

# 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

正員 佐藤 仁樹\* 非会員 葛西 大介\*\* 非会員 佐藤 雅子\*\*\*

## Characteristic Collection of Taste Sensor Based on Sensory Evaluation and its Application to Food Discrimination

Hideki Satoh\*, Member, Daisuke Kasai\*\*, Non-member, Masako Satoh\*\*\*, Non-member

(2015年10月21日受付, 2016年1月21日再受付)

A function that expresses the relationship between data obtained from a sensory evaluation and those obtained from a taste sensor was approximated using a polynomial. First, the basis functions of a polynomial were set to a Legendre function, and the number of basis functions and their degrees were selected using a genetic algorithm. Next, the coefficients of the polynomial were calculated using multiple regression analysis. The approximation error was estimated using cross-validation, and the number of basis functions and their degrees were optimized so as to maximize the generalization of the polynomial. As a result, sufficiently small approximation errors were obtained for test data, and the explicit relationship between the data obtained from a sensory evaluation and those obtained from a taste sensor were clarified using the polynomial. The discrimination rate for food was improved using the estimated data of a sensory evaluation compared to a method that used only the data obtained from a taste sensor.

キーワード: 味覚センサ, 関数近似, 多項式, 基底関数, 遺伝的アルゴリズム, パターン認識

**Keywords:** taste sensor, function approximation, polynomial, basis, genetic algorithm, pattern recognition

### 1. はじめに

食品の品質管理や賞味期限の設定など, 食品の製造販売には, 味の評価が重要である。食品業界では, パネルと呼ばれる官能評価のプロフェッショナルが実際に食品を試食して, 味の評価を行う。パネルは, 非常に繊細な味の差も識別できる。しかし, 味の感じ方には, 年齢, 地域, 食文化の違いがある。そのため, 優れたパネルでもこれらの全てに対応することは困難であり, パネル間でも意見の食い違いがしばしば発生する。

この問題を解決するために, (株)インテリジェントセンサー

テクノロジーにより味覚センサが開発された。味覚センサにより, 対象となる製品の甘味, 塩味, 旨味, 酸味などの味の強さを数値(味データ)として客観的に表現できる。そのため, 現在, 味覚センサは, 食品, 飲料, 医薬品などの様々な分野で, 研究は元より, マーケティング, 商品開発, 営業, および品質保証のために幅広く活用されている<sup>(1)</sup>。

(株)インテリジェントセンサーテクノロジーにより開発された味覚センサの精度および再現性は非常に高い<sup>(2)(3)</sup>。しかし, その一方で, 大きく異なる味の相対評価を行う際には, パネルによる官能評価の結果と異なる場合がある。この問題は, 味覚センサにより測定された甘味, 塩味, 旨味, 酸味等の味データ(説明変数)と, パネルによる官能評価のデータ(目的変数)の関係が明らかになれば解決できる。また, この関係が明らかになれば, 官能評価の代わりに味覚センサの測定結果を利用できる。そのため, 官能評価に伴う作業の繁雑さを軽減し, パネルの心理・健康状態に左右されない安定した評価が可能となる。

観測された目的変数と説明変数のデータから, データを生成したシステムの構造やパラメータを推定する問題は, 関数近似あるいはシステム同定と呼ばれ, 予測, パターン認識, 制御など, 様々な分野に係わる重要な問題である。この問題を解決するために, ニューラルネットワーク<sup>(4)</sup>, サポートベクターマシン(SVM)<sup>(5)</sup>, 遺伝的アルゴリズム(GA)<sup>(6)(7)</sup>

\* 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

\*\* 公益財団法人 とかち財団  
〒080-2462 北海道帯広市西 22 条北 2-23-10  
Tokachi Foundation

West22-North2-23-10, Obihiro, Hokkaido 080-2462, Japan

\*\*\* (株)インテリジェントセンサーテクノロジー  
〒243-0032 神奈川県厚木市恩名 5-1-1

Intelligent Sensor Technology, Inc.

5-1-1, Onna, Atsugi, Kanagawa 243-0032, Japan

等, 様々な手法が用いられてきた。

一般に, 近似精度を改善するためには, 目的変数に関係する可能性のあるできるだけ多くの種類の観測データを説明変数として用いることが望ましい。しかし, 説明変数の種類 (次元) の増加に従い, 目的変数と説明変数の非線形な関係を表す基底関数の数が指数関数的に増加する問題 (次元の呪い) が発生する<sup>(9)</sup>。そのため, 説明変数と目的変数の間の全ての非線形性を評価できない。

さらに, 多くの場合, 説明変数の種類 (次元) に対して十分な数のデータを用意できないため, データを生成する関数やシステムの複雑な構造を十分表現できない。そのため, あるデータ集合 (学習データ) に対して近似誤差が最小となるように多項式の基底関数と係数を決めたととしても, それがデータを生起するシステムを正しく表しているとは限らない。その結果, 学習データとは異なるデータ集合 (評価データ) に対する近似誤差が大きくなってしまいうという過学習がしばしば発生する。

多くの物理システムでは, 未知の関数  $f(x)$  を通常の Fourier 級数により展開した場合, Fourier 係数の多くが 0 またはほぼ 0 になる。これをスパース性と呼ぶ。また, 一般に, スパース性は説明変数  $x$  の次元が高くなるほど顕著になる。そこで, 次元の呪いと過学習という二つの問題を同時に解決するために, データのスパース性を利用した様々な解法が提案されている。

目的変数と説明変数の関係が線形の場合には, 高次元の説明変数に対してスパースな解を得る方法が提案された<sup>(9)</sup>。しかし, 一般に, 目的変数と説明変数の関係は線形とは限らない。目的変数と説明変数の関係が非線形となるような問題を解決するために, GA<sup>(6)(7)</sup> を用いて基底関数の数と次数を選択し, ベイズ情報量基準 (BIC)<sup>(10)</sup> を用いてその優劣を決めることにより, 過学習を防ぎ, 汎化能力を最大化する方法が提案された<sup>(11)(12)</sup>。BIC や赤池情報量基準 (AIC)<sup>(13)</sup> は, 汎化能力を評価するための評価関数である。しかし, データの分布やサンプル数が情報量基準に合わない場合, 情報量基準が正しく機能しないことがある。そのため, 汎化能力を最大化するためには, 扱うデータの分布やサンプル数に応じて, 用いる情報量基準を適切に変更しなければならない<sup>(14)</sup>。そのため, 最適な情報量基準の選択は試行錯誤にならざるを得ない。

これらの問題を回避するために, 本論文では, 味覚センサの複数の出力 (高次元の説明変数) から官能評価の結果 (目的変数) を推定する関数を GA を用いて構築し, 近似誤差を交差検証<sup>(15)</sup> を用いて評価する。また, 食品の種類を識別する際に官能評価の推定値を用いることにより, 認識率を改善できることを示す。

本論文では, まず, 2 章で, GA を用いた関数近似手法とパターン認識への応用手法を提案する。次に, 3 章で, 提案手法を用いて味データから官能評価データを再現し, 食品の種類を識別した結果を示す。最後に, 4 章で, 本論文の提案とその評価結果をまとめる。

## 2. 遺伝的アルゴリズムを用いた関数近似とパターン認識への応用

**(2・1) 関数近似** ベクトル  $x$  を未知の関数  $f(x)$  の入力, ベクトル  $y$  をその出力,  $x$  を説明変数,  $y$  を目的変数として, 観測されたデータの集合  $\{(x_{(1)}, y_{(1)}), (x_{(2)}, y_{(2)}), \dots\}$  から  $f(x)$  の近似関数  $\tilde{f}(x)$  を同定する問題を関数近似問題という。 $f(x)$  の近似関数  $\tilde{f}(x)$  は, 次式で表される。

$$\begin{aligned} \tilde{y} &= \tilde{f}(x) \\ &= \sum_{i=1}^N \alpha_i \phi_i(x) + b \dots\dots\dots (1) \end{aligned}$$

ここで,  $N$  は展開次数,  $\{\phi_i(x)\}$  は基底関数,  $\{\alpha_i\}$  および  $b$  は学習データから求められる係数ベクトルである。 $\{\phi_i(x)\}$  がガウシアンカーネル<sup>(5)</sup> の場合,  $\{\alpha_i\}$  を学習データから決める代表的な手法として, SVM が挙げられる<sup>(6)</sup>。SVM はパターン認識や予測の分野において最も優れた手法の一つであり, 線形問題のみならず, 非線形問題に対する能力も高い。また, 説明変数  $x$  が高次元の場合でも高い性能が得られるという優れた特徴がある。

しかし, 基底関数  $\{\phi_i(x)\}$  をガウシアンカーネルとした場合, 目的変数と説明変数の陽な関係が得られるとは限らない。そのため, 説明変数  $x$  から目的変数の近似値  $\tilde{y}$  を計算するだけでなく, 例えば, 目的変数が  $d$  番目の説明変数の 2 乗に反比例し,  $d'$  番目の説明変数に比例するといった近似関数  $\tilde{f}(x)$  の構造を知りたい場合には, SVM は適さない。

基底関数  $\{\phi_i(x)\}$  が正規直交基底の場合, 式 (1) は Fourier 級数展開であり,  $\{\alpha_i\}$  は Fourier 係数ベクトルとなる。この場合,  $\{\alpha_i\}$  は重回帰分析<sup>(6)</sup> により得られる。また,  $\{\phi_i(x)\}$  が Legendre 関数 (付録 1 参照) の場合, 目的変数を説明変数の冪乗で近似できるため, 目的変数と説明変数の高次の関係が明確になる。

しかし, 通常の Fourier 級数の正規直交基底を用いた場合, 説明変数  $x$  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, 十分な近似精度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関数近似の大きな問題点の一つである<sup>(8)</sup>。

本論文では, 味覚センサの出力から官能評価の結果を推定する関数を Legendre 関数を基底関数とする多項式で近似することにより, 味覚センサの出力と官能評価結果の陽な関係を明らかにする。また, 次元の呪いを回避し, 汎化能力を最大化するために, GA を用いて基底関数の数と次数を選択し, 近似精度を交差検証により評価する関数近似手法を提案する。さらに, パターン認識の認識率を改善するために, 提案した関数近似手法により得られた近似関数を, パターン認識に適用する方法を提案する。以下に, その手法を示す。

**(2・2) 遺伝的アルゴリズム** GA 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<sup>(6)(7)</sup>。i番目の個体は染色体に対応する  $m_{\max}$  次元のベクトル  $\mathbf{g}_i$  を持ち、個体の優劣は  $\mathbf{g}_i$  から計算される適応度  $J_i$  で表される。本論文では、染色体  $\mathbf{g}_i$  の  $m$  番目の遺伝子  $g_{i,m}$  を  $g_{i,m} \in \{0, 1, \dots, g_{\max}\}$  で定義した。GA は、優秀な個体を得るために、選択、交叉、突然変異を繰り返すことにより、染色体  $\mathbf{g}_i$  を改良するアルゴリズムである。GA は解空間を大域的に探索できるため、局所解に陥りにくいという特徴がある。

GA の最も基本的な戦略であるエリート戦略では、個体の適応度が高い順に  $\mathbf{g}_i$  をソートする。その後、 $\mathbf{g}_1, \dots, \mathbf{g}_{N_{\text{renew}}}$  を残し、 $\mathbf{g}_{N_{\text{renew}}+1}, \dots, \mathbf{g}_{b_{\max}}$  を廃棄する。ここで、 $b_{\max}$  は個体数である。次に、 $\mathbf{g}_1, \dots, \mathbf{g}_{N_{\text{renew}}}$  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  $\mathbf{g}_{N_{\text{renew}}+1}, \dots, \mathbf{g}_{b_{\max}}$  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、 $0, 1, 2, \dots, g_{\max}$  の一様乱数で決められる。

**〈2・3〉 非線形重回帰分析**  $d_x$  次元の説明変数ベクトル  $\mathbf{x} = (x_1, \dots, x_{d_x})^T$  と  $d_y$  次元の目的変数ベクトル  $\mathbf{y} = (y_1, \dots, y_{d_y})^T$  の関係を表す関数  $f$  の近似関数  $\tilde{f}$  を推定する問題を考える。 $f$  が線形の場合には、 $f$  が重回帰モデルで表されると仮定し、データ集合  $\mathbf{D} \stackrel{\text{def}}{=} \{(\mathbf{x}_{(n)}, \mathbf{y}_{(n)}) | n = 1, \dots, N_{\text{data}}\}$  から、重回帰モデルのパラメータを推定する重回帰分析<sup>(16)</sup> が多くの分野で用いられている。また、線形の重回帰分析を拡張した非線形重回帰分析により、 $f$  が非線形の場合でも、その近似関数  $\tilde{f}$  を推定できる。

$\mathbf{k}_i$  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)、正規直交基底関数を  $\{K(\mathbf{x}, \mathbf{k}_1), \dots, K(\mathbf{x}, \mathbf{k}_N)\}$ 、 $\mathbf{Z} \stackrel{\text{def}}{=} \{\mathbf{k}_1, \dots, \mathbf{k}_N\}$  をインデックスベクトルの集合、 $\mathbf{K}(\mathbf{x}, \mathbf{Z}) \stackrel{\text{def}}{=} (K(\mathbf{x}, \mathbf{k}_1), \dots, K(\mathbf{x}, \mathbf{k}_N))^T$  とすると、目的変数  $\mathbf{y}$  の推定値  $\tilde{\mathbf{y}}$  を計算するための近似関数  $\tilde{f}$  は、次式で表される。

$$\tilde{\mathbf{y}} = \tilde{f}(\mathbf{x}) \stackrel{\text{def}}{=} \mathbf{A}\mathbf{K}(\mathbf{x}, \mathbf{Z}) + \mathbf{b} \dots \dots \dots (2)$$

ここで、 $d_y \times d_x$  次元の行列  $\mathbf{A}$  と  $d_y$  次元のベクトル  $\mathbf{b}$  は、上記非線形重回帰モデルの回帰係数である。

学習データ集合を  $\mathbf{D}_L \stackrel{\text{def}}{=} \{(\mathbf{x}_{L(n)}, \mathbf{y}_{L(n)}) | n = 1, \dots, N_{L\text{data}}\}$ 、 $E[\mathbf{y}_{L(n)}]$  を  $\{\mathbf{y}_{L(n)}\}$  の平均、 $E[\mathbf{K}(\mathbf{x}_{L(n)}, \mathbf{Z})]$  を  $\{\mathbf{K}(\mathbf{x}_{L(n)}, \mathbf{Z})\}$  の平均、 $\text{COV}_{\mathbf{K}\mathbf{K}}$  を  $\{\mathbf{K}(\mathbf{x}_{L(n)}, \mathbf{Z})\}$  の共分散行列 (すなわち、 $\mathbf{K}(\mathbf{x}_{L(n)}, \mathbf{Z})$  の  $i$  番目の要素  $K(\mathbf{x}_{L(n)}, \mathbf{k}_i)$  と  $j$  番目の要素  $K(\mathbf{x}_{L(n)}, \mathbf{k}_j)$  の共分散を要素とする行列)、 $\text{COV}_{\mathbf{y}\mathbf{K}}$  を  $\{\mathbf{y}_{L(n)}\}$  と  $\{\mathbf{K}(\mathbf{x}_{L(n)}, \mathbf{Z})\}$  の共分散行列 (すなわち、 $\mathbf{y}_{L(n)}$  の  $i$  番目の要素  $y_{L(n),i}$  と  $\mathbf{K}(\mathbf{x}_{L(n)}, \mathbf{Z})$  の  $j$  番目の要素  $K(\mathbf{x}_{L(n)}, \mathbf{k}_j)$  の共分散を要素とする行列) とすると、式 (2) で定義された近似関数  $\tilde{f}$  (非線形重回帰モデル) の回帰係数は、次式により得られる。

$$\begin{cases} \mathbf{A} = \text{COV}_{\mathbf{y}\mathbf{K}}\text{COV}_{\mathbf{K}\mathbf{K}}^{-1} \\ \mathbf{b} = E[\mathbf{y}_{L(n)}] - \mathbf{A}E[\mathbf{K}(\mathbf{x}_{L(n)}, \mathbf{Z})] \end{cases} \dots \dots \dots (3)$$

本論文では、式 (2) の近似精度を、学習データ集合  $\mathbf{D}_L$  と異

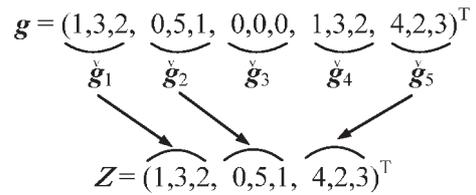


Fig. 1. Example of a map from  $\mathbf{g}$  to  $\mathbf{Z}$ .

なる評価データ集合  $\mathbf{D}_E \stackrel{\text{def}}{=} \{(\mathbf{x}_{E(n)}, \mathbf{y}_{E(n)}) | n = 1, \dots, N_{E\text{data}}\}$  に対する正規化誤差  $\bar{\varepsilon}$  により評価する。

$$\bar{\varepsilon} \stackrel{\text{def}}{=} E[(y_{E d(n)} - \tilde{y}_{E d(n)})^2 / (y_{E \max;d} - y_{E \min;d})^2] \dots \dots \dots (4)$$

ここで、 $y_{E d(n)}$  と  $\tilde{y}_{E d(n)}$  は、各々  $\mathbf{y}_{E(n)}$  と  $\tilde{\mathbf{y}}_{E(n)}$  の  $d$  番目の要素、 $y_{E \max;d}$  と  $y_{E \min;d}$  は、各々  $\mathbf{y}_{E d(n)}$  の最大値と最小値である。

また、得られた近似関数  $\tilde{f}$  (非線形重回帰モデル) から、説明変数と目的変数の陽な関係を明らかにするために、基底関数として Legendre 関数 (付録 1 参照) を用いる。

**〈2・4〉 交差検証による適応度の計算** 〈2・1〉節で述べた次元の呪いと過学習の問題を解決するために、本論文では、非線形重回帰モデル (式 (2)) の基底関数の数と次数を GA の染色体で表現し、近似精度を交差検証<sup>(15)</sup> により評価する。

$N_{\max} = m_{\max}/d_x$  を展開次数  $N$  の最大値、 $\mathbf{g}$  を  $N_{\max}$  個に分割して作られた染色体の断片を  $\check{\mathbf{g}}_i (1 \leq i \leq N_{\max})$  とする。染色体  $\mathbf{g}$  を基底関数のインデックスベクトル集合  $\mathbf{Z} \stackrel{\text{def}}{=} \{\mathbf{k}_1, \dots, \mathbf{k}_N\}$  に写像するアルゴリズムを以下に示す。

**[Algorithm 1]** 基底関数のインデックスベクトルの設定

- (1-1)  $\mathbf{g}$  から、 $\check{\mathbf{g}}_1, \check{\mathbf{g}}_2, \dots, \check{\mathbf{g}}_{N_{\max}}$  を作成し、 $\check{\mathbf{G}} \stackrel{\text{def}}{=} \{\check{\mathbf{g}}_1, \check{\mathbf{g}}_2, \dots, \check{\mathbf{g}}_{N_{\max}}\}$  とする。
- (1-2)  $\check{\mathbf{G}}$  から、 $\check{\mathbf{g}}_i = \mathbf{0}$  なる  $\check{\mathbf{g}}_i$  を削除する ( $1 \leq i \leq N_{\max}$ )。
- (1-3)  $\check{\mathbf{G}}$  から、 $\forall i > j$  に対して、 $\check{\mathbf{g}}_i = \check{\mathbf{g}}_j$  なる  $\check{\mathbf{g}}_i$  を削除する ( $1 \leq j \leq N_{\max}$ )。
- (1-4)  $N$  を  $\check{\mathbf{G}}$  の要素数とする。
- (1-5)  $\check{\mathbf{G}}$  の要素を、順に  $\mathbf{k}_1, \dots, \mathbf{k}_N$  に割り当てる。

$\mathbf{g} = (1, 3, 2, 0, 5, 1, 0, 0, 0, 1, 3, 2, 4, 2, 3)^T$  に対する Algorithm 1 の動作と  $\mathbf{Z}$  を Fig. 1 に示す。ここで、染色体の長さ  $m_{\max} = 15$ 、 $\mathbf{x}$  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)  $g_{\max} = 5$  であり、Algorithm 1 の結果として得られた展開次数  $N = 3$  である。Fig. 1 では、 $\check{\mathbf{g}}_3$  は  $\check{\mathbf{g}}_3 = \mathbf{0}$  であるため削除される。また、 $\check{\mathbf{g}}_4$  は  $\check{\mathbf{g}}_4 = \check{\mathbf{g}}_1$  であるため削除される。その結果、 $\check{\mathbf{g}}_1, \check{\mathbf{g}}_2, \check{\mathbf{g}}_5$  が  $\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3$  に割り当てられる。すなわち、 $\mathbf{k}_1 = \check{\mathbf{g}}_1, \mathbf{k}_2 = \check{\mathbf{g}}_2, \mathbf{k}_3 = \check{\mathbf{g}}_5, N = 3$  となる。Algorithm 1 で染色体  $\mathbf{g}$  から基底関数のインデックスベクトル集合  $\mathbf{Z}$  を計算する際に、基底関数の数  $N$  は、 $1 \leq N \leq N_{\max}$  の範囲で染色体  $\mathbf{g}$  に応じて決まる。そのため、適応度の計算に交差検証を導入し、評価データの誤差から染色体  $\mathbf{g}$  の適応

度を計算することにより, 過学習を押さえ汎化能力が最大となるように, 基底関数の次数のみならず基底関数の数  $N$  を決められる。

交差検証を行うためには, まず, 目的変数と説明変数のデータ集合  $\mathbf{D} \stackrel{\text{def}}{=} \{(\mathbf{x}_{(n)}, \mathbf{y}_{(n)}) | n \in \{1, \dots, N_{\text{data}}\}\}$  を  $\eta$  分割し, 各々を  $\mathbf{D}_1, \dots, \mathbf{D}_\eta$  とする。すなわち,  $\mathbf{D} = \{\mathbf{D}_1, \dots, \mathbf{D}_\eta\}$  である。次に,  $\mathbf{D}_{L;\ell} \stackrel{\text{def}}{=} \mathbf{D}_\ell$  を学習データ集合,  $\mathbf{D}_{E;\ell} \stackrel{\text{def}}{=} \mathbf{D} - \mathbf{D}_\ell$  を評価データ集合とする。以下に, 染色体  $\mathbf{g}$  から基底関数のインデックスベクトル集合  $\mathcal{Z}$  を計算し, 非線形重回帰分析により得られた近似関数の平均正規化誤差を交差検証により計算する手順を示す。ここで, 学習データ  $\mathbf{D}_{L;\ell}$  を用いて得られた近似関数を  $\tilde{\mathbf{f}}_\ell$ , 評価データ  $\mathbf{D}_{E;\ell}$  に対する  $\tilde{\mathbf{f}}_\ell$  の正規化誤差を  $\tilde{\epsilon}_\ell$ , 全評価データ ( $\ell = 1, \dots, \eta$ ) に対する平均正規化誤差を  $E[\tilde{\epsilon}_\ell]$  とすると, 染色体  $\mathbf{g}$  の適応度  $J$  は  $E[\tilde{\epsilon}_\ell] \times (-1)$  で定義される。

**[Algorithm2]** 交差検証を導入した適応度の計算

- (2-1) Algorithm 1 に  $\mathbf{g}$  と  $N_{\text{max}}$  を入力し,  $N$  および  $\mathcal{Z}$  を計算する。
- (2-2)  $\ell = 1, \dots, \eta$  に対して, 以下の計算を行う。
  - (i) 学習データ集合  $\mathbf{D}_{L;\ell}$  から, 式 (3) (重回帰分析) により回帰係数を計算する。
  - (ii) 式 (4) により, 評価データ集合  $\mathbf{D}_{E;\ell}$  に対する正規化誤差  $\tilde{\epsilon}_\ell$  を計算する。
- (2-3)  $J = -E[\tilde{\epsilon}_\ell]$

**〈2・5〉 関数近似とパターン認識への応用** 〈2・3〉節で述べた非線形重回帰分析および〈2・4〉節で述べた適応度の計算手法を GA に適用し, 〈2・4〉節で定義されたデータ集合  $\mathbf{D}$  に対して関数近似を行う。これにより, 次元の呪いを回避し, 汎化能力を最大にするために, 基底関数の数と次元を最適化する。以下に, そのアルゴリズムを示す。

**[Algorithm3]** GA を用いた関数近似—交差検証による汎化能力の最適化—

- (3-1) 染色体  $\mathbf{g}_1, \dots, \mathbf{g}_{b_{\text{max}}}$  を初期化する。
- (3-2) Algorithm 2 に,  $\mathbf{g}_1, \dots, \mathbf{g}_{b_{\text{max}}}$  および  $N_{\text{max}}$  を入力し, データ集合  $\mathbf{D}$  に対する適応度  $J_1, \dots, J_{b_{\text{max}}}$  を計算する。
- (3-3) 適応度の値により  $J_1, \dots, J_{b_{\text{max}}}$  および  $\mathbf{g}_1, \dots, \mathbf{g}_{b_{\text{max}}}$  をソートする。
- (3-4)  $J_1, \dots, J_{b_{\text{max}}}$  が収束していれば終了。そうでなければ, Step (3-5) へ。
- (3-5)  $\mathbf{g}_1, \dots, \mathbf{g}_{N_{\text{renew}}}$  を交叉し, 次世代の染色体  $\mathbf{g}_{N_{\text{renew}}+1}, \dots, \mathbf{g}_{b_{\text{max}}}$  を作成する。
- (3-6)  $\mathbf{g}_{N_{\text{renew}}+1}, \dots, \mathbf{g}_{b_{\text{max}}}$  の遺伝子を突然変異により変更する。
- (3-7) Step (3-2) へ戻る。

Algorithm 3 により, 目的変数の推定値を求める近似関数 (式 (2)) の基底関数のインデックスベクトル集合  $\mathcal{Z}$  が得られる。また, Step (3-2) では, Algorithm 2 の Step(2-2) により各学習データ集合  $\mathbf{D}_{L;\ell}$  毎に式 (2) の回帰係数が得られる。 $\mathbf{D}_{L;\ell}$  に対する近似関数  $\tilde{\mathbf{f}}_\ell$  は, 回帰係数  $A_\ell, \mathbf{b}_\ell$  を式 (2) に代入すると, 次式で表される。

$$\tilde{\mathbf{y}} = \tilde{\mathbf{f}}_\ell(\mathbf{x}) \stackrel{\text{def}}{=} A_\ell \mathbf{K}(\mathbf{x}, \mathcal{Z}) + \mathbf{b}_\ell \dots \dots \dots (5)$$

ここで, 交差検証により汎化能力を最適化しているため, インデックスベクトルの集合  $\mathcal{Z}$  は, 全ての  $\tilde{\mathbf{f}}_\ell$  に対して共通である。

次に, Algorithm 3 により得られた近似関数  $\tilde{\mathbf{f}}_\ell$  (式 (5)) を用いて, パターン認識における認識率を改善する手法を述べる。パターン認識は, 説明変数  $\mathbf{x}$  から説明変数に対応するクラスの推定値  $\tilde{\Omega}$  を計算する問題である。パターン認識器は, まず, 学習データとして与えられた説明変数とクラスの集合を用いて, 説明変数とクラスの間を学習する。その後, 未学習の評価データの説明変数に対応するクラスを推定する。

本論文では, 学習データに対してのみ, 認識率の改善に寄与するような補助情報を利用できる状況を仮定する。この状況では, 評価データに対しては補助情報を利用できない。そこで, 補助情報を Algorithm 3 における目的変数と考え, 説明変数と補助情報の関係を表す近似関数  $\tilde{\mathbf{f}}_\ell$  を Algorithm 3 を用いて同定し,  $\tilde{\mathbf{f}}_\ell$  により推定された補助情報を用いる。すなわち, Algorithm 3 の説明変数  $\mathbf{x}$  および Algorithm 3 により得られた  $\tilde{\mathbf{f}}_\ell$  により推定された補助情報の推定値  $\tilde{\mathbf{y}}$  から, 新たに説明変数  $\mathbf{x}'$  を作成し,  $\mathbf{x}'$  をパターン認識器の説明変数とすることにより, 認識率を改善する。

Algorithm 3 で使用されたデータ集合  $\mathbf{D}$  の説明変数  $\mathbf{x}$  を用いて, 〈2・4〉節で述べた  $\eta$ -分割交差検証により, パターン認識器の学習および評価を行うアルゴリズムを以下に示す。ここで,  $\mathbf{x}' \stackrel{\text{def}}{=} (\mathbf{x}, \tilde{\mathbf{y}})$  をパターン認識器の説明変数,  $\Omega$  を  $\mathbf{x}'$  に対応するクラス,  $\mathbf{D}' \stackrel{\text{def}}{=} \{(\mathbf{x}'_{(n)} \stackrel{\text{def}}{=} (\mathbf{x}_{(n)}, \tilde{\mathbf{y}}_{(n)}), \Omega_{(n)}) | \mathbf{x}_{(n)} \in \mathbf{D}, n \in 1, \dots, N_{\text{data}}\}$  をパターン認識器の説明変数とクラスのデータ集合とする。また, 〈2・4〉節の  $\eta$ -分割交差検証と同様に  $\mathbf{D}'$  を  $\eta$  分割し, 各々を  $\mathbf{D}'_1, \dots, \mathbf{D}'_\eta$  とする。すなわち,  $\mathbf{D}' = \{\mathbf{D}'_1, \dots, \mathbf{D}'_\eta\}$ ,  $\mathbf{D}'_{L;\ell} \stackrel{\text{def}}{=} \mathbf{D}'_\ell$  を学習データ集合,  $\mathbf{D}'_{E;\ell} \stackrel{\text{def}}{=} \mathbf{D}' - \mathbf{D}'_\ell$  を評価データ集合とする ( $\ell = 1, \dots, \eta$ )。

**[Algorithm4]** 学習データの補助情報を用いたパターン認識アルゴリズム

- (4-1) Algorithm 3 により, 近似関数  $\tilde{\mathbf{f}}_\ell$  (式 (5)) を同定する ( $\ell = 1, \dots, \eta$ )。
- (4-2)  $\ell = 1, \dots, \eta$  に対して, 以下の計算を行う。
  - (i) 式 (5) により,  $\{\mathbf{x}_{L(n)} \in \mathbf{D}_{L;\ell}\}$  から  $\{\tilde{\mathbf{y}}_{L(n)} \in \mathbf{D}'_{L;\ell}\}$  を計算する。
  - (ii)  $\mathbf{D}'_{L;\ell}$  からパターン認識器の学習およびパラメータ

の最適化を行う。

- (iii)  $\tilde{f}_\ell$  により,  $\{x_{E(n)} \in D_{E;\ell}\}$  から  $\{\tilde{y}_{E(n)} \in D'_{E;\ell}\}$  を計算する。
  - (iv) パターン認識器により,  $\{x'_{(n)} \in D'_{E;\ell}\}$  に対応するクラスの推定値  $\tilde{\Omega}$  および認識率  $r_{\Omega;\ell}$  を計算する。
- (4-3)  $r_{\Omega;\ell}$  の平均値  $E[r_{\Omega;\ell} | \ell \in \{1, \dots, m\}]$  を  $D'$  に対する認識率とする。

### 3. 性能評価

#### (3・1) 味覚センサの特性補正 (株)インテリジェント

センサーテクノロジーにより開発された味覚センサの精度および再現性は非常に高い<sup>(2)(3)</sup>。しかし、その一方で、大きく異なる味の相対評価を行う際には、味覚センサの分析結果は、パネルと呼ばれるプロフェッショナルの官能評価と異なる場合がある。味覚センサの特性を補正し、この問題を解決するために、味覚センサにより測定された味データとパネルによる官能評価の関係を導く。

味覚センサと官能評価の違いを明らかにするために、食品のサンプルとして、セミハードタイプ (18 サンプル)、フレッシュタイプ (10 サンプル)、およびソフトタイプ (6 サンプル) のチーズを用いた。これらのチーズの味データのうち、酸味、苦味雑味、旨味、苦味、及び旨味コクを味覚センサにより測定した。また、チーズの味に関わる重要な成分である乳酸と L-グルタミン酸の分量を、各々高速液体クロマトグラフィー (HPLC) とバイオセンサにより分析した (付録(2・1)参照)。さらに、これらのチーズの旨味、酸味、苦味雑味、旨味コク、及び苦味を官能評価により評価した (付録(2・2)参照)。

本節では、味覚センサと化学分析 (HPLC, バイオセンサ) により測定された味データ (以下では、客観的味データと呼ぶ) を説明変数、官能評価により得られた味データを目的変数として、(2・5)節で提案した関数近似手法 (Algorithm 3) を用いて、説明変数と目的変数の関係を導出する。尚、本論文で用いられる全ての味データ (説明変数及び目的変数) は、標準偏差で正規化された味データを表す<sup>†</sup>。味覚センサと化学分析により測定された客観的味データの分布の一例として、旨味コクと乳酸の分布を Fig. 2 に示す。また、官能評価により得られた旨味系の味 (以下では、補正旨味指標 M1 と呼ぶ) と化学分析により得られた乳酸の分布を Fig. 3 に示す。補正旨味指標 M1 は、官能評価により得られた旨味系の味の順位の関係から導出された旨味系の味の強度に対応する。

Fig. 2 より、味覚センサと化学分析では、3つのタイプ (セミハード、フレッシュ、ソフト) のチーズの旨味コクと乳酸の値が混じり合っていることが分かる。他の味データに対しても同様であり、味覚センサや化学分析により測定された客観的味データでは、チーズのタイプを完全に識別

<sup>†</sup> 味データを正規化することにより、回帰係数の絶対値が、基底関数の目的変数に対する影響度に比例する。

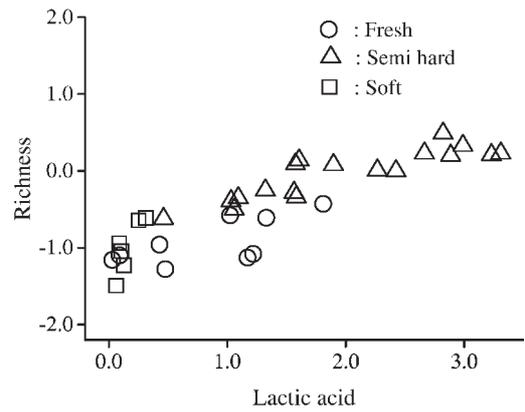


Fig. 2. Distribution of richness and lactic acid, where the former was obtained using taste sensor and the latter was obtained using HPLC.

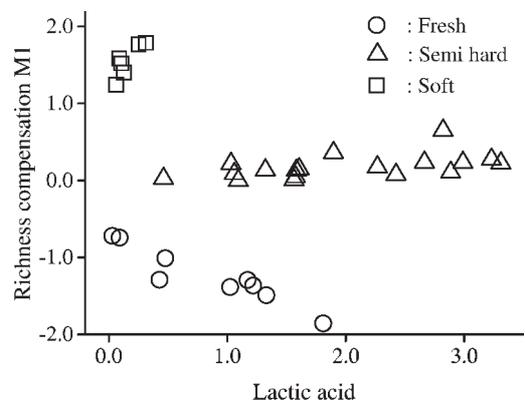


Fig. 3. Distribution of richness compensation M1 and lactic acid, where the former was obtained using sensory evaluation and the latter was obtained using HPLC.

できない。Fig. 2 の旨味コクは、チーズの懸濁液 (付録(2・1)参照) を遠心分離して得られた水溶性成分の味であり、チーズが本来持つ、不溶性ペプチドや脂質に由来する旨味が含まれていないためである。

一方、パネルが全ての成分を感じた上で総合的に味を評価している官能評価では、上記3タイプのチーズを明確に識別できる。特に、旨味系の評価に関しては明らかな違いがあり、Fig. 3 から分かるように、3つのタイプのチーズの補正旨味指標 M1 の値が分離している。このように、3つのタイプのチーズ (セミハードタイプ、フレッシュタイプ、ソフトタイプ) は、大きく味が異なるため、それらの味を同一の基準で相対的に評価すると、Fig. 3 の様に各々が異なるクラスタを構成する。

そのため、客観的味データから補正旨味指標 M1 を再現する法則が分かれば、客観的味データを基に、チーズの味を官能評価と同等に評価し、チーズを分類できる。その結果、手間のかかる官能評価の代わりに、味覚センサや化学分析により測定された客観的味データを、チーズの味の評価に利用できる。

そこで、2章で提案した Algorithm 3 を用いて、客観的味データから、補正旨味指標 M1 を再現する法則を導出す

る。まず、チーズの客観的味データ (酸味, 苦味雑味, 旨味, 苦味, 旨味コク, 乳酸, L-グルタミン酸) を説明変数  $x_1, \dots, x_7$ , 目的変数  $y_1$  を補正旨味指標 M1, 先に述べたチーズのサンプルの目的変数と説明変数のデータ集合を  $D$  とする。

Algorithm 3 の交差検証を実施するために, 全データ集合  $D$  を分割する。本章の実験では, データ数を考慮して, 全データ集合  $D$  を, 各々セミハードタイプ 9 種類, フレッシュタイプ 5 種類, ソフトタイプ 3 種類からなるデータ集合  $D_1$  と  $D_2$  に 2 分割し ( $\eta = 2$ ), Algorithm 3 により,  $g_{\max} = 1, 2, 4, 8$  に対して補正旨味指標 M1 の推定値を計算した。ここで, Algorithm 3 の主なパラメータは,  $N_{\max} = 16$ ,  $m_{\max} = 112$ ,  $b_{\max} = 2048$  である。 $N_{\max}$  は, 式 (3) において  $COV_{KK}$  の逆行列が安定に求められる最大値であり, データ集合  $D_1$  と  $D_2$  の要素数から決まる。 $m_{\max}$  は, (2.4) 節に記載の関係式 ( $N_{\max} = m_{\max}/d_x$ ),  $N_{\max} = 16$ , および  $d_x = 7$  から得られる。また,  $b_{\max}$  は, GA の性質上, 十分大きな値とすることにより最適解に近い解が得られ可能性が高くなるため, 実行時間などを考慮して 2048 とした<sup>(17)</sup>。

Algorithm 3 の動作を確認するために, Algorithm 3 の  $g_{\max}$  に対する平均正規化誤差および線形重回帰分析 (Linear MRA) による平均正規化誤差を Table 1 に示す。ここで, 平均正規化誤差は, (2.4) 節で定義されたように, 評価データに対する正規化誤差の平均値である。本実験では, Algorithm 3 の基底関数として Legendre 関数 (付録 1 参照) を用いている。また,  $g_{\max}$  は遺伝子の最大値であり, インデックスベクトルの要素の最大値を表す。そのため,  $g_{\max}$  は式 (5) の近似関数における説明変数  $x_d$  の最大次数となる。従って,  $g_{\max}$  の増加に従い基底関数の選択肢が増えるため, 近似関数の精度が上がり, 平均正規化誤差の減少が期待される。しかし, その一方で, 得られる近似関数が複雑になるために, 過学習による平均正規化誤差の増加が懸念される。しかし, Table 1 より,  $g_{\max}$  の増加に従い, 平均正規化誤差が小さくなっていることが分かる。また, Algorithm 3 の平均正規化誤差は, 線形重回帰分析と比べて 1 桁以上小さい。さらに, Table 1 より, 全ての  $g_{\max}$  に対して, 展開次数  $N$  はその最大値 ( $N_{\max} = 16$ ) より小さいことが分かる。Algorithm 3 は, 様々な値の  $N$  に対して正規化誤差を交差検証により評価しているため, Algorithm 3 により選ばれた  $N$  の値が Table 1 に記載の通り最大値より小さいことは, Algorithm 3 により選ばれた  $N$  の値が汎化能力を最大化するために最適であることを意味する。

比較のため, Algorithm 3 で交差検証を用いない場合<sup>†</sup> について, 評価データに対する平均正規化誤差と展開次数  $N$  を Table 2 に示す。一般に, 学習データに対する平均正規化誤差は, 展開次数が大きいくほど小さくなる。また, Algorithm 3 では, 平均正規化誤差が小さくなるように展開

Table 1. Mean normalized error between richness obtained from sensory evaluation and richness compensation M1 obtained from Algorithm 3.

	Linear MRA	Algorithm 3			
	---	1	2	4	8
$g_{\max}$	---	1	2	4	8
Error	$4.3 \times 10^{-2}$	$4.5 \times 10^{-3}$	$9.5 \times 10^{-4}$	$2.5 \times 10^{-4}$	$8.5 \times 10^{-5}$
$N$	7	11	14	15	15

Table 2. Mean normalized error between richness obtained from sensory evaluation and richness compensation M1 obtained from Algorithm 3 that does not use cross-validation<sup>†</sup>.

	Algorithm 3 (with $\bar{\varepsilon}_\ell$ for $D_{1:\ell}$ ) <sup>†</sup>			
	1	2	4	8
$g_{\max}$	1	2	4	8
Error	$2.7 \times 10^{-1}$	$7.5 \times 10^{-1}$	$5.5 \times 10^{-1}$	$2.4 \times 10^{-1}$
$N$	16	16	16	16

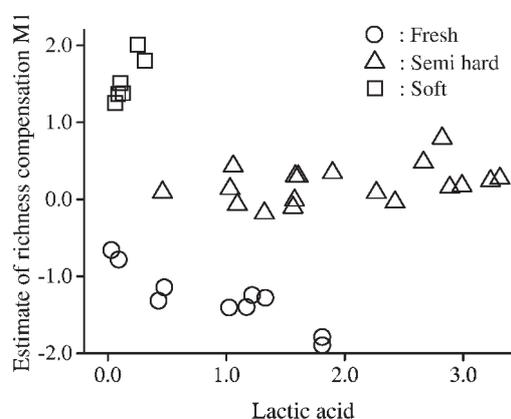


Fig. 4. Distribution of estimate of richness compensation M1 and lactic acid, where the former was obtained using Algorithm 3 and the latter was obtained using HPLC.

次数が選択される。そのため, Algorithm 3 で交差検証を用いない場合, すなわち, 学習データに対する平均正規化誤差を GA の適応度に用いる場合には, 展開次数はその最大値 ( $N_{\max} = 16$ ) となるはずである。これらの理由から, Algorithm 3 で交差検証を用いない場合には, Table 2 に記載の通り, 展開次数が最大値 16 となる。また, Algorithm 3 で交差検証を用いない場合には, 展開次数と回帰係数が, 学習データに対する平均正規化誤差のみを小さくするように決定される。この場合, 一般に過学習, すなわち, 評価データに対する誤差が大きくなる可能性が高い。そのため, Table 2 に記載の通り, 評価データに対する平均正規化誤差が Table 1 のものより大きい。

これらの結果により, 基底関数の数と次数を表す染色体の適応度を交差検証により計算する効果および Algorithm 3 の汎化能力を確認できた。

Algorithm 3 により得られた  $g_{\max} = 2$  の場合の補正旨味指標 M1 の推定値と乳酸の分布を, Fig. 4 に示す。ここで, Fig. 4 には,  $D_1$  を学習データ集合,  $D_2$  を評価データ集合とした時の  $D_2$  の補正旨味指標 M1 の推定値と乳酸値, お

<sup>†</sup> Algorithm 2, Step (2-2)(ii) において, 学習データ集合  $D_{1:\ell}$  に対する正規化誤差  $\bar{\varepsilon}_\ell$  を計算する。

Table 3. Index vectors of the bases and coefficients obtained by Algorithm 3.

$i$	$k_i$	$a_i$
1	$(0200020)^T$	$-5.71 \times 10^0$
2	$(0010000)^T$	$1.50 \times 10^2$
3	$(0000010)^T$	$-2.72 \times 10^0$
4	$(0200000)^T$	$7.58 \times 10^1$
5	$(0000100)^T$	$1.72 \times 10^2$
6	$(0001020)^T$	$-3.90 \times 10^0$
7	$(0001000)^T$	$1.78 \times 10^2$
8	$(0000001)^T$	$3.35 \times 10^1$
9	$(2120200)^T$	$-2.58 \times 10^2$
10	$(0021201)^T$	$1.91 \times 10^2$
11	$(0201221)^T$	$-6.45 \times 10^2$
12	$(0000220)^T$	$6.56 \times 10^1$
13	$(2000210)^T$	$-1.74 \times 10^2$
14	$(2112202)^T$	$2.84 \times 10^0$
		$b$
		$3.90 \times 10^{-1}$

よび  $D_2$  を学習データ集合,  $D_1$  を評価データ集合とした時の  $D_1$  の補正旨味指標 M1 の推定値と乳酸値が記載されている。Algorithm 3 の平均正規化誤差 (補正旨味指標 M1 (Fig. 3) に対する補正旨味指標 M1 の推定値 (Fig. 4) の誤差) は,  $9.5 \times 10^{-4}$  と十分小さい (Table 1,  $g_{\max} = 2$  を参照)。また, Fig. 4 の分布は Fig. 3 とほぼ一致しており, 3つのタイプのチーズの補正旨味指標 M1 の推定値が分離している。このように, Fig. 4 の補正旨味指標 M1 の推定値は, Fig. 3 の補正旨味指標 M1 を精度良く再現しており, 補正旨味指標 M1 の推定値により, チーズのタイプを識別できることが分かる。

また, Fig. 4 における補正旨味指標 M1 の推定値 (目的変数の推定値  $\tilde{y}_1$ ) と客観的味データ (説明変数  $x$ ) の関係は, 次式により表される。

$$\tilde{y}_1 = \sum_{i=1}^{14} a_i K(x, k_i) + b \dots \dots \dots (6)$$

上式において,  $K(x, k_i)$  は式 (付 3) の Legendere 関数,  $a_i$  および  $b$  は回帰係数,  $k_i$  は Legendere 関数のインデックスベクトルである。Algorithm 3 によって得られた  $a_i, b$ , および  $k_i$  の値を Table 3<sup>†</sup> に示す。Legendere 関数は, 説明変数  $x_d$  の冪乗で表される。そのため, 式 (6) および Table 3 より, 説明変数  $x_d$  と目的変数の推定値  $\tilde{y}_1$ , すなわち, 補正旨味指標 M1 の推定値と客観的味データの陽な関係 ( $x_d$  の積和) が明らかになった。

〈3・2〉 味による食品の認識率の改善

食品を味により識別する場合, 味覚センサや化学分析により測定された客観的味データが有効である。しかし, 〈3・1〉節の Fig. 2 に示

したように, 一部の食品では, 客観的味データによる識別が困難な場合がある。本節では, このような場合でも, 学習データに対してのみ官能評価を行うことができれば, その結果を用いて Algorithm 4 により, 食品の認識率を改善できることを示す。

本節では, パターン認識器として SVM<sup>(6)</sup> を用いる。SVM は最も優れたパターン認識器の一つであり, 線形のパターン認識問題のみでなく, 非線形問題に対するパターン認識能力も高い。また, 〈3・1〉節で定義されたデータ集合  $D_1, D_2$  を使い,  $D_1, D_2$  の説明変数ベクトル  $x$  に対応するクラス  $\Omega$  を  $D_1, D_2$  のチーズのタイプ (セミハードタイプ, フレッシュタイプ, ソフトタイプ) として, チーズのタイプの識別問題を考える。

まず, 以下の手順により, 客観的味データ (説明変数ベクトル  $x$ ) のみを用いた場合の SVM の認識率を調べた。(1)  $D_1$  を学習データ集合として, SVM の学習およびパラメータ (Lagrange 乗数の上限値  $C$ , ガウシアンカーネルにおけるガウス関数の広がりを表す  $\sigma$ )<sup>(6)</sup> の最適化を行う。(2)  $D_2$  を評価データ集合として認識率を求める。(3)  $D_1$  と  $D_2$  を入れ替えて同様に認識率を求め, 両者を平均する。その結果, 認識率は 91% であった。

次に, 学習データに対してのみ実施した官能評価により得られた補正旨味指標 M1 を補助情報  $y_1$ , パターン認識器を SVM, 説明変数ベクトル  $x$  の旨味コク ( $x_5$ ) を補助情報  $y_1$  の推定値  $\tilde{y}_1$  に置き換えたベクトル  $x' = (x_1, \dots, x_4, \tilde{y}_1, x_6, x_7)^T$  を SVM の説明変数,  $x'$  および  $\Omega$  からなるデータ集合を  $D'_1, D'_2$  として, Algorithm 4 の認識率を調べた。ここで, Algorithm 4 における SVM の学習およびパラメータの最適化手順は, 前述の SVM と同様である。その結果, 認識率は 100% となった。

以上の結果より, 学習データに対してのみ官能評価を行うことができれば, その結果を利用して Algorithm 4 を実行することにより, 客観的味データのみを用いる場合と比べて, チーズの3つのタイプ (セミハード, フレッシュ, ソフト) の認識率を改善できることが分かった。

一般に, 客観的味データと官能評価の関係は, (6) 式で表されるように, 非線形になる。そのため, 全ての食品に対する客観的味データと官能評価の関係を一つの式で表すことは, 必要となるデータの量および関係式の複雑さを考えると, 現実的でない。しかし, 対象となる食品を特定し, 学習データに対してのみ実施された官能評価の結果を利用して Algorithm 4 を実行することにより, 客観的味データのみを用いる場合と比べて, 対象とした食品 (例えば, 本論文におけるチーズの3つのタイプ) の認識率を改善できる。

これは, 評価データに対して官能評価を行うことなく, 食品の識別精度を改善できることを意味している。そのため, 官能評価に伴う作業の複雑さを軽減し, パネルの心理・健康状態に左右されない安定した評価が可能となる。

<sup>†</sup> 7 次元の変数  $x$  に対する基底関数  $K(x, k_i)$  の振幅は, 式 (付 4) における  $P(\cdot)$  の係数  $(\sqrt{\frac{2k_d+1}{D_d}} |k_d = 0, 1, 2, D_d \approx 4)$  の 7 乗オーダーであり, 非常に小さい。振幅の小さな基底関数を用いて, 変動範囲が  $-2 \sim 2$  の目的変数を近似するために, 相対的に回帰係数  $a_i$  が大きくなる。そのため, 一部の回帰係数  $a_i$  の値が, 目的変数の変動範囲 ( $-2 \sim 2$ ) と比較して非常に大きく ( $10^2$  オーダー) になっている。

4. まとめ

(株)インテリジェントセンサーテクノロジーにより開発された味覚センサは、非常に高い精度および再現性を持つ一方、大きく異なる味の相対評価を行う際には、パネルによる官能評価の結果と異なる場合がある。この問題を解決するために、味覚センサや化学分析により測定された客観的味データから官能評価の結果を推定する関数を多項式で表し、GAと交差検証を用いることにより、汎化能力が最大となるように基底関数の数と次数を最適化した。また、Legendre関数を多項式の基底関数とすることにより、客観的味データと官能評価の結果の陽な関係を明らかにできた。この関係を利用し、官能評価の代わりに客観的味データを用いることにより、パネルの心理や健康状態に左右されない安定した評価が可能となる。さらに、再現された官能評価のデータを用いてチーズのタイプを識別した場合、客観的味データのみを用いる場合と比べて、認識率を改善できた。

文 献

(1) K. Toko: "Biochemical Sensors: Mimicking Gustatory and Olfactory Senses: Order", Pan Stanford Publishing (2013)  
 (2) 都甲潔 編:「味覚センサ」, 朝倉書店 (1993)  
 (3) H. Ikezaki etc: "Techniques to control sensitivity and selectivity of multi-channel taste sensor using lipid membranes", Transducers '99 (1999)  
 (4) I. Matsuba: "Nonlinear Time Series Prediction using Neural Networks", Trans. REAJ, Vol.28, No.7, pp.442-450 (2006) (in Japanese)  
 松葉育雄:「ニューラルネットワークによる非線形時系列予測」, 日本信頼性学会誌, Vol.28, No.7, pp.442-450 (2006)  
 (5) N. Cristianini and J. S. Taylor: "An Introduction to Support Vector Machines and Other Kernel-based Learning Methods", Cambridge Univ. Press, UK (2000)  
 (6) D. E. Goldberg: "Genetic Algorithms in Search Optimization and Machine Learning", Addison-Wesley, USA (1989)  
 (7) J. H. Holland: "Genetic algorithm", Scientific American, pp.66-72 (1992-7)  
 (8) Sutton, R.G. and Barto, A.G.: "Reinforcement Learning", MIT Press, USA (1998)  
 (9) T. Zhang: "Adaptive Forward-Backward Greedy Algorithm for Learning Sparse Representations", IEEE Trans. Info. Th, Vol.57, pp.4689-4708 (2011)  
 (10) G. Schwarz: "Estimating the Dimension of a Model", The Annals of Statistics, Vol.6, No.2, pp.461-464 (1978)  
 (11) Z.-J. Yang, T. Fujimoto, and K. Kumamaru: "A genetic algorithm approach to identification of nonlinear polynomial models", 12th IFAC Sympo. on System Identification, pp.793-798 (2000-6)  
 (12) K. Nakajima, J. Nakajima, T. V. Khoa, and S. Hashimoto: "Identification method of nonlinear systems with friction based on genetic algorithm", 3rd International Symposium on Knowledge Acquisition and Modeling (KAM), pp.55-58 (2010-10)  
 (13) H. Akaike: "Information Theory and an Extension of the Maximum Likelihood Principle", Proc. 2nd Inter. Symposium on Information Theory, pp.267-281 (1973)  
 (14) N. Nakamura and S. Konishi: "Estimation of Number of Components for Multivariate Normal Mixture Models Based on Information Criteria", Trans. JSAS, Vol.27, No.3, pp.165-180 (1998) (in Japanese)  
 中村永友・小西貞則:「情報量規準に基づく多変量正規混合分布モデルのコンポーネント数の推定」, 応用統計学, Vol.27, No.3, pp.165-180 (1998)  
 (15) R. Kohavi: "A study of cross-validation and bootstrap for accuracy estimation and model selection", Proceedings of the 14th international joint conference on Artificial intelligence, IJCAI'95, Vol.2, pp.1137-1143 (1995)  
 (16) R. A. Johnson and D. W. Wichern: "Applied Multivariate Statistical Analysis 5th ed.", Pearson Education (Prentice Hall), USA (2001)  
 (17) M. Satoh, H. Satoh, and H. Ikezaki: "Optimization of Food Ingredients and their Blend Ratios Based on Taste Sensor Output", IEEJ Trans. SM, Vol.135,

No.2, pp.57-64 (2015-2) (in Japanese)

佐藤雅子・佐藤仁樹・池崎秀和:「味覚センサの出力データに基づく原料・ブレンド比の最適化」, 電学論 E, Vol.135, No.2, pp.57-64 (2015-2)

(18) 佐藤 信:「官能検査入門 第7刷」, (株)日科技連出版社 (東京都), pp.63-69 (1993)  
 (19) I.N. Bronshtein and K.A. Semendyayev: "Handbook of Mathematics", Springer-Verlag, UK (1997)

付 録

1. 正規直交基底

$\mathbf{x} \stackrel{\text{def}}{=} (x_1, \dots, x_{d_x})^T$ ,  $\mathcal{D}_x \stackrel{\text{def}}{=} \{\mathbf{x} | x_{\min d} \leq x_d \leq x_{\max d}, 1 \leq d \leq d_x\}$ とする。本節では、 $\mathbf{x} \in \mathcal{D}_x$ の正規直交基底について述べる。 $\mathbf{k} \stackrel{\text{def}}{=} (k_1, \dots, k_{d_x})^T \in \mathcal{Z}$ をインデックスベクトル、 $\mathcal{Z}$ を $\mathbf{k}$ の集合、 $h(\mathbf{k})$ を $\mathbf{k}$ に対するFourier係数とする。関数 $f(\mathbf{x})$ のFourier級数展開は、次式により定義される<sup>(9)</sup>。

$$f(\mathbf{x}) = \sum_{\mathbf{k} \in \mathcal{Z}} h(\mathbf{k}) K(\mathbf{x}, \mathbf{k}) \dots\dots\dots (付 1)$$

$$h(\mathbf{k}) \stackrel{\text{def}}{=} \int_{\mathcal{D}_x} f(\mathbf{x}) K^*(\mathbf{x}, \mathbf{k}) d\mathbf{x} \dots\dots\dots (付 2)$$

ここで、上付き添え字\*は複素共役、 $\{K(\mathbf{x}, \mathbf{k})\}$ 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、 $\{K(\mathbf{x}, \mathbf{k})\}$ は次式で定義される。

$$K(\mathbf{x}, \mathbf{k}) \stackrel{\text{def}}{=} \prod_{d=1}^{d_x} K_d(x_d, k_d) \dots\dots\dots (付 3)$$

$\{K_d(x_d, k_d)\}$ 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。

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

ここで、 $D_d \stackrel{\text{def}}{=} x_{\max d} - x_{\min d}$ 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。

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

基底関数 $\phi_i(\cdot)$ を次式で定義する。

$$\phi_i(\mathbf{x}) \stackrel{\text{def}}{=} K(\mathbf{x}, \mathbf{k}) \dots\dots\dots (付 6)$$

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、 $\mathcal{Z}_d \stackrel{\text{def}}{=} \{0, 1, \dots, N_d\}$ 、 $\mathcal{Z}$ は $\mathcal{Z}_d$ の直積により $\mathcal{Z} \stackrel{\text{def}}{=} \mathcal{Z}_1 \times \mathcal{Z}_2 \times \dots \times \mathcal{Z}_{d_x}$ で与えられるため、 $i$ は $\mathbf{k}$ を用いて次式で表される。

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

すなわち、通常のFourier級数は、Fourier係数を $\alpha_i$ 、基底関数 $\{\phi_i(\mathbf{x})\}$ 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,

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

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

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

**2. 味データの測定**

**(2・1) 味覚センサと化学分析による測定** まず, チーズの外皮を除き, チーズ全体を均等となるようにカットしてサンプルを作成した。次に, サンプル 6g に蒸留水 24 mL を加えてミルサーで均質化し, 5 倍懸濁液を作成した。これを遠心分離 (5°C, 18,000 rpm, 20 min) 後, ろ紙にてろ過し, 試験原液を抽出した。試験原液をさらに蒸留水で 10 倍に希釈し, 酸味, 苦味雑味, 旨味, 苦味, 及び旨味コクを味覚センサ (㈱インテリジェントセンサーテクノロジー TS-5000Z) により測定した。

また, 試験原液を DISMIC フィルター (0.25 μm) でろ過後, 乳酸を HPLC (東ソー㈱ 8020 シリーズ) により分析した (測定条件: カラム Shodex, RS-PAK KC-LG811 300 mm × 2 本, 移動相 25 mM HClO<sub>4</sub>, 流量 1.0 mL/min, 50°C, RI 検出器, 注入量 20 μL)。同様に, 検体に応じて試験原液を 5 ~ 250 倍に希釈して供試試料を作成し, L-グルタミン酸をバイオセンサ (王子計測機器㈱ BF-5iD) により分析した (測定条件: 検量線 10 ~ 500 mg/L, グルコース電極, 専用緩衝液 (37°C, PH7.0), 流量 1.0 mL/min)。

**(2・2) 官能評価による測定** 官能評価は, 普段からチーズを食べ慣れ, タイプの異なるチーズの標準的な味を認識しているパネル 4 名により実施された。セミハードタイプ (エメンタルチーズ, チェダーチーズ), フレッシュタイプ (モッツァレラチーズ, シェーブルチーズ), 及びソフトタイプ (カマンベールチーズ) のチーズに対して, 順位法<sup>(8)</sup>により, 味覚センサーや化学分析と比較可能な尺度である旨味, 酸味, 苦味雑味 (先に感じる苦味), 旨味コク

(後に残る旨味), 及び苦味 (後に残る苦味) の 5 項目を評価した。

本評価では, ヘルシンキ宣言に則り, パネルには事前に研究内容について十分に説明し, 試験に対する同意を確認した。また, 実験内容を, とち財団の関係部署で稟議の上, 承認を得た。

**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 packets 化, ATM 網のトラフィック制御, 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。

**葛西大介** (非会員) 1989 年弘前大学農学部園芸化学科卒業。同年よつ葉乳業㈱入社。チーズ, 乳製品の製造に従事。1994 年 4 月より (財) 十勝圏振興機構



(現: 公益財団法人とち財団) 入団。1999 年 4 月より研究開発課にて, 主に食品加工技術および食品評価技術の研究に従事。岩手連合大学院農学研究科 (生物資源科学専攻生物資源利用学講座) 博士課程在学中。

**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。原料・ブレンド比最適化ツール開発に従事。

