電気学会論文誌 E（センサ・マイクロマシン部門誌）
IEEJ Transactions on Sensors and Micromachines
Vol.138 No.4 pp.123–131　DOI: 10.1541/ieejsmas.138.123

論　文

非線形写像による高次元センサ情報の可視化とクラス構造の解析

正　員　佐藤　仁樹*　非会員　佐藤　雅子**　非会員　高尾　佳史***

Visualization of High-dimensional Sensor Data and Analysis of their Class Structure Using Nonlinear Map

Hideki Satoh*, Member, Masako Satoh**, Non-member, Yoshifumi Takao***, Non-member

（2017年6月16日受付，2017年10月30日再受付）

An algorithm that constructs a nonlinear map from a high-dimensional feature space into a low-dimensional space was developed to enable analysis of the structure of data with high-dimensional characteristic features and their class information obtained using various sensors and analyzers. First, a nonlinear map is defined by summing nonlinear basis functions, and their optimal combination is derived using a genetic algorithm to avoid the "curse of dimensionality." Next, the coefficients of the basis functions are derived using the Nelder-Mead method to flexibly cope with the various demands for the map that cannot always be expressed using statistics of the characteristic features. As a result, nine-dimensional sake data can be mapped into a two-dimensional space so as not only to discriminate the classes but also to preserve the order of distances between classes as much as possible.

キーワード：遺伝的アルゴリズム，Nelder-Mead 法，基底関数，非線形，写像，可視化
Keywords: genetic algorithm, Nelder-Mead method, basis, nonlinear, map, visualization

1. はじめに

化学分析，味覚センサ，匂いセンサ等を用いて，様々な食品の特徴量を測定し，それらを 2 または 3 次元空間上に配置した食品の相関図を作成する問題を考える。この相関図により多次元の食品の特徴を視覚的に把握できるため，消費者は自分のニーズに合わせて食品を選択できる。この相関図には様々な利用目的が想定される。例えば，食品の栄養成分を特徴量とした相関図は，バランスが取れた栄養を摂取するための食事の提供に利用できる。また，清酒やワイン等の嗜好品の場合，味や匂いを特徴量として作成された相関図に自分の好みの銘柄があれば，試飲することなく好みに近い銘柄を探し出せる。さらに，料理との相性，製造法，地域による味の違いなどを特徴量とすることにより，様々な応用が考えられる。

従来から，様々な目的に対して多くの相関図が作られてきた。しかし，それらの多くには何らかの形で明示的でない制作者の手作業が入っているため，客観的な基準が明確でない場合が多い。誰が見ても納得できる相関図を作成するためには，データと客観的な基準のみに基づき，手作業を排除して相関図を自動的に作成するためのアルゴリズムが必要となる。

高次元の特徴量を 2 または 3 次元の低次元空間に写像し，高次元の特徴量の分布を視覚的に把握するために，様々な手法が提案されてきた。正準判別分析 [1] は，特徴量が属するクラス情報が与えられている際，低次元空間における各クラスの特徴量の分布ができるだけ分離するように写像を構築する手法である。多次元尺度構成法 [2] は，高次元における特徴量の相対距離を保存するように低次元空間に写像する手法である。主成分分析 [3] は，低次元空間の座標軸上

* 公立はこだて未来大学 システム情報科学部
〒 041-8655 北海道函館市亀田中野町 116-2
School of Systems Information Science, Future University Hakodate
116-2, Kamedanakano-cho, Hakodate, Hokkaido 041-8655, Japan
** 情報ノ宮蕗の下工房
〒 041-0833 北海道函館市陣川町 80-33
Johonomiya Fukinoshita Studio
80-33, Jinkawa-cho, Hakodate, Hokkaido 041-0833, Japan
*** 菊正宗酒造㈱総合研究所
〒 658-0026 兵庫県神戸市東灘区魚崎西町 1-8-6
General research laboratory, Kiku-masamune sake brewing Co. Ltd.
1-8-6, Uozaki-nishimachi, Higashinada-ku, Kobe, Hyogo 658-0026, Japan

での特徴量の分散ができるだけ大きくなるように低次元空間の座標軸を構成する。これは，低次元空間に写像された特徴量から高次元の特徴量を復元した際，できるだけ誤差が少なくなる写像と等価であり，誤差の観点から情報を圧縮するための代表的な手法である。これらの手法を拡張し，非線形に歪んだデータの分布に対応するために，カーネル正準判別分析[4]，ISOMAP[5]，非線形主成分分析[6]等の様々な手法が提案されている。

これらの手法は，データを変換（情報圧縮）する際の基準（目的関数）をデータの統計量に基づく数式で表現しているため，少ない計算量で分析できる。例えば，正準判別分析は，共分散によって定義されるクラスの分離度とクラス内のデータの散らばりの比が最大になるように写像を決める。このデータ変換基準は，固有値問題として定式化されるため，問題を高速に解ける。しかし，実際のデータのデータ数や分布がこれらの手法の想定と異なる場合には，必ずしもクラスが分離する写像を構築できるとは限らない。また，低次元空間に写像された特徴量の分布に対する要求は必ずしもデータの統計量で表されるとは限らない。そのため，低次元空間に写像された特徴量に対する様々な要求に柔軟に対応できない。また，非線形に拡張された手法では，測定された特徴量と相関図の座標軸の関係が必ずしも明確でないため，相関図の物理的な解釈が困難である。

そこで，我々は，これらの問題を解決するために，データと客観的な基準に基づいて，様々な目的に柔軟に対応できる手法を開発した。

本論文では，まず，2章で，高次元の特徴量を低次元空間に写像・圧縮する手法を定式化し，その後，従来手法の問題点をまとめる。次に，3章で，それらを解決するために開発した手法を述べる。提案手法では，非線形写像を高次元の特徴量の陽な関数で表すために，非線形写像の基底関数としてLegendre関数(付録1参照)を用いる。また，特徴量の次元の増大に伴い必要となる基底関数の組合せが爆発的に増えてしまう次元の呪い[7]の問題を解決するために，非線形写像を構成する基底関数の膨大な組合せの中から，遺伝的アルゴリズム（GA）[8][9]を用いて，適切な組合せを選択する[10]。さらに，データを変換する際の目的関数がデータの統計量で表されないような場合にも柔軟に対応するために，基底関数の係数をNelder-Mead法（NMM）[11]で求める。

上記提案手法を，複数のクラスに分けられた清酒の味データの相関図を作成する問題に適用した。この問題では，各クラスの分布が重複しないように相関図に表示されなければならないだけでなく，各クラスの相互距離の情報もできるだけ保存されるべきである。さらに，味の違いを直感的に把握するためには，相関図の座標が人間の味の感覚に合っている必要がある。4章では，提案手法により，これらの要求が満たされることを示す。

最後に，5章で，本論文の提案とその評価結果をまとめる。

2. 情報圧縮と非線形写像

〈2·1〉 情報圧縮と正準判別分析　標本データベクトル $\boldsymbol{x} \in \mathcal{R}^{d_x}$ と写像データベクトル $\boldsymbol{y} \in \mathcal{R}^{d_y}$ の関係を次式で表す。

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x}) \cdots\cdots (1)$$

$\boldsymbol{f}$ が線形の場合，上式は次式により表される。

$$\boldsymbol{y} = M^{\mathrm{T}}\boldsymbol{x} \cdots\cdots (2)$$

ここで，T は転置，$M \stackrel{\text{def}}{=} [\boldsymbol{e}_1, \cdots, \boldsymbol{e}_{d_y}]$ は $d_x \times d_y$ の係数行列，$\|\boldsymbol{e}_1\| =, \cdots, = \|\boldsymbol{e}_{d_y}\| = 1$ である。$d_x > d_y$ の場合，上式は情報圧縮であり，目的に応じて様々な解法が提案されている[1][2][3]。

$\boldsymbol{x}_{(n)}$ を n 番目の標本データベクトル，$\boldsymbol{y}_{(n)}$ を式(2)により得られる n 番目の写像データベクトル，N_{class} をクラス数，$a_{(n)}$ を $\boldsymbol{x}_{(n)}$ および $\boldsymbol{y}_{(n)}$ が属しているクラスの番号，N_{data} を $\boldsymbol{x}_{(n)}$ と $\boldsymbol{y}_{(n)}$ の標本数とする。与えられた標本データベクトル集合 $\mathcal{D} \stackrel{\text{def}}{=} \{\boldsymbol{x}_{(n)}, a_{(n)} | n = 1, \cdots, N_{\mathrm{data}}\}$ を，写像データベクトル集合 $\tilde{\mathcal{D}} \stackrel{\text{def}}{=} \{\boldsymbol{y}_{(n)}, a_{(n)} | n = 1, \cdots, N_{\mathrm{data}}\}$ に写像した際，データが属するクラスの集合が，できるだけ分離することを目的として行列 M を構築する方法に，正準判別分析[1]がある。

各クラスに属する標本データベクトルが多次元正規分布に従うと仮定する。$\bar{\boldsymbol{x}}^{(\ell)}$ をクラス ℓ に属する標本データベクトルの平均，$S^{(\ell)}$ をクラス ℓ に属する標本データベクトルの共分散行列，$\bar{\boldsymbol{x}}$ を全標本データベクトルの平均，$N_{\mathrm{data}}^{(\ell)}$ をクラス ℓ に属する標本データベクトルの標本数とすると，群間変動行列 B と郡内変動行列 W は次式で表される。

$$B \stackrel{\text{def}}{=} \sum_{\ell=1}^{N_{\mathrm{class}}} N_{\mathrm{data}}^{(\ell)} (\bar{\boldsymbol{x}}^{(\ell)} - \bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}^{(\ell)} - \bar{\boldsymbol{x}})^{\mathrm{T}} \cdots\cdots (3)$$

$$W \stackrel{\text{def}}{=} \sum_{\ell=1}^{N_{\mathrm{class}}} (N_{\mathrm{data}}^{(\ell)} - 1) S^{(\ell)} \cdots\cdots (4)$$

ここで，写像データ空間 $\mathcal{R}^{d_y}$ の i 番目の座標軸 y_i における群間変動と郡内変動の比（群間変動/ 郡内変動）λ_i を

$$\lambda_i \stackrel{\text{def}}{=} \frac{\boldsymbol{e}_i{}^{\mathrm{T}} B \boldsymbol{e}_i}{\boldsymbol{e}_i{}^{\mathrm{T}} W \boldsymbol{e}_i} \cdots\cdots (5)$$

で定義すると，λ_i および $\boldsymbol{e}_i$ は，各々 $W^{-1}B$ の固有値と固有ベクトルとなる。ここで，上式分子は y_i 軸におけるクラスの散らばり具合を，上式分母は y_i 軸における各クラス内のデータの散らばり具合を表す。また，$\lambda_i \geq \lambda_{i+1}$ とする。

このように，正準判別分析では，写像データ空間 $\mathcal{R}^{d_y}$ においてデータが属するクラスの集合ができるだけ分離するという問題を，クラス相互の散らばり具合を表す群間変動とクラス内部の散らばり具合を表す郡内変動の比で表される目的関数を最大にする最適化問題に置き換える。この置き換えにより，目的関数が標本データベクトルの平均や共

分散などの統計量で記述されるため，この最適化問題は最終的に固有値問題に帰着される。その結果，固有値問題を解くアルゴリズムを用いることにより高速に最適解が得られる。

一方，データが属するクラスの集合ができるだけ分離することと，群間変動と郡内変動の比を最大化することは必ずしも一致しない。また，上述の様に，正準判別分析は，目的関数を統計量（群間変動と群内変動の比）で表し，標本データが多次元正規分布に従うことを想定している。そのため，標本データの分布が多次元正規分布と異なる場合には，データが属するクラスの集合ができるだけ分離するという本来の目的が必ずしも達成されるとは限らない。

〈2·2〉 非線形写像 標本データベクトルの分布および写像問題の目的によっては，線形写像の代わりに非線形写像を用いる必要がある。そこで，本節では，正規直交基底関数を用いて非線形写像を定式化する[10]。

$\boldsymbol{k}_i$ を i 番目の基底関数のインデックスベクトル，N を展開次数（基底関数の数），基底関数を $\{K(\boldsymbol{x},\boldsymbol{k}_1),\cdots,K(\boldsymbol{x},\boldsymbol{k}_N)\}$，$\boldsymbol{\mathcal{Z}} \stackrel{\text{def}}{=} \{\boldsymbol{k}_1,\cdots,\boldsymbol{k}_N\}$ をインデックスベクトルの集合，$\boldsymbol{K}(\boldsymbol{x},\boldsymbol{\mathcal{Z}}) \stackrel{\text{def}}{=} (K(\boldsymbol{x},\boldsymbol{k}_1),\cdots,K(\boldsymbol{x},\boldsymbol{k}_N))^{\mathrm{T}}$ とすると（付録 1 参照），式 (1) は次式（式 (6)）で表される。基底関数が Legendre 関数の場合，インデックスベクトル $\boldsymbol{k}_i$ の要素は，Legendre 関数を構成する Legendre 多項式の次数 $0, 1, 2, \cdots$ (付録 1 参照) に対応するため，0 を含む自然数となる。

$$\boldsymbol{y} = M^{\mathrm{T}}\boldsymbol{K}(\boldsymbol{x},\boldsymbol{\mathcal{Z}}) \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (6)$$

ここで，$M \stackrel{\text{def}}{=} [\boldsymbol{e}_1,\cdots,\boldsymbol{e}_{d_y}]$ は $N \times d_y$ の係数行列，$\|\boldsymbol{e}_1\| = ,\cdots, = \|\boldsymbol{e}_{d_y}\| = 1$ とする。

$\{K(\boldsymbol{x},\boldsymbol{k}_1),\cdots,K(\boldsymbol{x},\boldsymbol{k}_N)\}$ が正規直交基底の場合，式 (6) は多次元の Fourier 級数展開であり，係数行列 M は Fourier 係数となる。

しかし，通常の Fourier 級数の正規直交基底を用いた場合，標本データベクトル $\boldsymbol{x}$ の次元 d_x の増加に従い，必要となる基底関数の数（展開次数 N ）が指数関数的に増加する（付録 1，式 (付 9) 参照）。すなわち，高い精度を持つ写像を得るために必要となる展開次数 N の値は d_x に対して指数関数的に増加する。これは次元の呪いと呼ばれ非線形写像を構築する際の大きな問題点の一つである[7]。

3. 非線形写像のパラメータ最適化

本論文では，〈2·2〉節で述べた次元の呪いを回避し，適切な展開次数で十分な精度を持つ写像を構築するために，GA を用いて基底関数の数と次数を選択する[10]。また，写像データ空間の座標軸の物理的な意味を明確にするために，式 (6) の基底関数 $\{K(\boldsymbol{x},\boldsymbol{k}_1),\cdots,K(\boldsymbol{x},\boldsymbol{k}_N)\}$ を Legendre 関数 (付録 1 参照) とする。これにより，写像データベクトル $\boldsymbol{y}$ を標本データベクトル $\boldsymbol{x}$ の要素の冪乗で表せるため，標本データベクトルと写像データベクトルの関係が明確になる。さらに，目的関数が標本データベクトルの統計量で表されない場合にも柔軟に対応するために，目的関数の最適化問題を NMM[11] を用いて解く。以下に，その方法を述べる。

〈3·1〉 GA GA[8][9] は，最適化問題を解決するための進化的アルゴリズムの一つである。i 番目の個体は染色体に対応する $m_{\max}$ 次元のベクトル $\boldsymbol{g}_i$ を持ち，解の候補は各個体の染色体により表される。また，個体の優劣は $\boldsymbol{g}_i$ から計算される適応度 J_i で表される。GA は，交叉と突然変異により染色体を改良し，適応度が高い個体を選択する。そのため，解を大域的に探索し，局所解に陥りにくいという特徴がある。

本論文では，各個体の染色体 $\boldsymbol{g}_i$ の要素（遺伝子）をインデックスベクトル $\boldsymbol{k}_1 \sim \boldsymbol{k}_N$ の要素に対応させることにより，基底関数の数 N とインデックスベクトルを改良する。インデックスベクトルの要素は，〈2·2〉節で述べたように 0 を含む自然数であるため，染色体 $\boldsymbol{g}_i$ の m 番目の遺伝子 $g_{i;m}$ を $g_{i;m} \in \{0, 1, \cdots, g_{\max}\}$ で定義した。ここで，$g_{\max}$ はインデックスベクトルの要素の最大値，すなわち，Legendre 関数を構成する Legendre 多項式の最大次数に対応する。$g_{\max}$ を問題に応じて適切に設定し，解の探索空間を制限することにより，効率良く解を求められる。

GA の最も基本的な戦略である切捨て選択 (truncation selection)[12] では，個体の適応度が高い順に $\boldsymbol{g}_i$ をソートする。その後，$\boldsymbol{g}_1,\ldots,\boldsymbol{g}_{N_{\text{renew}}}$ を残し，$\boldsymbol{g}_{N_{\text{renew}}+1},\ldots,\boldsymbol{g}_{b_{\max}}$ を廃棄する。ここで，$b_{\max}$ は個体数である。次に，$\boldsymbol{g}_1,\ldots,\boldsymbol{g}_{N_{\text{renew}}}$ からランダムに選ばれた染色体を交叉することにより，新たに $\boldsymbol{g}_{N_{\text{renew}}+1},\ldots,\boldsymbol{g}_{b_{\max}}$ を作成する。その後，新たに作成された染色体の遺伝子は確率 p_{M} で突然変異を起こす。突然変異を起こした遺伝子の値は，$0, 1, 2, \cdots, g_{\max}$ の一様乱数で決められる。

〈3·2〉 基底関数のインデックスベクトルの設定 〈3·1〉節で述べたように，本論文では，染色体を表すベクトル $\boldsymbol{g}$ [†] の要素（遺伝子）をインデックスベクトル $\boldsymbol{k}_1 \sim \boldsymbol{k}_N$ の要素に対応させている。そのため，$\boldsymbol{g}$ の次元（遺伝子の数）$m_{\max}$ を標本データベクトル $\boldsymbol{x}$ の次元 d_x で割った値が，基底関数の数（展開次数）N が取り得る最大値 $N_{\max}$ となる (すなわち，$N_{\max} = m_{\max}/d_x$)。染色体 $\boldsymbol{g}$ を $N_{\max}$ 個に分割して作られた染色体の断片を $\check{\boldsymbol{g}}_j (1 \le j \le N_{\max})$ とする。$\check{\boldsymbol{g}}_j$ は d_x 次元のベクトルであり，本論文では，$\check{\boldsymbol{g}}_j$ をインデックスベクトルに対応させる。

〈3·1〉節で定義された染色体を通常の GA に従って改良した場合，一つの染色体の中に同一の染色体の断片が現れることがある。例えば，染色体の長さ $m_{\max} = 15$，$\boldsymbol{x}$ の次元 $d_x = 3$，展開次数の最大値 $N_{\max} = 5$，遺伝子の最大値（インデックスベクトルの要素の最大値）$g_{\max} = 5$ として，GA により $\boldsymbol{g} = (1,3,2,\ 0,5,1,\ 0,0,0,\ 1,3,2,\ 4,2,3)^{\mathrm{T}}$ が得られた場合，$\boldsymbol{g}$ と $\check{\boldsymbol{g}}_j$ の対応関係は，Fig. 1 の様になる。Fig. 1 から分かるように，$\check{\boldsymbol{g}}_1 = \check{\boldsymbol{g}}_4$ であり，これらは同一のインデックスベクトルに対応するため，一方は不要である。また，$\check{\boldsymbol{g}}_3$

[†] 煩雑な表記を避けるため，〈3·1〉節における染色体 $\boldsymbol{g}_i$ を，〈3·2〉節では，$\boldsymbol{g}$ と表記した。

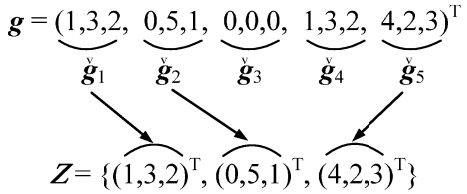


Fig. 1. Example of a map from $\boldsymbol{g}$ to $\mathcal{Z}$.

をインデックスベクトルとするLegendre関数は定数に対応するため，標本データベクトル $\boldsymbol{x} \in \mathcal{R}^{d_x}$ を写像データベクトル $\boldsymbol{y} \in \mathcal{R}^{d_y}$ に写像する際に，$\check{\boldsymbol{g}}_3$ は不要である。これら不要となる染色体の断片（Fig. 1の例では，$\check{\boldsymbol{g}}_3$ と $\check{\boldsymbol{g}}_4$）を除去し，$\boldsymbol{g}$ を基底関数のインデックスベクトル集合 $\mathcal{Z} \stackrel{\text{def}}{=} \{\boldsymbol{k}_1, \cdots, \boldsymbol{k}_N\}$ に写像するアルゴリズムを以下に示す。

〔**Algorithm1**〕　基底関数のインデックスベクトルの設定

(1-1) $\boldsymbol{g}$ から，$\check{\boldsymbol{g}}_1, \check{\boldsymbol{g}}_2, \cdots, \check{\boldsymbol{g}}_{N_{\max}}$ を作成し，$\check{\boldsymbol{G}} \stackrel{\text{def}}{=} \{\check{\boldsymbol{g}}_1, \check{\boldsymbol{g}}_2, \cdots, \check{\boldsymbol{g}}_{N_{\max}}\}$ とする。

(1-2) $\check{\boldsymbol{G}}$ から，$\check{\boldsymbol{g}}_j = \boldsymbol{0}$ なる $\check{\boldsymbol{g}}_j$ を削除する $(1 \le j \le N_{\max})$。

(1-3) $\check{\boldsymbol{G}}$ から，$\forall j > j'$ に対して，$\check{\boldsymbol{g}}_j = \check{\boldsymbol{g}}_{j'}$ なる $\check{\boldsymbol{g}}_j$ を削除する（$1 \le j' \le N_{\max}$）。

(1-4) N を $\check{\boldsymbol{G}}$ の要素数とする。

(1-5) $\check{\boldsymbol{G}}$ の要素を，順に $\boldsymbol{k}_1, \cdots, \boldsymbol{k}_N$ に割り当てる。

Fig. 1の例では，Algorithm 1の結果，$\check{\boldsymbol{g}}_1$，$\check{\boldsymbol{g}}_2$，$\check{\boldsymbol{g}}_5$ が $\boldsymbol{k}_1$，$\boldsymbol{k}_2$，$\boldsymbol{k}_3$ に割り当てられる。すなわち，$\boldsymbol{k}_1 = \check{\boldsymbol{g}}_1$，$\boldsymbol{k}_2 = \check{\boldsymbol{g}}_2$，$\boldsymbol{k}_3 = \check{\boldsymbol{g}}_5$，展開次数 $N = 3$ となる。このように，基底関数の数 N も，Algorithm 1によりインデックスベクトルを計算する際に，染色体 $\boldsymbol{g}$ に応じて決められる。

〈3·3〉　目的関数と基底関数　　複数のクラスに分けられた高次元の特徴量の相対的な違いを把握するために，2または3次元の相関図を作成する問題を考える。この問題は，高次元の特徴量が分布する空間を標本データ空間，2または3次元の相関図を写像データ空間として，式(6)の係数行列 $\boldsymbol{M}$，展開次数 N，およびインデックスベクトル集合 $\mathcal{Z}$ を求める問題である。この問題では，各クラスの分布が重複しないように写像データ空間に表示されることが重要である。また，標本データ空間における各クラスの位置関係が写像データ空間においてできるだけ保存されていることが望ましい。

そこで，本論文では，第1の目的を写像データ空間でクラスの集合が分離することとした。第1の目的が達成される写像にある程度の自由度がある場合には，その範囲内で，写像データ空間で各クラスの位置関係ができるだけ保存されることを第2の目的とした。これらの目的を達成するために，まず，写像データ空間において各クラスの分布が重複している程度を表す指標 $\varepsilon_{\text{class}}$，および標本データ空間における各クラスの位置関係と写像データ空間における各クラスの位置関係の違いを表す指標 $\varepsilon_{\text{dist}}$ を定義する。

$\bar{\boldsymbol{x}}^{(\ell)}$ を標本データ空間におけるクラス ℓ の重心（すなわち，クラス ℓ に属する標本データベクトルの平均），$\bar{\boldsymbol{y}}^{(\ell)}$ を写像データ空間におけるクラス ℓ の重心（すなわち，クラス ℓ に属する写像データベクトルの平均）とする。写像データ空間では，各クラスの境界を直感的に把握できることが望ましい。そこで，各クラスの境界がシンプルなマハラノビス距離に基づくパターン認識器[1]を採用した。写像データ空間における全標本点の共分散行列を S とすると，マハラノビス距離 $dist_{(n)}^{(\ell)}$ は次式で定義される。

$$dist_{(n)}^{(\ell)} \stackrel{\text{def}}{=} (\boldsymbol{y}_{(n)} - \bar{\boldsymbol{y}}^{(\ell)})^{\mathrm{T}} S^{-1} (\boldsymbol{y}_{(n)} - \bar{\boldsymbol{y}}^{(\ell)}) \cdots\cdots\cdots\cdots (7)$$

n 番目の写像データベクトル $\boldsymbol{y}_{(n)}$ は，$dist_{(n)}^{(\ell)} (1 \le \ell \le N_{\text{class}})$ が最小となるクラス $\ell_{\min(n)}$ に判別される（すなわち，$\ell_{\min(n)} = \arg[\min_{\ell}[dist_{(n)}^{(\ell)}]]$）。$\boldsymbol{y}_{(n)}$ が属するクラス番号 $a_{(n)}$ およびクロネッカーの記号 δ_{ij}（$i = j$ の場合1，それ以外は0）[13]を用いると，誤り数 $\varepsilon_{\text{class}}$ は，

$$\varepsilon_{\text{class}} = \sum_{n=1}^{N_{\text{data}}} (1 - \delta_{a_{(n)} \ell_{\min(n)}}) \cdots\cdots\cdots\cdots\cdots\cdots\cdots (8)$$

となる。本論文では，写像データ空間において各クラスの分布が重複している程度を表す指標として，上記誤り数 $\varepsilon_{\text{class}}$ を用いる。

次に，標本データ空間における各クラスの位置関係と写像データ空間における各クラスの位置関係の違いを示す指標 $\varepsilon_{\text{dist}}$ を定義する。一般に，クラス ℓ とクラス m のクラス間距離として，各々のクラスの重心の距離が用いられる。しかし，写像データ空間において，標本データ空間におけるデータの特徴を適切に把握するためには，重心の距離よりも，重心の距離の順序の方が影響が大きい場合がある。例えば，N_{class} 種類のクラス（1,2,⋯,N_{class}）の製品があり，本来はクラス1に属する製品を用いたいところであるが，クラス1に属する製品を購入できないため，代替え品を探すという問題を考える。この場合には，クラス1に近いクラスの中から製品を選択すればよいため，(1) クラス1と他の全てのクラスの重心の距離を測り，(2) 重心の距離が小さい順にクラスをソートし，(3) 重心の距離が最も小さいクラスに属する製品から優先的に採用するという手続きが取られる。この手続きでは，クラス1との重心の距離の順序が製品選択の基準になる。そのため，写像データ空間に写像されたデータにより上記手続きを実行する場合を考えると，写像データ空間において，標本データ空間における重心の距離の順序が保存されていることが望ましい。

本論文では，このような問題に対応することを目的として，標本データ空間における各クラスの位置関係と写像データ空間における各クラスの位置関係の違いを表す指標 $\varepsilon_{\text{dist}}$ を重心の距離の順序を基に定義する。まず，標本データ空間において，あるクラス ℓ に対するクラス $m \in \{1, \cdots, N_{\text{class}}\}$ の中での $\bar{\boldsymbol{x}}^{(\ell)}$ と $\bar{\boldsymbol{x}}^{(m)}$ の重心の距離の順序を $r_{\ell m}$ とする。例えば，$\bar{\boldsymbol{x}}^{(\ell)}$ と $\bar{\boldsymbol{x}}^{(m)}$ の距離が最も短い場合は $r_{\ell m} = 1$，2番目に短い場合は $r_{\ell m} = 2$，最も遠い場合は $r_{\ell m} = N_{\text{class}}$ となる。同様に，写像データ空間において，$\bar{\boldsymbol{y}}^{(\ell)}$ と $\bar{\boldsymbol{y}}^{(m)}$ の重

心の距離の順序を $\tilde{r}_{\ell m}$ とする。$r_{\ell m}$ を ℓ 行 m 列の要素とする行列 $\boldsymbol{R}$ を標本データ空間における距離順序行列，$\tilde{r}_{\ell m}$ を ℓ 行 m 列の要素とする行列 $\tilde{\boldsymbol{R}}$ を写像データ空間における距離順序行列として，本論文では，$\varepsilon_{\rm dist}$ を $\boldsymbol{R}$ と $\tilde{\boldsymbol{R}}$ の要素の最大誤差で定義する。

$$\varepsilon_{\rm dist} \stackrel{\rm def}{=} \max_{\ell m} |r_{\ell m} - \tilde{r}_{\ell m}| \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (9)$$

ここで，$\max_{\ell m}(\cdot)$ は，$\ell, m \in \{1, \cdots, N_{\rm class}\}$ に関して最大値を取る関数である。

本節の冒頭で述べた本論文の目的（クラスの集合が分離すること，および各クラスの位置関係が保存されること）を達成するために，上記，$\varepsilon_{\rm class}$ および $\varepsilon_{\rm dist}$ を用いて目的関数を計算する手順を Algorithm 2 に示す。

〔**Algorithm2**〕　目的関数 $f(M|\mathcal{D}, N, \mathcal{Z})$

(2-1)　式 (6) により $\mathcal{D}$ から $\tilde{\mathcal{D}}$ を作成する。

(2-2)　$\tilde{\mathcal{D}}$ をパターン認識器に学習させ，式 (8) により誤り数 $\varepsilon_{\rm class}$ を計算する。

(2-3)　$\tilde{\boldsymbol{R}}$ と $\boldsymbol{R}$ との最大誤差 $\varepsilon_{\rm dist}$ を式 (9) により計算する。

(2-4)　$f(M|\mathcal{D}, N, \mathcal{Z}) = \varepsilon_{\rm class} + w_{\rm dist}\varepsilon_{\rm dist}$

ここで，$w_{\rm dist} = 1/N_{\rm class}$ とすることにより，$w_{\rm dist}\varepsilon_{\rm dist} < 1$ となる。また，$\varepsilon_{\rm class}$ は自然数である。そのため，$\varepsilon_{\rm dist}$ よりも $\varepsilon_{\rm class}$ が優先的に最小化される。

また，特徴量の違いを直感的に把握するためには，写像データ空間の座標軸と標本データ空間の座標軸の関係が明確であることが望ましい。本論文では，式 (6) の基底関数 $\{K(\boldsymbol{x}, \boldsymbol{k}_1), \cdots, K(\boldsymbol{x}, \boldsymbol{k}_N)\}$ を Legendre 関数 (付録 1 参照) とする。これにより，写像データベクトル $\boldsymbol{y}$ を標本データベクトル $\boldsymbol{x}$ の要素の冪乗で表せるため，標本データベクトルと写像データベクトルの関係が明確になる。

〈3·4〉　GA と NMM を用いた非線形写像の最適化

〈2·2〉節で述べた写像および〈3·2〉節で述べた基底関数のインデックスベクトルの設定手法を GA に適用し，非線形写像のパラメータ（M_i，N_i，$\mathcal{Z}_i$）を最適化する。ここで，M_i は i 番目の染色体に対する式 (6) の係数行列，N_i は i 番目の染色体に対する展開次数，$\mathcal{Z}_i$ は i 番目の染色体に対するインデックスベクトル集合である。

〔**Algorithm3**〕　非線形写像の最適化

(3-1)　染色体 $\boldsymbol{g}_1, \cdots, \boldsymbol{g}_{b_{\max}}$ を初期化する。

(3-2)　Algorithm 1 に $\boldsymbol{g}_i$ と $N_{\max}$ を入力し，N_i および $\mathcal{Z}_i$ を得る $(1 \le i \le b_{\max})$。

(3-3)　Algorithm 2 で定義された目的関数 $f(M_i|\mathcal{D}, N_i, \mathcal{Z}_i)$ を M_i について最小化し，$J_i = -f(M_i|\mathcal{D}, N_i, \mathcal{Z}_i)$ とする $(1 \le i \le b_{\max})$。

(3-4)　適応度の値により $J_1, \cdots, J_{b_{\max}}$ および $\boldsymbol{g}_1, \cdots, \boldsymbol{g}_{b_{\max}}$ をソートする。

(3-5)　$J_1, \cdots, J_{b_{\max}}$ が収束していれば終了。そうでなければ，Step (3-6) へ。

(3-6)　一点交叉により $\boldsymbol{g}_1, \cdots, \boldsymbol{g}_{N_{\rm renew}}$ を交叉し，次世代の染色体 $\boldsymbol{g}_{N_{\rm renew}+1}, \cdots, \boldsymbol{g}_{b_{\max}}$ を作成する。交叉点は，2 から $m_{\max} - 1$ までの一様乱数で設定される。

(3-7)　$\boldsymbol{g}_{N_{\rm renew}+1}, \cdots, \boldsymbol{g}_{b_{\max}}$ の遺伝子を突然変異により変更する。

(3-8)　Step (3-2) へ戻る。

Algorithm 3 により適応度が最も大きい 1 番目の個体に対応するパラメータ M_1，N_1，および $\mathcal{Z}_1$ を式 (6) に適用することにより，最適な非線形写像が得られる。

また，$b_{\max} = 1$，$N_{\max} = N = d_{\rm x}$, $\boldsymbol{k}_1 = (1, 0, 0, \cdots, 0)^{\rm T}$, $\boldsymbol{k}_2 = (0, 1, 0, \cdots, 0)^{\rm T}, \cdots$, $\boldsymbol{k}_{d_{\rm x}} = (0, 0, 0, \cdots, 1)^{\rm T}$ として，上記アルゴリズムを Step(3-3) で終了すると，$\boldsymbol{k}_1$，$\cdots$，$\boldsymbol{k}_{d_{\rm x}}$ に対応する基底関数（Legendre 関数）が線形であるため，最適な線形写像が得られる。

〈2·1〉節で述べたように，目的関数を統計量で表した場合，データの分布が想定と異なる場合には，必ずしも本来の目的が達成されるとは限らない。また，クラスができるだけ分離するという要求のみならず，高次元空間でのクラス間距離の順序が保存されるといったような要求がある場合は，目的関数が統計量で表されるとは限らない。そのため，最適化問題を効率的な解法が存在する固有値問題の様な問題に帰着させられるとは限らない。

そこで，本論文では，様々な要求に柔軟に対応するために，上記アルゴズムの Step(3-3) において，係数行列 M_i の要素を $N \times d_{\rm y}$ 次元の変数，$\|\boldsymbol{e}_1\| =, \cdots, = \|\boldsymbol{e}_{d_{\rm y}}\| = 1$ を制約条件として，目的関数 $f(M_i|\mathcal{D}, N_i, \mathcal{Z}_i)$ が最小となる M_i の要素を，Nelder-Mead 法 (NMM) [11] で求める。

4.　性能評価

本章では，様々な標本データベクトル集合の相関図における，クラス判別誤り率 $r_{\rm class}$（$\stackrel{\rm def}{=}$(クラス判別誤り数 $\varepsilon_{\rm class}$(式 (8))/ 総データ数)×100[%]) と距離順序行列の最大誤差 $\varepsilon_{\rm dist}$（式 (9)）を調べることにより，提案手法 Algorithm 3 の有効性を示す。

〈4·1〉　正準判別分析と Algorithm 3 の比較

本節では，3 次元，5 クラス，各クラスのデータ数 40，総データ数 200 の標本データベクトル集合を用いて，正準判別分析 (CDA) [1] と Algorithm 3 の目的関数の違いが性能に与える影響を調べる。ここで，標本データは，Table 1 に示す一様乱数により作成され，その分布は Fig. 2 の通りである。また，Table 1 の $U(a, b)$ は，a から b の一様乱数である。

Table 1.　Uniform random variables used for experimental data.

axis	class 1	class 2	class 3	class 4	class 5
x_1	U(0.0,0.3)	U(0.6,0.9)	U(0.0,0.3)	U(0.0,0.3)	U(1.0,1.3)
x_2	U(0.0,0.3)	U(0.0,0.3)	U(0.6,0.9)	U(0.0,0.3)	U(1.0,1.3)
x_3	U(0.0,0.3)	U(0.0,0.3)	U(0.0,0.3)	U(1.0,1.3)	U(0.0,0.3)

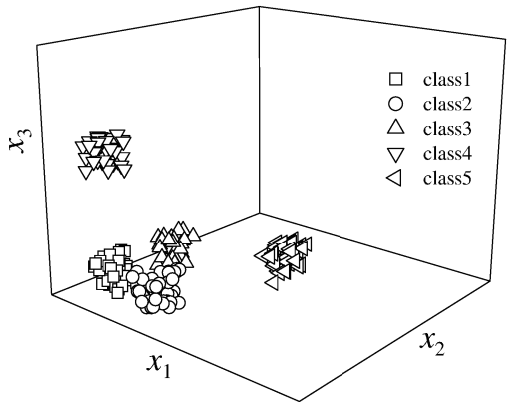


Fig. 2. Three-dimensional distribution with five classes.

Table 2. Difference between CDA and Algorithm 3 (linear) due to objective function.

method	objective function	r_{class}[%]	$\varepsilon_{\text{dist}}$
CDA	λ_1 and λ_2	6.5	1
Algorithm 3 (linear)	$\varepsilon_{\text{class}}$	0	3
Algorithm 3 (linear)	$\varepsilon_{\text{class}} + w_{\text{dist}}\varepsilon_{\text{dist}}$	0	0

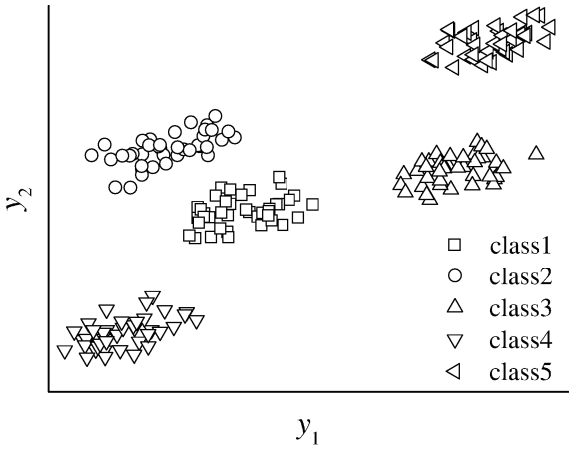


Fig. 3. Map of three-dimensional distribution on two-dimensional space obtained by Algorithm 3 (linear) (objective function: $\varepsilon_{\text{class}} + w_{\text{dist}}\varepsilon_{\text{dist}}$).

CDA と Algorithm 3 を比較するために，本節では，Algorithm 3 が最適化する写像は CDA と同様に線形（〈3·4〉節参照）とする（以下，Algorithm 3 (linear) と記す）。また，Algorithm 3 の目的関数における $\varepsilon_{\text{dist}}$ の効果を検証するために，Algorithm 2 の Step(2-4) において，$f(M|\mathcal{D}, N, \mathcal{Z}) = \varepsilon_{\text{class}}$ として（すなわち，目的関数を $\varepsilon_{\text{class}}$ として）r_{class} と $\varepsilon_{\text{dist}}$ を調べる。その結果，Table 2 に示すように，Algorithm 3 (linear) では 2 種類の目的関数のいずれの場合でも $r_{\text{class}} = 0\%$ であるにも拘わらず，CDA では 6.5% となった。また，Algorithm 3 において，目的関数に $\varepsilon_{\text{dist}}$ を導入することにより，$\varepsilon_{\text{dist}}$ が 3 から 0 に減少した。目的関数を $\varepsilon_{\text{class}} + w_{\text{dist}}\varepsilon_{\text{dist}}$ として Algorithm 3 (linear) により得られた写像ベクトル集合を Fig. 3 に示す。

他の様々な標本データベクトル集合に対して同様な実験を行った結果，Algorithm 3 の r_{class} が 0% の場合でも，CDA では 0% とならない場合があった。しかし，その逆の例は現時点では見つかっていない。これらの結果から，$\varepsilon_{\text{class}}$ を目的関数としている Algorithm 3 は，群間変動と郡内変動の比（λ_1 および λ_2）を目的関数としている CDA と比較して，クラス分離性能が優れていることが分かった。また，$\varepsilon_{\text{dist}}$ を目的関数に導入することにより，r_{class} を悪化させることなく $\varepsilon_{\text{class}}$ を改善できることが分かった。

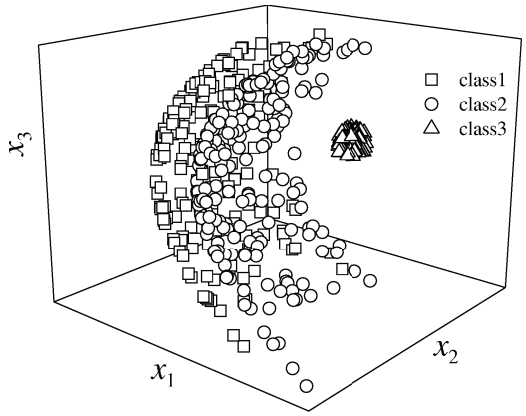


Fig. 4. Nonlinear three-dimensional distribution with three classes.

Table 3. Uniform random variables and equations used for experimental data.

axis	class 1	class 2	class 3
x_1	$(x_2-1)^2+(x_3-1)^2$	$(x_2-1)^2+(x_3-1)^2+0.5$	$U(1.0,1.2)$
x_2	$U(0.0,1.6)$	$U(0.0,1.6)$	$U(1.8,2.0)$
x_3	$U(0.0,2.0)$	$U(0.0,2.0)$	$U(0.9,1.1)$

Table 4. Difference between Algorithm 3 (linear) and Algorithm 3 (nonlinear) for nonlinear distribution.

method	objective function	r_{class}[%]	$\varepsilon_{\text{dist}}$
Algorithm 3 (linear)	$\varepsilon_{\text{class}} + w_{\text{dist}}\varepsilon_{\text{dist}}$	20.2	0
Algorithm 3 (nonlinear)	$\varepsilon_{\text{class}} + w_{\text{dist}}\varepsilon_{\text{dist}}$	0	0

〈4·2〉 非線形最適化の効果の検証 Algorithm 3 の非線形最適化の効果を確認するために，Fig. 4 に示すような非線形に歪んだ分布を持つ標本データ集合を用いて Algorithm 3 を評価した。標本データ集合は，Table 3 に示す一様乱数と方程式により作成され，3 次元，3 クラス，各クラスのデータ数 200 である。その結果を，Table 4 に示す。ここで，Table 4 において，Algorithm 3 (linear) は〈4·1〉節と同様に Algorithm 3 による線形最適化を，Algorithm 3 (nonlinear) は Algorithm 3 による非線形最適化（最大次数 $g_{\max} = 2$，展開次数の最大値 $N_{\max} = 8$，突然変異の確率 $p_{\text{M}} = 10^{-3}$）を表す。Table 4 から分かるように，線形最適化と非線形最適化の両者とも $\varepsilon_{\text{dist}} = 0$ である。また，線形最適化により得られた r_{class} は非常に大きいにも拘わらず，非線形最適化では $r_{\text{class}} = 0\%$ になっている。さらに，Fig. 5 に示した非線形最適化により得られた相関図からも，Fig. 4 に示すような非線形に歪んだ分布を持つ標本データ集合がきれいに分離していることが分かる。これらの結果から，Algorithm 3

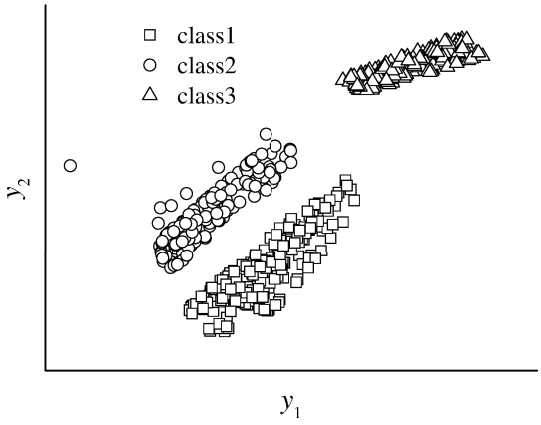


Fig. 5. Map of nonlinear three-dimensional distribution on two-dimensional space obtained by Algorithm 3 (nonlinear).

Table 5. Difference between Algorithm 3 (linear) and Algorithm 3 (nonlinear) for sake data.

method	objective function	r_{class}[%]	$\varepsilon_{\mathrm{dist}}$
Algorithm 3 (linear)	$\varepsilon_{\mathrm{class}} + w_{\mathrm{dist}}\varepsilon_{\mathrm{dist}}$	5.9	2
Algorithm 3 (nonlinear)	$\varepsilon_{\mathrm{class}} + w_{\mathrm{dist}}\varepsilon_{\mathrm{dist}}$	0	3

による非線形最適化の効果を確認できた。

〈4･3〉 清酒データの相関図 2015～2016年に製造され，製法（精米歩合，添加物）と味（甘口/ 辛口）の組合せにより7クラスに分類された清酒82サンプルの中から，標本データ空間でクラスが分離しているサンプルとして計51サンプルを選び，これらを標本データ集合として，Algorithm 3を用いて清酒の標本データ集合の相関図を作成した。ここで，清酒の特徴を表す標本データベクトルは，味覚センサ[14]により測定された，酸味，苦味雑味，渋味刺激，旨味，塩味，苦味，渋味，旨味コク，および化学分析により測定されたグルコース濃度を要素とする9次元のベクトルである（標本データ集合とクラス分類の詳細は，付録2参照）。その結果得られた r_{class} と $\varepsilon_{\mathrm{dist}}$ を Table 5 に示す。ここで，Table 5 において，Algorithm 3 (linear) は〈4･1〉節と同様にAlgorithm 3による線形最適化を，Algorithm 3 (nonlinear) は Algorithm 3 による非線形最適化（最大次数 $g_{\max} = 2$，展開次数の最大値 $N_{\max} = 16$，突然変異の確率 $p_{\mathrm{M}} = 10^{-3}$）を表す。

Table 5 から分かるように，線形最適化では $r_{\mathrm{class}} = 5.9\%$ であるにも拘わらず，非線形最適化では0% である。すなわち，線形最適化では分離できなかった清酒データの分布が，非線形最適化により分離できた。一方，$\varepsilon_{\mathrm{dist}}$ は，線形最適化より非線形最適化の方が大きくなっている。これは，Algorithm 3 の目的関数では，$\varepsilon_{\mathrm{class}}$(すなわち r_{class}) が $\varepsilon_{\mathrm{dist}}$ より優先されているためである。

Fig. 6 は，非線形最適化により得られた相関図である。最適な非線形写像を構成するインデックスベクトル集合 $\mathcal{Z}_1$，係数行列 M_1，および展開次数 N_1 から，Fig. 6 の横軸の主な要素は苦味と甘味，縦軸の主な要素は酸味と塩味となった。Fig. 6 における清酒の7つのクラスの分布，および縦軸と横軸の主な構成要素に関する考察を以下にまとめる。

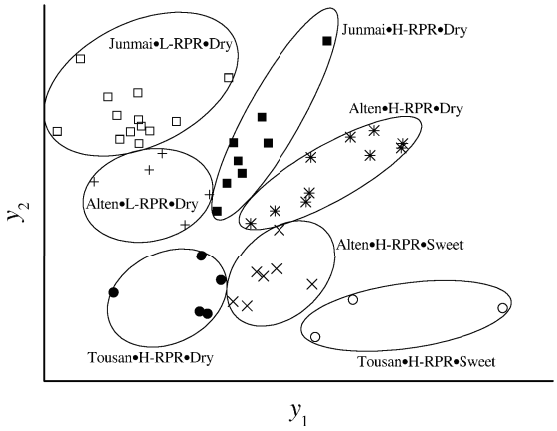


Fig. 6. Map of nine-dimensional sake data on two-dimensional space obtained by Algorithm 3 (nonlinear).

- 苦味と甘味が同じ軸（横軸）の主な構成要素となったことは，今回用いられたデータでは両者に相関があることを示している。これは，清酒における経験的な知見（両者には負の相関があること）と一致する。
- 塩味センサは有機酸などにも反応するため，塩化ナトリウムがほとんど存在しない清酒の分析では経験的に味の濃さとされる。一方，清酒における味の濃さと酸味には相関があることが示されており[16]，そのために酸味と塩味が同じ軸（縦軸）の主な構成要素となったと考えられる。
- 純米酒が相関図の上に分布し，糖類酸味料添加酒が下に分布していることは，清酒における経験的な知見（酸味や味の濃さ（塩味センサ分析値）は純米酒で強く，味の濃さ（塩味センサ分析値）は糖類酸味料添加酒では低い）と一致する。
- 甘口/ 辛口の味の違いは，苦味，甘味，酸味，および塩味の総合評価として，右斜め下の軸になっている。
- 辛口同士，甘口同士が隣接している。また，縦軸に沿って，純米吟醸酒→吟醸酒（アルコール添加）→糖類酸味料添加酒，純米酒→アルコール添加酒→糖類酸味料添加酒などのように清酒のランク順に各クラスが並んでいる。
- 経験的に味が似ていると感じられるクラスが隣接している。

これらの結果から，提案手法により得られた相関図の座標軸は明解であり，経験や分析により得られた知見と矛盾しないことが分かる。

5. ま と め

高次元の特徴量の分布を視覚的に把握するために，高次元特徴空間から低次元の相関図への非線形写像を構築するアルゴリズムを開発した。提案手法では，非線形写像を構築する際の次元の呪いの問題を回避し，相関図の座標軸と

高次元の特徴量の関係を明確にするために，非線形写像をLegendre 関数の和で定義し，最適な Legendre 関数の組合せを GA を用いて求めた。また，必ずしも統計量で表現できるとは限らない様々な要求に応じて相関図を作成するために，非線形写像を構成する Legendre 関数の係数を NMM により求めた。その結果，統計量を目的関数としたアルゴリズムと比較して，相関図に対する目的を高精度で達成できた。また，データと客観的な目的関数のみに基づき，手作業を排除して，誰が見ても納得できる相関図を作成できることを示すために，提案手法を 7 クラスに分類された清酒データに適用した。その結果，座標軸が明解かつ経験や分析により得られた知見と矛盾しない相関図を作成できた。

今後，清酒データの数を増やし，相関図の汎用性を高めていく。また，提案手法を様々な製品の評価や品質管理に適用するために，相関図の作成および未評価の製品の評価（データを相関図に写像し，製品の品質を視覚的に把握する）等の一連の作業を効率化するためのインタフェースを開発する予定である。

文　　献

（1） 小西貞則：「多変量解析入門」，岩波書店 (2010)
（2） R. O. Duda, P. E. Hart, and D. G. Stork: “Pattern Classification, 2nd Edition”, Jhon Wiley & Suns (2000)
（3） R. A. Johnson and D. W. Wichern : “Applied Multivariate Statistical Analysis 5th ed.”, Pearson Education (Prentice Hall), USA (2001)
（4） S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K. R. Muller: “Fisher Discriminant Analysis with Kernels”, In IEEE Neural Networks for Signal Processing Workshop, pp. 41–48 (1999)
（5） J. B. Tenenbaum, V. de Silva, and J. C. Langford:“A Global Geometric Framework for Nonlinear Dimensionality Reduction”, Science 22, Vol.290, Issue 5500, pp. 2319–2323 (2000)
（6） B. Scholkopf, A. Smola, and KR Muller: “Nonlinear Component Analysis as a Kernel Eigenvalue Problem”, Neural Computation, Vol.10, pp. 1299–1319 (1998)
（7） R. G. Sutton and A. G. Barto : “Reinforcement Learning”, MIT Press, USA (1998)
（8） D. E. Goldberg: “Genetic Algorithms in Search Optimization and Machine Learning”, Addison-Wesley, USA (1989)
（9） J. H. Holland: “Genetic Algorithm”, Scientific American, pp.66–72 (1992)
（10） H. Satoh, D. Kasai, and M. Satoh: “Characteristic Collection of Taste Sensor Based on Sensory Evaluation and its Application to Food Discrimination”, IEEJ Trans. on Sensors and Micromachines, Vol.136, No.7, pp. 303–311 (2016) (in Japanese)
佐藤仁樹・葛西大介・佐藤雅子：「官能評価に基づく味覚センサの特性補正と食品識別への応用」，電学論 E，Vol.136, No.7, pp. 303–311 (2016)
（11） J. A. Nelder and R. Mead: “A Simplex Method for Function Minimization”, Computer Journal Vol.7, Issue 4, pp. 308–313 (1965)
（12） H. Muhlenbein and D. Schlierkamp-Voosen: “Predictive Models for the Breeder Genetic Algorithm I. Continuous Parameter Optimization”, Evolutionary Computation, Vol.1, No.1, pp. 25–49 (1993)
（13） I. N. Bronshtein and K. A. Semendyayev: “Handbook of Mathematics”, Springer-Verlag, UK (1997)
（14） K. Toko: “Biochemical Sensors: Mimicking Gustatory and Olfactory Senses: Order”, Pan Stanford Publishing (2013)
（15） 石川雄章編：「増補改訂清酒製造技術 新版」，日本醸造協会 (2009)
（16） S. Sato, H. Kawashima, and Y. Maruyama: “Studies on the Taste of Sake Part III. Application of Regression Models Relating Sweetness, Fullness and Chemical Date”, J. Soc. Brew. Japan, Vol.69, No.11, pp. 774–777 (1974) (in Japanese)
佐藤　信・川島　宏・丸山良光：「清酒の味覚に関する研究 (第 3 報) 甘辛と濃さに関する重回帰式」，J. Soc. Brew. Japan, Vol.69, No.11, pp. 774–777 (1974)
（17） H. Utsunomiya, A. Isogai, and H. Iwata: “Amakara Categories for Type Designation”, J. of Brewing Society of Japen, Vol.99, No.12, pp. 882–889 (2004) (in Japanese)
宇都宮仁・磯谷敦子・岩田　博：「清酒の甘辛区分表示について」，日本醸造 協会誌，Vol.99, No.12, pp. 882–889 (2004)

付　　録

1. 正規直交基底

$\boldsymbol{x} \overset{\text{def}}{=} (x_1, \cdots, x_{d_x})^{\mathrm{T}}$，$\mathcal{D}_{\mathbf{x}} \overset{\text{def}}{=} \{\boldsymbol{x} | x_{\min d} \le x_d \le x_{\max d},\ 1 \le d \le d_x\}$ とする。本節では，$\boldsymbol{x} \in \mathcal{D}_{\mathbf{x}}$ の正規直交基底について述べる。$\boldsymbol{k} \overset{\text{def}}{=} (k_1, \cdots, k_{d_x})^{\mathrm{T}} \in \mathcal{Z}$ をインデックスベクトル，$\mathcal{Z}$ を $\boldsymbol{k}$ の集合，$h(\boldsymbol{k})$ を $\boldsymbol{k}$ に対する Fourier 係数とする。関数 $f(\boldsymbol{x})$ の Fourier 級数展開は，次式により定義される [13]。

$$f(\boldsymbol{x}) = \sum_{\boldsymbol{k} \in \mathbf{Z}} h(\boldsymbol{k}) K(\boldsymbol{x}, \boldsymbol{k}) \cdots\cdots (\text{付 } 1)$$

$$h(\boldsymbol{k}) \overset{\text{def}}{=} \int_{\mathcal{D}_{\mathbf{x}}} f(\boldsymbol{x}) K^*(\boldsymbol{x}, \boldsymbol{k}) \mathrm{d}\boldsymbol{x} \cdots\cdots (\text{付 } 2)$$

ここで，上付き添え字 $*$ は複素共役，$\{K(\boldsymbol{x}, \boldsymbol{k})\}$ は多次元正規直交基底である。$\{K_d(x_d, k_d)\}$ を 1 次元の正規直交基底とすると，$\{K(\boldsymbol{x}, \boldsymbol{k})\}$ は次式で定義される。

$$K(\boldsymbol{x}, \boldsymbol{k}) \overset{\text{def}}{=} \prod_{d=1}^{d_x} K_d(x_d, k_d) \cdots\cdots (\text{付 } 3)$$

$\{K_d(x_d, k_d)\}$ が Legendre 関数を用いた正規直交基底の場合，$K_d(x_d, k_d)$ は次式により定義される。

$$K_d(x_d, k_d) = \sqrt{\frac{2k_d+1}{D_d}} P\left(2\frac{x_d - x_{\min d}}{D_d} - 1, k_d\right) \cdots\cdots (\text{付 } 4)$$

ここで，$D_d \overset{\text{def}}{=} x_{\max d} - x_{\min d}$ である。また，$P(x, k)$ は $x \in [-1, 1]$ の Legendre 多項式であり，次式で表される。

$$\begin{cases} P(x, 0) = 1, \\ P(x, 1) = x, \\ P(x, 2) = (3x^2 - 1)/2, \\ P(x, 3) = (5x^3 - 3x)/2, \\ \quad \vdots \end{cases} \cdots\cdots (\text{付 } 5)$$

基底関数 $\phi_i(\cdot)$ を次式で定義する。

$$\phi_i(\boldsymbol{x}) \overset{\text{def}}{=} K(\boldsymbol{x}, \boldsymbol{k}) \cdots\cdots (\text{付 } 6)$$

ここで，i は基底のインデックスである。

通常の Fourier 級数の場合，N_d を x_d の展開次数とすると，$\mathcal{Z}_d \overset{\text{def}}{=} \{0, 1, \cdots, N_d\}$，$\mathcal{Z}$ は $\mathcal{Z}_d$ の直積により $\mathcal{Z} \overset{\text{def}}{=} \mathcal{Z}_1 \times \mathcal{Z}_2 \times, \cdots, \times \mathcal{Z}_{d_x}$ で与えられるため，i は $\boldsymbol{k}$ を用いて次式で表される。

$$i = \sum_{d=1}^{d_x} k_d \prod_{d'=d+1}^{d_x} (N_{d'} + 1) \cdots\cdots (\text{付 } 7)$$

すなわち，通常の Fourier 級数は，Fourier 係数を α_i，基底関数 $\{\phi_i(\boldsymbol{x})\}$ を式 (付 3)，(付 6)，(付 7) で表される正規直交基底で与えることにより，

app. Table 1. Definitions of brewing method and taste of sake.

Label	Definition
Alten	includes brewer's alcohol
Junmai	made from only rice and water
Tousan	includes brewer's alcohol, saccharides, organic acids, amino acid salts
L-RPR	ginjō-shu or daiginjō-shu; rice-polishing ratio below 60%
H-RPR	rice-polishing ratio greater than 60%
Dry	AV=Glc.- TA ≤ 1.0
Sweet	AV=Glc.- TA > 1.0

app. Table 2. Classes and numbers of samples of sake.

Class information		Number of samples
Class No.	Combination of brewing method and taste	
1	Alten・L-RPR・Dry	5
2	Alten・H-RPR・Sweet	7
3	Alten・H-RPR・Dry	10
4	Junmai・L-RPR・Dry	13
5	Junmai・H-RPR・Dry	8
6	Tousan・H-RPR・Sweet	3
7	Tousan・H-RPR・Dry	5

$$f(\boldsymbol{x}) = \sum_{i=0}^{N} \alpha_i \phi_i(\boldsymbol{x}) \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \text{(付 8)}$$

で定義される。また，$\boldsymbol{x}$ の展開次数 N は次式により得られる。

$$N = \prod_{d=1}^{d_x} (N_d + 1) - 1 \cdots\cdots\cdots\cdots\cdots\cdots\cdots \text{(付 9)}$$

2. 清酒の標本データ

〈2･1〉 清酒のクラス 清酒は，製法品質表示基準に示された精米歩合や麹米の使用割合，および醸造アルコールなどの使用の有無などにより，8 種類の特定名称と特定名称がつけられないものに分けられている[15]。さらに，味わいに関する区分の仕方として，これまでに濃淡度[16]，甘辛度[16]，新甘辛度[17] などが提案されてきた。本論文ではこれらのうち，精米歩合，添加物，新甘辛度を組み合わせて Table 2 の様に清酒を 7 クラスに分類した。ここで，Table 1 において，AV は新甘辛度[17]，Glc[g/dL] は化学分析によって得られたグルコース濃度，TA[mL] は酸度を表す。

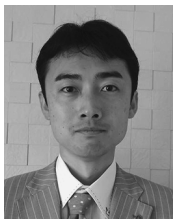
〈2･2〉 清酒の標本データ空間 本論文では，味覚センサ[14] により測定された，酸味，苦味雑味，渋味刺激，旨味，塩味，苦味，渋味，旨味コク，および化学分析により測定されたグルコース濃度† により清酒の味を表し，これらを標本データベクトル $\boldsymbol{x}$ の要素とする。

まず，あらかじめクラス分けされた清酒 82 サンプルに対してこれらの味データを測定した。次に，〈3･3〉節に記載のマハラノビス距離に基づくパターン認識器を用いてクラスを識別した。最後に，識別誤りを起こしたサンプルをあらかじめ用意された 82 サンプルから取り除き，その結果残った 51 サンプルを標本データ集合 $\mathcal{D} \stackrel{\text{def}}{=} \{(\boldsymbol{x}_{(n)}, a_{(n)}) | n = 1, \cdots, N_{\text{data}}\}$ とした。すなわち，クラス数 $N_{\text{class}} = 7$，$\boldsymbol{x}_{(n)}$ の標本数 $N_{\text{data}} = 51$ である。各クラスにおける製法と味の組合せ，および各クラスの標本数 $N_{\text{data}}^{(\ell)}$ は Table 2 の通りである。この操作により，標本データ空間では，各クラスの清酒は完全に分離している。

† 甘味センサを使用する際には，試料に加水してアルコール度数を合わせる必要がある。加水により味自体が変化するため，甘味センサはアルコール飲料の甘味の測定には適さない。一般に清酒の甘味はグルコースによるところが大きいため，甘味センサの代わりに化学分析により測定されたグルコース濃度を採用した。

佐 藤 仁 樹（正員） 1987 年早稲田大学理工学研究科修士課程修了。同年㈱東芝研究開発センター入社。音声のパケット化，ATM 網のトラヒック制御，およびインターネットの輻輳制御の研究に従事。2000 年 4 月より㈱ワイ・アール・ビー移動通信基盤技術研究所に出向。移動通信網の送信電力制御，輻輳制御，およびインターネット TV 会議システムの研究に従事。2002 年より公立はこだて未来大学にて，非線形システムの解析および最適化の研究に従事。博士 (情報科学) 早稲田大学。

佐 藤 雅 子（非会員） 1989 年筑波大学第三学群情報学類卒業。同年㈱東芝研究開発センター入社。ATM 交換機の OS 開発およびモバイル端末のアプリケーション開発に従事。1997 年同社退職。2012 年より㈱インテリジェントセンサーテクノロジーテクニカルスタッフ。情報ノ宮蕗の下工房にて，原料・ブレンド比最適化ツール開発に従事。



高 尾 佳 史（非会員） 2007 年山口大学大学院農学研究科修士課程修了。同年菊正宗酒造㈱入社。総合研究所において樽酒や酵母に関する研究に従事。現在に至る。博士 (生命科学) 山口大学。