要もない．さらに，提案手法の有効性を，数値実験を通して実証した．

キーワード　非線形主成分分析，階層構造，砂時計型パーセプトロン

1. まえがき

データを解析する上で，データの次元数を削減することは大切である．小さい次元数でのデータ表現は，データの理解を容易にし，データを計算処理するコストを抑えるからである．このような次元数を削減する手法として，主成分分析（Principal Component Analysis: PCA）が，パターン認識や画像処理などの広い分野で用いられている [1], [2].

PCA は，データの座標系を，データの表現が効率的である低次元の座標系に直交変換する．この直交変換の基底は，特に，データが線形多様体上に分布する場合に，データを効率的に記述する．しかしながら，もしデータが非線形多様体上に分布する場合には，その非線形多様体の次元数以上の次元数がデータを記述するのに必要となるので，データの記述は非効率的である．

この問題を解決するために，いくつかの非線形主成分分析（Nonlinear Principal Component Analysis: NLPCA）の手法が提案されている [3]- [7]．第 1 の手法は，入江らが提案した砂時計型多層パーセプトロン [8] を応用した手法である．第 2 の手法は，Hastie らによる部分線形近似に基づく手法である [9]．しかしながら，これらのいずれの手法においても，非線形 PCA のパラメータを調整する前に, 主成分数を決定しなければならない．従って，適当な主成分数を決定するには，パラメータの調整を試行錯誤的に繰り返すことになる．さらに，適切な主成分数を決定したとしても，これらの手法では，従来の PCA における寄与率に対応するパラメータがないため，主成分は順位付けられて構成されない．これらの欠点は，実世界の問題を扱う上で，その手法の応用を制限することになる．

また，第 3 の手法に，近年 Schölkopf によって提案された Kernel PCA [10] がある．Kernel PCA は，非線形写像による入力データの像に対して主成分分析を行なう手法であり，主成分の順位付けが定義できる．しかしながら，非線形写像のパラメータをデータの分

[†] 早稲田大学大学院理工学研究科，東京都
Graduate School of Science and Engeering, Waseda University, Okubo 3–4–1, Shinjuku-ku, Tokyo, 169–8555 Japan
[††] 株式会社 NTT データ，東京都
NTT Data Corp., Kayaba-cho tower 1–21–2, Shinkawa, Chuo-ku, Tokyo, 104–0033 Japan
[†††] 早稲田大学理工学部，東京都
School of Science and Engineering, Waseda University, Okubo 3–4–1, Shinjuku-ku, Tokyo, 169–8555 Japan

布に対して適切に決定する方法は未知であり，データの像に関する主成分分析の結果をもとに決定するにしても，非線形写像のパラメータを変更するごとに主成分分析の再計算が必要となるなどの欠点がある．さらに，Kernel PCA では，固有ベクトルを求めるためにデータの像の共分散行列に対する固有方程式を解く必要があり，主成分の計算には対象となるデータと全ての学習データとの Kernel 関数を計算する必要がある．このため，学習データを全て記憶する必要があり，学習データの数が多い場合には，固有方程式を解くことや主成分を計算するための処理量が非常に大きくなる問題がある．

本論文では，主成分の順位を保持した状態でデータを主成分に非線形変換するアルゴリズムを検討し，そのアルゴリズムを実現するニューラルネットワークを提案する．提案手法は，主成分の順位に対応した階層的な構造をもつため，その主成分の順位を保持することが可能である．さらに，提案手法では，パラメータの調整前に主成分数を決定する必要がない．これらの有効性を，数値実験を通して実証する．

以降，2. では，提案する非線形 PCA モデルを定式化する．3. では，その有効性を数値実験により示す．4. では，結論と今後の課題を述べる．

2. 非線形 PCA の定式化

2.1 線形 PCA から非線形 PCA への拡張

$\boldsymbol{x}, \boldsymbol{y}, \hat{\boldsymbol{x}}$ を，それぞれ，データ，特徴（主成分），復元データを表現するベクトルとする．簡単のため，$\boldsymbol{x}$ の平均を $E[\boldsymbol{x}] = 0$ とする．

自然数 $m, n \in N$ $(m \leqq n)$ に対して，データ空間 $\boldsymbol{x} \in U_x \subseteqq R^n$ から特徴空間 $\boldsymbol{y} \in U_y \subseteqq R^m$ 上への線形写像 $L^T : U_x \mapsto U_y$ を，

$$\boldsymbol{y} = L^T \boldsymbol{x} \tag{1}$$

$$= (\boldsymbol{e}_1^T \boldsymbol{x}, \boldsymbol{e}_2^T \boldsymbol{x}, \cdots, \boldsymbol{e}_m^T \boldsymbol{x})^T \tag{2}$$

と定義し，特徴空間 $\boldsymbol{y} \in U_y \subseteqq R^m$ から復元データ空間 $\hat{\boldsymbol{x}} \in U_{\hat{x}} \subseteqq R^n$ 上への線形写像 $L : U_y \mapsto U_{\hat{x}}$ を，

$$\hat{\boldsymbol{x}} = L\boldsymbol{y} \tag{3}$$

$$= \sum_{i=0}^{m} \boldsymbol{e}_i (\boldsymbol{e}_i^T \boldsymbol{x}) \tag{4}$$

と定義する．ここで，$\boldsymbol{e}_i$ は L の i 番目の列ベクトルである．

PCA では，データの復元に関する平均自乗誤差，

$$J = E[||\boldsymbol{x} - \hat{\boldsymbol{x}}||^2] \tag{5}$$

$$= E[||\boldsymbol{x} - LL^T \boldsymbol{x}||^2] \tag{6}$$

を最小化する L を決定する．この最小化基準は，L の列ベクトル $\{\boldsymbol{e}_i\}_{i=1,\cdots,m}$ が，正規直交基底，

$$\boldsymbol{e}_i^T \boldsymbol{e}_j = \delta_{ij} \quad i, j = 1, \cdots, m \tag{7}$$

である拘束のもとで，特徴の分散 $E[||\boldsymbol{y}||^2]$ を最大化する基準と等価である [3]．上述の PCA では，データ空間 U_x から特徴空間 U_y 上への写像 L^T は線形であり，同様に，特徴空間 U_y から復元データ空間 $U_{\hat{x}}$ 上への写像 L も線形である．以降では，このようにデータの変換写像が線形である PCA を線形 PCA と呼び，前者の写像 L^T をデータの特徴を抽出する意味で線形抽出写像，後者の写像 L をデータを復元する意味で，線形復元写像と呼ぶことにする．

線形変換によるデータ表現は，データの分布により，その効率性が異なる．データ $\boldsymbol{x} \in R^n$ が次元数 $m < n$ の線形多様体上に分布する場合，線形 PCA は次元数 m でデータを記述できるので，その記述は効率的である．例えば，$\boldsymbol{x}$ が R^n 上の超平面上に分布する場合がその例である．しかしながら，$\boldsymbol{x} \in R^n$ が次元数 $m < n$ の非線形多様体上に分布する場合には，線形 PCA では m 以上の次元数がデータを記述するのに必要となるので，その記述は非効率的である．例えば，$\boldsymbol{x}$ が R^n 上の超曲面上に分布する場合がその例である．この場合，データが分布する非線形多様体の座標系でデータを表現する，記述の冗長性が少ない非線形 PCA が有利である．

そこで，データ空間 $\boldsymbol{x} \in U_x \subseteqq R^n$ から特徴空間 $\boldsymbol{y} \in U_y \subseteqq R^1$ 上への非線形抽出写像 $\boldsymbol{\phi} : U_x \mapsto U_y$ を，

$$\boldsymbol{y} = \boldsymbol{\phi}\ (\boldsymbol{x}), \quad \boldsymbol{\phi} \in S_e \tag{8}$$

と定義し，特徴空間 $\boldsymbol{y} \in U_y \subseteqq R^1$ から復元データ空間 $\hat{\boldsymbol{x}} \in U_{\hat{x}} \subseteqq R^n$ 上への非線形復元写像 $\boldsymbol{\psi} : U_y \mapsto U_{\hat{x}}$ を，

$$\hat{\boldsymbol{x}} = \boldsymbol{\psi}\ (\boldsymbol{y}), \quad \boldsymbol{\psi} \in S_r \tag{9}$$

と定義する．ここで，S_e と S_r は，それぞれ，$\boldsymbol{\phi}$ と $\boldsymbol{\psi}$ が属する関数族である．変換写像を非線形にすることで，データと特徴（主成分）は非線形に対応する．

我々の問題は，データの復元に関する平均自乗誤差：

$$J = E[||\boldsymbol{x} - \hat{\boldsymbol{x}}||^2] \tag{10}$$

$$= E[||\boldsymbol{x} - \boldsymbol{\psi}(\boldsymbol{\phi}(\boldsymbol{x}))||^2] \quad (11)$$

を最小化する $\boldsymbol{\psi}$ と $\boldsymbol{\phi}$ を S_e と S_r の中から求めることである．J を最小化する $\boldsymbol{\psi}$ と $\boldsymbol{\phi}$ がデータの分布に応じて求まれば，非線形 PCA は線形 PCA の主成分数以下の主成分数でデータを記述できる．

2.2 非線形主成分の順位付け

PCA には，主成分数が多ければその記述能力は高まるが，必要なメモリや計算処理のコストも大きくなる，という関係がある．このため，使用する主成分数を目的に応じて調整することになる．線形 PCA の場合は，主成分の重要度が寄与率によって順位付けされるので，重要度順に主成分を用いることで，使用する主成分数の調整が可能である．一方，従来の非線形 PCA の場合は，主成分の重要度が陽に順位付けされていないので，使用する主成分数を変更するために変換写像のパラメータを再調整する必要があり，効率的ではない．

ここでは，主成分の重要度を陽に順位付けるために，主成分番号 $i = 1, \cdots, m$ に対して，データ空間 $\boldsymbol{x} \in U_x \subseteqq R^n$ から第 i 特徴空間 $y_i \in U_y^i \subseteqq R^1$ 上への非線形抽出写像 $\phi_i : U_x \mapsto U_x^i$ を，

$$y_i = \phi_i(\boldsymbol{x}) \quad i = 1, \cdots, m \quad (12)$$

と定義し，上位 i までの特徴直積空間 $(y_1, \cdots, y_i) \in \prod_{j=1}^{i} U_y^j \subseteqq R^i$ から第 i 復元データ空間 $\hat{\boldsymbol{x}}_i \in U_{\hat{x}}^i \subseteqq R^n$ 上への非線形復元写像 $\boldsymbol{\psi}_i : \prod_{j=1}^{i} U_y^j \mapsto U_{\hat{x}}^i$ を，

$$\hat{\boldsymbol{x}}_i = \boldsymbol{\psi}_i(y_1, \cdots, y_i) \quad i = 1, \cdots, m \quad (13)$$

と定義する．以降では，主成分の順位について，i が小さければ上位，大きければ下位と呼ぶことにする．

ここで，主成分と復元データは，

$$\hat{\boldsymbol{x}}_1 = \boldsymbol{\psi}_1(\phi_1(\boldsymbol{x})) \quad (14)$$
$$\hat{\boldsymbol{x}}_2 = \boldsymbol{\psi}_2(y_1, \phi_2(\boldsymbol{x})) \quad (15)$$
$$\vdots$$
$$\hat{\boldsymbol{x}}_i = \boldsymbol{\psi}_m(y_1, \cdots, y_{i-1}, \phi_i(\boldsymbol{x})) \quad (16)$$
$$\vdots$$
$$\hat{\boldsymbol{x}}_m = \boldsymbol{\psi}_m(y_1, \cdots, y_{m-1}, \phi_m(\boldsymbol{x})) \quad (17)$$

のように，同じ主成分番号の写像を組 $(\phi_i, \boldsymbol{\psi}_i)$ として，番号 i が小さい組から順に計算することにする．主成分の計算は主成分の上位下位にかかわらず独立に行なえるが，復元データの計算はそれより上位の主成分が必要である．

$(\phi_i, \boldsymbol{\psi}_i)_{i=1,\cdots,m}$ は，その組の復元データに関する平均自乗誤差，

$$J_i = E[||\boldsymbol{x} - \hat{\boldsymbol{x}}_i||^2] \quad (18)$$
$$= E[||\boldsymbol{x} - \boldsymbol{\psi}(y_1, \cdots, y_{i-1}, \phi_i(\boldsymbol{x}))||^2] \quad (19)$$

を最小化するように調整される．この調整は，一括処理の場合は番号 i の小さい順に，逐次処理の場合は i の小さい順または並列して行なわれる．

提案手法では，主成分の順位を，主成分を組み合わせて用いた場合にデータの記述が最も良くなる主成分の順序とする．提案手法は，非線形抽出写像 $\phi_1, \phi_2, \cdots, \phi_{k-1}$ と k 番目の非線形抽出写像 ϕ_k を組み合わせて用いたときに，データの記述が最良となるように ϕ_k を調整する．従って，$\phi_1, \phi_2, \cdots, \phi_{k-1}$ と組み合わせて用いる非線形抽出写像を $\phi_k, \phi_{k+1}, \cdots, \phi_m$ の中から選択する場合，データの記述能力は，k 番目の ϕ_k が最も良く，$k+1$ 番目以降の $\phi_{k+1}, \phi_{k+2}, \cdots, \phi_m$ はこれと同等か劣る．以上を k について帰納的に考えると，$\phi_1, \phi_2, \cdots, \phi_k, \cdots, \phi_m$ の順序が，主成分を組み合わせて用いた場合にデータの記述が最も良くなる順序，即ち，主成分の順位に相当するものと考えられる．

2.3 非線形 PCA の実装

2.2 では，非線形主成分を順位付ける手順の数理的枠組みを示した．本節では，これを具体的に実装する方法を示す．非線形主成分の構成法には様々な方法があるが，我々は，中間層素子数が無限の場合，任意の関数近似能力のあることが知られている多層パーセプトロン（Multi Layered Perceptron: MLP）に着目した．

まず，式 (8)，式 (9) の非線形 PCA を実現するニューラルネットワークとして，文献 [8] で提案されている図 1 の砂時計型 MLP について検討する．砂時計型 MLP は，その入出力関係が恒等写像となるように調整されることで，入力データを記述するのに適した内部表現を獲得する．図 1 において，最も左の層，中央の層，最も右の層を，それぞれ，入力層，抽出層，出力層と呼ぶことにする．入力層と出力層の素子数は等しい．抽出層の素子数は入力層および出力層の素子数よりも少ないため，抽出層でデータの特徴（主成分）が抽出される．入力層から抽出層までの部分と抽出層から出力層までの部分は，それぞれ，非線形抽出写像 $\boldsymbol{\phi}$ と非線形復元写像 $\boldsymbol{\psi}$ として機能する．砂時計型 MLP では抽出層の素子数が主成分数に対応するため，使用する

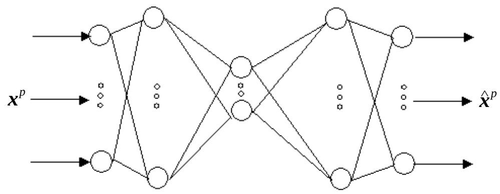

図 1　砂時計型 MLP の構造
Fig. 1　The structure of the sandglass-type MLP.

主成分数はパラメータの調整前に決定されなければならない．また，抽出層の各素子は互いに同列に扱われているため，これらの素子間に陽な重要度の違いが生成されないという欠点がある．

そこで，本論文では，2.2 で提案したアルゴリズムを実現する，階層型非線形主成分ネットワーク（Hierarchical Nonlinear Principal Component Network: HNPCN）を提案する．HNPCN は，$(\phi_i, \boldsymbol{\psi}_i)_{i=1,\cdots,m}$ のパラメータを逐次的な調整で獲得するネットワークである．

HNPCN の構造を図 2 に示した．HNPCN は，複数の独立な部分ネットワークが階層的に接続された構造をもつ．HNPCN の部分ネットワークは，主成分番号 $i = 1, \cdots, m$ で番号付けられており，第 i 部分ネットワークは式（16）を実現する．部分ネットワークの数は使用する主成分数の上限とし，部分ネットワークの層の数は 5 層以上とする．砂時計型 MLP の場合と同様に，各部分ネットワークの最も左の層，中央の層，最も右の層を，入力層，抽出層，出力層と呼ぶことにする．第 i 部分ネットワークの入力層から抽出層までの部分と抽出層から出力層までの部分は，それぞれ，第 i 主成分に対応した非線形抽出写像 ψ_i と非線形復元写像 $\boldsymbol{\psi}_i$ として機能する．

全ての部分ネットワークにおいて，入力層および出力層の素子数を，データの次元数と等しくする．第 i 部分ネットワークの抽出層の素子数は，i とする．他の層の素子数は，写像の記述能力を高く設定する場合は多く，写像の記述能力を低く設定する場合は少なくする．全ての部分ネットワークにおいて，入力層，抽出層，出力層の素子の出力関数を，

$$f(u) = u \tag{20}$$

とし，他の層の素子の出力関数を，

$$f(u) = \frac{1}{1 + \exp(-\frac{u}{T})} \tag{21}$$

とする．T は，定数である．u は，素子の入力値の荷重和である．以下では，第 i 部分ネットワークの荷重係数を w_i で表す．

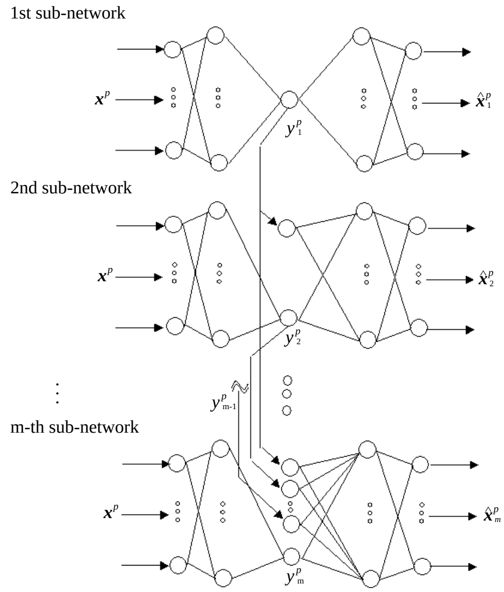

図 2　HNPCN の構造
Fig. 2　The structure of HNPCN.

HNPCN は，データの入力に対して，パラメータを逐次的に調整する．データ数 N のデータを $\{\boldsymbol{x}^p\}_{p=1,\cdots,N}$ とするとき，p 番目のデータ $\boldsymbol{x}^p$ に対する HNPCN の動作について説明する．

部分ネットワークの入出力動作は，上位の部分ネットワークから順に行なわれる．まず，$\boldsymbol{x}^p$ が第 1 部分ネットワークの入力層の素子に入力されると，第 1 部分ネットワークの抽出層の素子は，第 1 主成分，

$$y_1^p = \phi_1(\boldsymbol{x}^p) \in R^1, \tag{22}$$

を出力し，第 1 部分ネットワークの出力層の素子は，

$$\hat{\boldsymbol{x}}_1^p = \boldsymbol{\psi}_1(\boldsymbol{x}^p) \in R^n, \tag{23}$$

を出力する．次に，$\boldsymbol{x}^p$ が第 2 部分ネットワークの入力層の素子に入力されると，第 2 部分ネットワークの抽出層の第 2 素子は，第 2 主成分，

$$y_2^p = \phi_2(\boldsymbol{x}^p) \in R^1, \tag{24}$$

を出力する．第 2 部分ネットワークの抽出層の第 1 素子には，第 1 部分ネットワークで得た第 1 主成分が入力される．第 2 部分ネットワークの出力層の素子は，

$$\hat{\boldsymbol{x}}_2^p = \boldsymbol{\psi}_2(\boldsymbol{x}^p) \in R^n, \tag{25}$$

を出力する．以下同様に，第 m 部分ネットワークまで入出力動作は行なわれる．

HNPCN のパラメータは，入出力動作後に調整される．パラメータの調整は，上位の部分ネットワークから順に行なう．ただし，パラメータの調整は部分ネットワークごとに独立して行なえるので，部分ネットワークのパラメータを調整する順番は任意である．

データ $\boldsymbol{x}^p$ が入力されたときの第 i 部分ネットワークの荷重係数 w_i^p は，データの復元に関する平均自乗誤差，

$$J_i^p = E[||\boldsymbol{x}^p - \hat{\boldsymbol{x}}_i^p||^2] \tag{26}$$

$$= E[||\boldsymbol{x}^p - \boldsymbol{\psi}(y_1, \cdots, y_{i-1}, \phi_i(\boldsymbol{x}^p))||^2] \tag{27}$$

を最小化するように調整される．$\boldsymbol{x}^p$ に対する荷重係数の修正量 Δw_i^p は，

$$\Delta w_i^p = -\eta \frac{\partial J_i^p}{\partial w_i^p}, \tag{28}$$

である．荷重係数は，誤差逆伝播法により調整される．η は，一定な学習パラメータとする．

部分ネットワークのパラメータは，部分ネットワークごとに独立に調整されるため，上位の主成分が入力される下位の部分ネットワークから上位の部分ネットワークには，誤差信号が伝播されない．つまり，下位の部分ネットワークの調整は，上位主成分の入力を通して，上位部分ネットワークの抽出写像を補完するように強制される．

3. 数値実験

提案手法の有効性を確認するために，2 種類の人工データによる数値実験を行なった．一方は，分布の理解が容易である 3 次元データの実験である．他方は，波形データを用いた高次元データの実験である．本節では，これらの実験結果を示す．

3.1 3 次元データの実験

提案手法を用いて，R^3 における放物面上からサンプリングされたデータの主成分を獲得する実験を行なった．要素数 N のデータ $D_N = \{\boldsymbol{x}^p\}_{p=1,\cdots,N}$ の各成分は，関係式

$$x_3 = \frac{x_1^2}{a_1^2} + \frac{x_2^2}{a_2^2}, \quad x_1, x_2 \in [-1, 1] \tag{29}$$

を満たす．実験では，$a_1 = 1.0$，$a_2 = 3.0$ とした．

実験では，部分ネットワークの数を 2 とし，部分ネットワークの層の数は，5 とした．第 i 部分ネットワークの各層の素子数は，入力層から順に，$(3, 10, i, 10, 3)$ とした．他のパラメータは，表 1 に示した値を用いた．訓練データは，放物面上のランダムな点を，試験データは，図 3 に示される格子点とした．

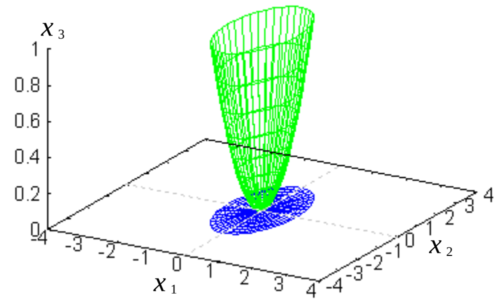

図 3　試験データ
Fig. 3　Test data.

訓練データによるパラメータの調整により，第 1 部分ネットワークの平均自乗誤差 $E[J_1^p]$ は正の値に収束し，第 2 部分ネットワークの平均自乗誤差 $E[J_2^p]$ は，0 に十分近い値に収束した．パラメータの調整後，HNPCN に試験データを入力した．

図 4 に，第 1 主成分から得られた復元データ $\hat{\boldsymbol{x}}_1$ を示す．図 5 に，第 2 主成分から得られた復元データ $\hat{\boldsymbol{x}}_2$ を示す．図 5 の $\hat{\boldsymbol{x}}_2$ は，第 2 部分ネットワークにおいて，第 1 主成分の値を 0 に固定して得られた復元データである．

式 (29) より，データの分布は放物線，

$$x_3 = \frac{x_2^2}{3^2}, \tag{30}$$

に沿って，最大の分散をもつ．図 4 は，第 1 主成分による復元データが式 (30) に対応することを示す．また，データの分布は放物線，

$$x_3 = x_1^2 \tag{31}$$

に沿って，2 番目に大きい分散を持つ．図 5 は，第 2

表 1　HNPCN のパラメータ
Table 1　Parameters of HNPCN

variable	value
η	0.05
T	0.1
N in training	20000
N in testing	200
initial w	random number over $[-0.1, 0.1]$

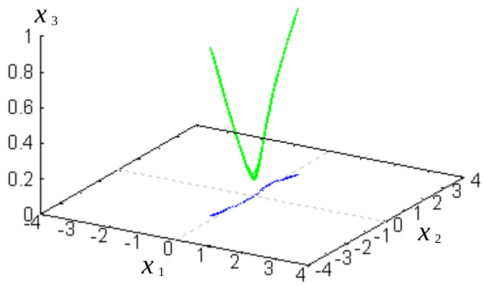

図 4　第 1 主成分による復元データ
Fig. 4　The reconstructed data from the first principal component.

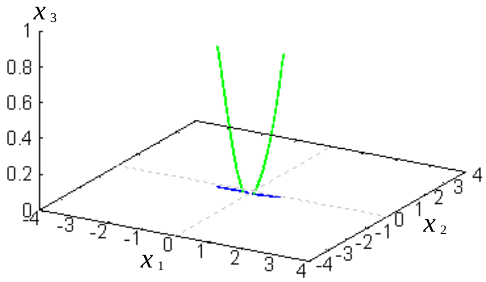

図 5　第 2 主成分による復元データ
Fig. 5　The reconstructed data from the second principal component.

主成分による復元データが式 (31) に対応することを示す．

図 6 に，第 1 主成分と第 2 主成分による復元データ $\hat{\boldsymbol{x}}_2$ を示す．HNPCN は 2 つの主成分で，データが分布する放物面の概形を得ている．一方，線形 PCA の場合は，データの表現に x_3 軸，x_2 軸，x_1 軸に対応する 3 つの主成分が必要である．提案手法では，データの分布する非線形多様体に適した特徴抽出が行なわれる．

図 7 に，第 1 主成分と第 2 主成分の分布を示す．水平軸は第 1 主成分であり，垂直軸は第 2 主成分である．この分布は，試験データの x_1x_2 平面への射影に類似していた．

次に，提案手法の部分ネットワーク数を 3 として，3 つの主成分を獲得する実験を行なった．部分ネットワークの数以外は，実験条件を上記の実験と同じ条件とした．表 2 に，データ数 1000 の試験データに対する各部分ネットワークの平均自乗誤差 (Mean Square Error: MSE) を示す．

表 2 によると，第 1，第 2，第 3 部分ネットワークによる平均自乗誤差は，それぞれ，0.18319, 0.00473, 0.00257 であるので，データの記述誤差は主成分数が多いほど小さい．

しかしながら，主成分数を 1 つから 2 つに増加させた場合の平均自乗誤差の減少量が 0.17846 であるのと比較して，主成分数を 2 つから 3 つに増加させた場合の平均自乗誤差の減少量は，0.00216 と小さい値であった．第 3 の主成分を加えても平均自乗誤差が大きく改善されないことから，データを記述するための主成分数は 2 つで充分であり，上位の非線形主成分ほど記述能力は高いことが分かる．

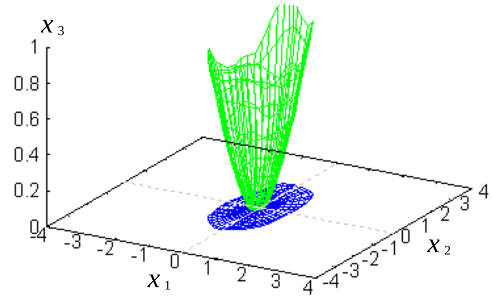

図 6　第 1，第 2 主成分による復元データ
Fig. 6　The reconstructed data from the first and second principal components.

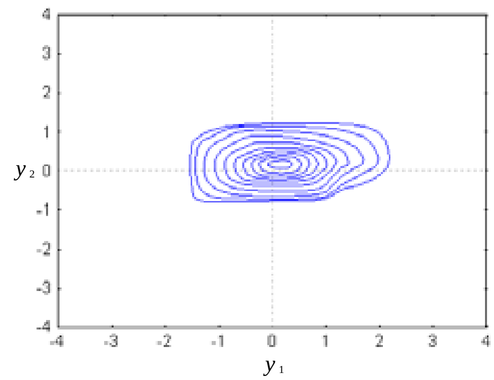

図 7　第 1 主成分と第 2 主成分の分布
Fig. 7　Distribution of the first and second principal component.

表 2　部分ネットワークの平均自乗誤差
Table 2　MSEs of sub-networks

# sub-network	MSE	diminution of MSE
1st	0.18319	-
2nd	0.00473	0.17846
3rd	0.00257	0.00216

さらに，上で述べた R^3 の放物面上のデータを，R^4, R^5, R^6 に埋め込んだデータについて，同様の実験を行なった．このような高次元の場合においても，HNPCN は R^3 の場合と同様に，試験データを 2 つの主成分から復元できた．

3.2 波形データの実験

提案手法を用いて，波形データからサンプリングされたデータの主成分を獲得する実験を行ない，高次元データに対する有効性を検討した．要素数 N のデータ $D_N = \{\boldsymbol{x}^p\}_{p=1,\cdots,N}$ の各成分は，3 つの三角関数を線形結合した関数，

$$f(s\tau+\theta) = \sum_{k=1}^{3} a_k \cdot sin(\omega_k(s\tau+\theta)) \qquad (32)$$

から，

$$(x_1, x_2, \ldots, x_n) = (f(\tau+\theta), f(2\tau+\theta), \ldots, f(n\tau+\theta)) \qquad (33)$$

のように等間隔でサンプリングした値とした．f の初期位相 θ を 0 から 2π の間で変動させて得た D_N は，N 次元データ空間の非線形多様体上に分布する．実験では，データ $\boldsymbol{x}^p$ の次元数を $n = 100$，f のパラメータを $(\omega_1, \omega_2, \omega_3) = (1.0, 2.0, 3.0)$，$(a_1, a_2, a_3) = (0.5, 0.3, 0.2)$，$\tau = \frac{2\pi}{100}$ とした．θ は，訓練データではランダムとし，試験データでは $[0, 2\pi]$ で等間隔な値とした．

実験では，部分ネットワークの数を 5 とし，部分ネットワークの層の数を 7 とした．部分ネットワークの抽出部と復元部の層の数を増やすことで，ネットワークの記述能力を高めた．第 i 部分ネットワークの各層の素子数は，入力層から順に $(100, 200, 200, i, 200, 200, 100)$ とした．他のパラメータは，表 3 に示した値を用いた．

訓練データによるパラメータの調整後，HNPCN に試験データを入力した．図 8 と図 9 は，初期位相が $\theta = 0$ と π の場合の復元データを示す．横軸は主成分番号，縦軸は成分の値である．

表 3 HNPCN のパラメータ
Table 3 Parameters of HNPCN

variable	value
η	0.001
T	0.1
N in training	50000
N in testing	100
initial w	random number over $[-0.1, 0.1]$

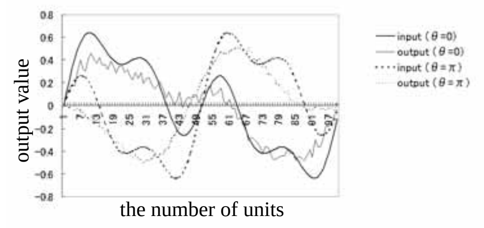

図 8 第 1 部分ネットワークの出力波形
Fig. 8 Output wave of the first sub-network.

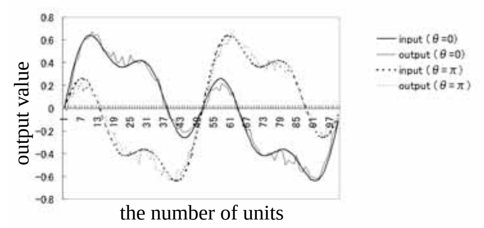

図 9 第 5 部分ネットワークの出力波形
Fig. 9 Output wave of the fifth sub-network.

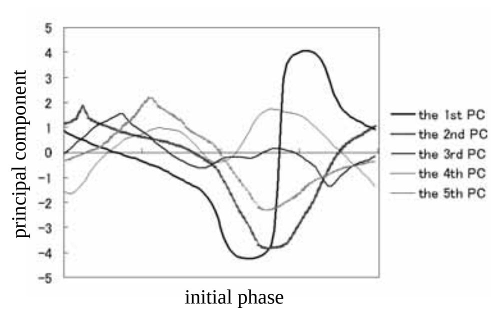

図 10 初期位相に対する主成分
Fig. 10 The principal components as to the initial phases.

図 8 と図 9 によると，5 つの主成分を用いた第 5 部分ネットワークの復元波形の方が，1 つの主成分を用いた第 1 部分ネットワークの復元波形よりも，試験データを良く復元している．

図 8 において，第 1 部分ネットワークの復元波形は，三角関数の概形に類似していることから，第 1 抽出写像は，f を構成する三角関数の中で最も周波数が小さい三角関数を獲得したものと考えられる．この三角関数は，最も大きい振幅を持つので，第 1 抽出写像は，他の抽出写像よりも効率的に波形を復元する．

図 10 は，初期位相 θ に対する各主成分の値を示す．図 10 によると，値の変動が小さい第 1 主成分は，最

も周波数の低い成分に対応する．下位の主成分は値の変動が大きいので，周波数の高い成分に対応する．この結果は，フーリエ解析による特徴抽出と比較できる．フーリエ解析は，三角関数を抽出写像として特徴を抽出する．一方，提案手法は，データの分布により抽出写像を調整するので，フーリエ解析では効率的に記述できない不連続信号のようなデータを，効率的に記述することが期待される．

4. むすび

主成分の順位を保持した状態でデータを主成分に非線形変換するアルゴリズムを導出し，そのアルゴリズムを実現するニューラルネットワークを提案した．提案手法は，主成分の順位に対応した階層的な構造を持つため，その主成分の順位を保持することが可能である．また，提案手法では，パラメータの調整前に使用する主成分数を決定する必要がない．

さらに，数値実験を通して，提案手法がデータの分布に応じて，特徴抽出とデータ復元を行なう非線形写像を獲得することを確認した．

本論文では，提案したアルゴリズムをニューラルネットワークで実装したが，このために，データの非線形性に関する性質が分かりにくいという問題がある．これを解決するには，アルゴリズムに用いる関数族を解析的に与えることが必要である．適切に関数族を与えるには，データと関数族の関係に関する検討が必要であるが，これは今後の課題としたい．また，提案手法が実世界のデータ圧縮やパターン認識の問題に対してどの程度有効かを評価することも課題である．

文　献

[1] R. Duda, P. Hart, Pattern classification theory and systems, Springer-Verlag, 1988.

[2] T. Sanger, Optimal unsupervised learning in a single-layer linear feedforward neural network, Neural Networks, Vol.2, pp.459-473, 1989.

[3] K. Diamantaras, S. Kung, Principal component neural networks theory and applications, John Wiley & Sons Inc, 1996.

[4] R. Gnanadesikan, Methods for statistical data analysis of multivariate observations, John Wiley & Sons Inc, 1977.

[5] J. Karhunen, J. Joutsensalo, Generalization of principal component analysis, optimization problems, and neural network, Neural Networks, Vol.8, No.4, pp.549-562, 1995.

[6] 高橋隆史，徳永隆治，平井有三，KL 変換を実現する 3 層線形パーセプトロンの教師付き学習則，信学論（D-II），Vol.J80–D–II, No.5, pp.1267–1275, May, 1997

[7] 高橋隆史，徳永隆治，重畳エネルギー関数による多層パーセプトロンの冗長性削減，信学論（D-II），Vol.J80–D–II, No.9, pp.2532–2540, Sep, 1997.

[8] 入江文平，川人光男，多層パーセプトロンによる内部表現の獲得，信学論（D-II），Vol.J73–D–II No.8, pp.1173-1178, 1990.

[9] T. Hastie, W. Stuetzle, Principal curves, Journal of the American statistical association, Vol.84, No.406, pp.502-516, 1989.

[10] B. Schölkopf, A. Smola, and K. Müller, Nonlinear component analysis as a kernel eigenvalue problem, Vol.10, No.5, pp.1299-1319, 1998.

（平成 x 年 xx 月 xx 日受付）

三枝　亮　（学生員）

平 10 早大・理工・応用物理卒．同大大学院理工学研究科博士課程在学中．ニューラルネットワークの研究に従事．

坂野　鋭　（正員）

昭 63 中大・理工・物理卒．平 2 埼玉大大学院理学研究科修士課程了．同年，NTT データ通信入社．以来，文字認識，バイオメトリクスなどパターン認識の研究に従事．複雑系の科学に興味をもつ．応用数理学会，物理学会各会員．早稲田大学客員研究員．

橋本　周司　（正員）

昭 45 早大・理工・応用物理卒．東邦大学講師，助教授を経て，現在，早稲田大学理工学部教授．確率過程の応用，画像処理，ロボティクス，音楽情報処理，などの研究を通して，メタアルゴリズム，感性情報処理，ヒューマンインタフェースに興味を持つ．工博．情報処理学会，日本バーチャルリアリティ学会，日本顔学会各会員．