

清酒の成分に対する味覚センサデータの非線形重回帰モデル

非会員 佐藤 雅子* 非会員 高尾 佳史** 正員 佐藤 仁樹***

A Nonlinear Multiple Regression Model of Taste Sensor Data for Components in Sake

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

(2018年8月27日受付, 2018年12月10日再受付)

A nonlinear function that expresses the relationship between taste sensor data and components in sake was approximated using a polynomial of Legendre functions. First, the number of components in sake was reduced using principal component analysis. Second, the number of Legendre functions of the polynomial and their degrees were selected using a genetic algorithm. Third, the coefficients of the polynomial were calculated using multiple regression analysis. The approximation error was estimated using cross-validation, and the number of Legendre functions and their degrees were optimized so as to maximize the generalization of the polynomial. As a result, sufficiently small approximation errors were obtained, and the explicit relationship between taste sensor data and components in sake was clarified using the polynomial. Furthermore, it was possible not only to confirm the taste sensor response but also to improve manufacturing processes of sake using the estimates of the variations in the taste sensor data.

キーワード: 味覚センサ, 化学分析, 関数近似, 遺伝的アルゴリズム, 主成分分析

Keywords: taste sensor, chemical analysis, function approximation, genetic algorithm, principal component analysis

1. はじめに

食品の品質管理や機能性の評価のために, 食品業界では様々な化学分析が行われている。測定可能な成分の種類は膨大であるため, 成分は目的に沿って選択される。これらの分析は, 分析機器の発達により迅速かつ高精度に行われている。

一方, 食品の製造販売には, 味の評価も重要である。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, 体調, 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⁽¹⁾⁽²⁾。味覚センサにより, 対象となる食品の甘味, 塩味, 旨味, 酸味などの味の強さを数値(味データ)として表現できる。そのため, 味の目標値と開発中の食品の味の違いを客観的かつ数値的に表現できるようになった。

しかし, 味の違いは食品の成分を表現していないため, 食品の製造プロセスの改良には直結しない。味と成分の関係が明らかになれば, 味の違いに影響する成分を推定し, その成分の含有量を修正するように製造プロセスを改良できる可能性が広がる。そこで, 本論文では, 清酒に着目し, 清酒の成分から味データを推定するための陽な関数を求め, それらの関係を明らかにする。

観測された説明変数と目的変数のデータから, データを生成したシステムの構造やパラメータを同定する問題は, 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⁽³⁾, サポー

* 情報ノ宮路の下工房

〒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(株)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

*** 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

トベクターマシン (SVM) (4), 遺伝的アルゴリズム (GA) (6) 等, 様々な手法が用いられてきた。

これらの手法により, 目的変数を高い精度で推定できる。しかし, その一方で, 説明変数と目的変数の関係が明示的に得られるとは限らない。そこで, 本論文では, GA により得られた最適な基底関数を用いた非線形重回帰分析 (NMRA) (6) により, 目的変数を説明変数の関数で表す。この手法では, 基底関数として Legendre 関数 (付録 1 参照) を用いるため, 目的変数を説明変数の冪乗で近似できる。そのため, 説明変数と目的変数の高次の関係が明確になる。

本論文では, まず, 2 章で, GA を用いた NMRA による関数近似について述べる。次に, 3 章で, GA を用いた NMRA により, 清酒の成分から味データを推定した結果を示す。また, 味データの変化量の推定値を, 味覚センサの応答の確認や製造プロセスの改良に応用する方法について述べる。最後に, 4 章で, 本論文の結論をまとめる。

2. GA を用いた NMRA による関数近似

本章では, GA を用いた NMRA (6) により, 説明変数と目的変数の関係を明らかにし, 説明変数から目的変数を推定する方法を述べる。

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

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

ここで, N は展開次数, $\{\phi_i(\mathbf{x})\}$ は基底関数, $\{\alpha_i\}$ および \mathbf{b} は観測されたデータから求められる係数ベクトルである。

基底関数 $\{\phi_i(\mathbf{x})\}$ 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 (6) により得られる。また, $\{\phi_i(\mathbf{x})\}$ 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, 十分な近似精度を得るために必要となる展開次数 N は説明変数 \mathbf{x} の次元 d_x に対して指数関数的に増加する (付録 1, 式 (付 9) 参照)。この問題は, GA を用いて基底関数の数と次数を選択することにより回避される (6)。

〈2・2〉 NMRA 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, データ集合 $\mathbf{D} \stackrel{\text{def}}{=} \{(\mathbf{x}_{(n)}, \mathbf{y}_{(n)}) | n = 1, \dots, N_{\text{data}}\}$ を用いて NMRA により同定する。ここで, \mathbf{T} は転置を表す。

\mathbf{k}_i を i 番目の基底関数のインデックスベクトル (付録 1 参照), 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, 目的変数ベクトル \mathbf{y} の推定値 $\tilde{\mathbf{y}}$ を計算するための近似関数 \tilde{f} は, 以下の非線形重回帰モデルで表される。

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

ここで, $d_y \times N$ 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。

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

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。

$$\bar{\varepsilon} \stackrel{\text{def}}{=} \frac{E[(y_{E(d(n))} - \tilde{y}_{E(d(n))})^2 | (y_{E\text{max};d} - y_{E\text{min};d})^2]}{d \in \{1, \dots, d_y\}, n \in \{1, \dots, N_{E\text{data}}\}} \dots \dots \dots (4)$$

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, $y_{E\text{max};d}$ と $y_{E\text{min};d}$ は, 各々 $y_{E(d(n))}$ 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。この場合, インデックスベクトル $\mathbf{k}_i (i \in \{1, \dots, N\})$ の要素は, Legendre 関数を構成する Legendre 多項式の次数 $0, 1, 2, \dots$ (付録 1 参照) に対応するため, 0 を含む自然数となる。

〈2・3〉 GA 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。 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。 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切り捨て選択 (truncation selection) (6) 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, $\mathbf{g}_{N_{\text{renew}}+1}, \dots, \mathbf{g}_{b_{\text{max}}}$ 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 $g_{N_{\text{renew}}+1}, \dots, g_{b_{\text{max}}}$ 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。

本論文では, 各個体の染色体 g_i の要素 (遺伝子) をインデックスベクトル $k_1 \sim k_N$ 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 g_i の m 番目の遺伝子 $g_{i,m}$ を $g_{i,m} \in \{0, 1, \dots, g_{\text{max}}\}$ で定義した。突然変異を起こした遺伝子の値は, $0, 1, 2, \dots, g_{\text{max}}$ の一様乱数で決められる。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。

(2.4) 基底関数のインデックスベクトルの設定と交差検証による適応度の計算 (2.3)節で述べたように, 本論文では, 染色体を表すベクトル g^\dagger の要素 (遺伝子) をインデックスベクトル $k_1 \sim k_N$ の要素に対応させている。そのため, g の次元 (遺伝子の数) m_{max} を説明変数ベクトル 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, $N_{\text{max}} = m_{\text{max}}/d_x$)。染色体 g を N_{max} 個に分割して作られた染色体の断片を $\check{g}_j (1 \leq j \leq N_{\text{max}})$ とする。 \check{g}_j は d_x 次元のベクトルであり, 本論文では, \check{g}_j をインデックスベクトルに対応させる。

例えば, 染色体の長さ $m_{\text{max}} = 15$, x の次元 $d_x = 3$, 展開次数の最大値 $N_{\text{max}} = 5$, 遺伝子の最大値 (インデックスベクトルの要素の最大値) $g_{\text{max}} = 5$ として, GA により $g = (1, 3, 2, 0, 5, 1, 0, 0, 0, 1, 3, 2, 4, 2, 3)^\top$ が得られた場合, g と \check{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, $\check{g}_1 = \check{g}_4$ 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{g}_3 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, \check{g}_3 は不要である。

g 中の不要となる染色体の断片 (Fig. 1 の例では, \check{g}_3 と \check{g}_4) を除去し, g を基底関数のインデックスベクトル集合 $Z \stackrel{\text{def}}{=} \{k_1, \dots, k_N\}$ 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集合 Z の設定

- (1-1) g から, $\check{g}_1, \check{g}_2, \dots, \check{g}_{N_{\text{max}}}$ を作成し, $\check{G} \stackrel{\text{def}}{=} \{\check{g}_1, \check{g}_2, \dots, \check{g}_{N_{\text{max}}}\}$ とする。
- (1-2) \check{G} から, $\check{g}_j = \mathbf{0}$ なる \check{g}_j を削除する ($1 \leq j \leq N_{\text{max}}$)。
- (1-3) \check{G} から, $\forall j > j'$ に対して, $\check{g}_j = \check{g}_{j'}$ なる \check{g}_j を削除する ($1 \leq j' \leq N_{\text{max}}$)。
- (1-4) N を \check{G} の要素数とする。
- (1-5) \check{G} の要素を, 順に k_1, \dots, k_N に割り当てる。

Fig. 1 の例では, Algorithm 1 の結果, $\check{g}_1, \check{g}_2, \check{g}_5$ が $k_1,$

[†] 煩雑な表記を避けるため, (2.3)節における染色体 g_i を, (2.4)節では, g と表記した。

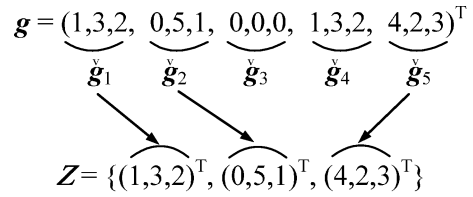


Fig. 1. Example of a map from g to Z .

k_2, k_3 に割り当てられる。すなわち, $k_1 = \check{g}_1, k_2 = \check{g}_2, k_3 = \check{g}_5, N = 3$ となる。Algorithm 1 で染色体 g から基底関数のインデックスベクトル集合 Z を計算する際に, 基底関数の数 N は, $1 \leq N \leq N_{\text{max}}$ の範囲で染色体 g に応じて決まる。そのため, 適応度の計算に交差検証⁹⁾を導入し, 評価データの誤差から染色体 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 $D \stackrel{\text{def}}{=} \{(x^{(n)}, y^{(n)}) | n \in \{1, \dots, N_{\text{data}}\}\}$ を η 分割し, 各々を D_1, \dots, D_η とする。すなわち, $D = \{D_1, \dots, D_\eta\}$ である。次に, $D_{L;\ell} \stackrel{\text{def}}{=} D_\ell$ を学習データ集合, $D_{E;\ell} \stackrel{\text{def}}{=} D - D_\ell$ を評価データ集合とする。以下に, 染色体 g から基底関数のインデックスベクトル集合 Z を設定し, 非線形重回帰分析により得られた近似関数の適応度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 $D_{L;\ell}$ を用いて得られた近似関数を \tilde{f}_ℓ , 評価データ $D_{E;\ell}$ に対する \tilde{f}_ℓ の正規化誤差を $\tilde{\varepsilon}_\ell$, 全評価データ ($\ell = 1, \dots, \eta$) に対する平均正規化誤差を $E[\tilde{\varepsilon}_\ell]$ として, 染色体 g の適応度 J を $E[\tilde{\varepsilon}_\ell] \times (-1)$ と定義する。

[Algorithm 2] 交差検証を導入した適応度の計算

- (2-1) Algorithm 1 に g と N_{max} を入力し, N および 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 $D_{L;\ell}$ から, 式 (3) 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 $D_{E;\ell}$ に対する正規化誤差 $\tilde{\varepsilon}_\ell$ を計算する。
- (2-3) $J = -E[\tilde{\varepsilon}_\ell]$

(2.5) 主成分分析 本論文は, 清酒の成分 (3 章参照) から味データを推定することを目的としている。成分を説明変数とした場合, 成分の次元が非常に高い (41 次元, 3 章参照) ため, 標本数が少ない場合には過学習に陥る可能性が高くなる。本論文では, 主成分分析 (PCA)¹⁰⁾ により次元を圧縮された清酒の成分を説明変数とすることによりこの問題を解決する。

d_u 次元の清酒の成分ベクトルを $u = (u_1, \dots, u_{d_u})^\top$, d_x 次元 ($d_x < d_u$) の説明変数ベクトルを $x = (x_1, \dots, x_{d_x})^\top$ とすると, PCA により x は次式で表される。

$$x = M^\top u \dots \dots \dots (5)$$

ここで, $\mathbf{M} \stackrel{\text{def}}{=} [\mathbf{e}_1, \dots, \mathbf{e}_{d_x}]$ は $d_u \times d_x$ の主軸行列, $\mathbf{e}_1, \dots, \mathbf{e}_{d_x}$ は \mathbf{u} の共分散行列の固有ベクトルである。また, $\|\mathbf{e}_1\| = \dots = \|\mathbf{e}_{d_x}\| = 1$ であり, \mathbf{e}_i の固有値 λ_i は $|\lambda_i| \geq |\lambda_{i+1}|$ である。本論文では, 41 次元の清酒の成分ベクトル \mathbf{u} を式 (5) により 16 次元 (3 章参照) に圧縮し, これを説明変数ベクトル \mathbf{x} とする。

〈2・6〉 PCA と GA を用いた NMRA (2・2) 節で述べた NMRA および (2・4) 節で述べた適応度の計算手法を GA に組み込み, (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。その際, 交差検証により GA の適応度を計算することにより, 過学習を回避し汎化能力が最大になるように基底関数の数と次数が最適化される。また, 式 (5) を用いて清酒の成分ベクトル \mathbf{u} を説明変数ベクトル \mathbf{x} に変換することにより, 過学習を防ぐと同時に GA の解空間を圧縮し, 解を効率良く探索する。以下に, そのアルゴリズムを示す。

[Algorithm 3] PCA による情報圧縮と GA を用いた NMRA による関数近似

- (3-1) PCA(式 (5)) により清酒の成分ベクトル \mathbf{u} を d_x 次元の説明変数ベクトル \mathbf{x} に変換し, データ集合 \mathbf{D} を作成する。
- (3-2) 染色体 $\mathbf{g}_1, \dots, \mathbf{g}_{b_{\max}}$ を初期化する。
- (3-3) Algorithm 2 に, $\mathbf{g}_1, \dots, \mathbf{g}_{b_{\max}}$ および N_{\max} を入力し, 適応度 $J_1, \dots, J_{b_{\max}}$ を計算する。
- (3-4) 適応度の値により $J_1, \dots, J_{b_{\max}}$ および $\mathbf{g}_1, \dots, \mathbf{g}_{b_{\max}}$ をソートする。
- (3-5) $J_1, \dots, J_{b_{\max}}$ 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 $\mathbf{g}_1, \dots, \mathbf{g}_{N_{\text{renew}}}$ を交叉し, 次世代の染色体 $\mathbf{g}_{N_{\text{renew}}+1}, \dots, \mathbf{g}_{b_{\max}}$ 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) $\mathbf{g}_{N_{\text{renew}}+1}, \dots, \mathbf{g}_{b_{\max}}$ の遺伝子を突然変異により変更する。
- (3-8) Step (3-3) 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 (3-3) 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{f}_ℓ は, 回帰係数 A_ℓ, \mathbf{b}_ℓ を式 (2) に代入すると, 次式で表される。

$$\tilde{\mathbf{y}} = \tilde{\mathbf{f}}_\ell(\mathbf{x}) \stackrel{\text{def}}{=} A_\ell \mathbf{K}(\mathbf{x}, \mathcal{Z}) + \mathbf{b}_\ell \dots \dots \dots (6)$$

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{f}_ℓ に対して共通である。

3. 性能評価

〈3・1〉 清酒の成分を用いた味データの推定 本章では, 2015~2017 年度に製造された清酒 113 サンプルをデータ集合 (付録 2 参照) として, 2 章で述べた Algorithm 3 を用いて清酒の成分と味データの関係を導出する。ここで, 清酒の成分ベクトル \mathbf{u} の要素は, 化学分析により測定された一般分析項目 (日本酒度, アルコール度数, 酸度など), 糖 (単糖, 二糖など), 有機酸 (クエン酸, ピルビン酸など), アミノ酸 (アスパラギン酸, トレオニンなど) の 38 種類 (Table 1,2 参照), および糖合計, 有機酸合計, アミノ酸合計の計 41 項目である。味データベクトル \mathbf{y} の要素は, 味覚センサにより測定された, 酸味, 苦味雑味, 渋味刺激, 旨味, 塩味, 苦味, 渋味, および旨味コクの 8 項目である。尚, 本論文で用いられる全てのデータは, 各々の標準偏差で正規化されている。

味データの測定は, 味認識装置 SA-402B (株インテリジェントセンサーテクノロジー) により行われた。まず, 人工脂質膜型味覚センサを基準液 (30 mM KCl, 0.3 mM 酒石酸) に浸して膜電位 V_f を測定した。次に, 清酒サンプルにセンサを浸して膜電位 V_s を測定した。その後, センサを基準液で洗浄し, 再び基準液中で膜電位 V_f' を測定した。これらの測定は, サンプルを配置した水槽に 30°C の水を循環させて行われた。測定に用いられた味覚センサは, 先味 (酸味, 塩味, 旨味, 苦味雑味, 渋味刺激) に対して SB2CA0, SB2CT0, SB2AAE, SB2C00, SB2AE1 の 5 種類, 後味 (旨味コク, 苦味, 渋味) に対して SB2AAE, SB2C00, SB2AE1 の 3 種類である。

測定された膜電位, センサ毎に決められた標準サンプルの膜電位, および官能評価のしきい値から味の強さ (味データ) が導かれる。味データの 1 単位は, 人間が識別できる最小濃度差を基に設定される⁽⁹⁾。

Algorithm 3 の主なパラメータは, $m_{\max} = 256, b_{\max} = 2048, N_{\max} = 16, g_{\max} = 2, p_M = 10^{-3}, \eta = 2, d_x = 16$ である。 N_{\max} の増加に従い $E[\varepsilon_\ell]$ は減少する。しかし, その一方で, 式 (3) において COV_{KK} の逆行列の計算が不安定になる。今回の実験では, 清酒のサンプル数, 逆行列計算の安定性, および近似精度を考慮して $N_{\max} = 16$ とした。また, 清酒の成分ベクトル \mathbf{u} の次元 ($d_u = 41$) に対して標本数が少ない (113 サンプル) ため, NMRA が過学習に陥る可能性が高い。この問題を回避するために, Algorithm 3 では PCA ((2・5) 節参照) により \mathbf{u} を圧縮し, NMRA の説明変数ベクトル \mathbf{x} とした。 \mathbf{x} の次元 d_x は, 式 (6) の汎用性と精度を考慮して 16 とした。 m_{\max} は, (2・4) 節に記載の関係式 ($N_{\max} = m_{\max}/d_x$) より求めた。 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る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⁽¹⁰⁾。

Algorithm 3 により得られた平均正規化誤差 $E[\varepsilon_\ell]$ を Table 3 に示す。なお, GA では, 近似精度が染色体 $\mathbf{g}_i (1 \leq i \leq b_{\max})$

Table 1. Components-1.

General analysis	Organic acid	Saccharide
Nihonshu-do	Citric acid	Monosaccharide
Alcohol content	Pyruvic acid	Disaccharide
Acidity	Malic acid	Trisaccharide
Amino acid content	Succinic acid	Tetrasaccharide
Glucose	Lactic acid	Pentasaccharide or more
Aldehyde	Acetic acid	Glycerol

Table 2. Components-2.

Amino acid		
Aspartic acid	Threonine	Serine
Asparagine	Glutamic acid	Glutamine
Proline	Glycine	Alanine
Valine	Cysteine	Methionine
Isoleucine	Leucine	Tyrosine
Phenylalanine	Gamma-Amino Butyric Acid	Lysine
Histidine	Arginine	

Table 3. Mean normalized errors.

Taste	Algorithm 3	MRA
Sourness	3.54×10^{-3}	3.01×10^{-2}
Bitterness	2.46×10^{-3}	2.26×10^{-2}
Astringency	2.39×10^{-3}	2.82×10^{-2}
Umami	3.74×10^{-3}	2.44×10^{-2}
Saltiness	8.49×10^{-4}	9.07×10^{-3}
Aftertaste of Bitterness	1.93×10^{-3}	4.62×10^{-2}
Aftertaste of Astringency	3.42×10^{-3}	2.67×10^{-2}
Richness	1.28×10^{-3}	2.46×10^{-2}

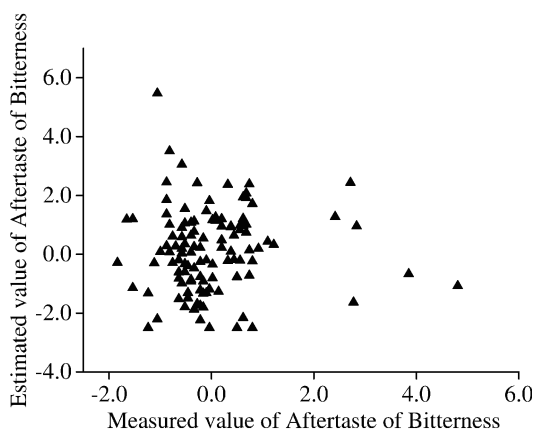


Fig. 2. Estimated values to measured values of Aftertaste of Bitterness (MRA).

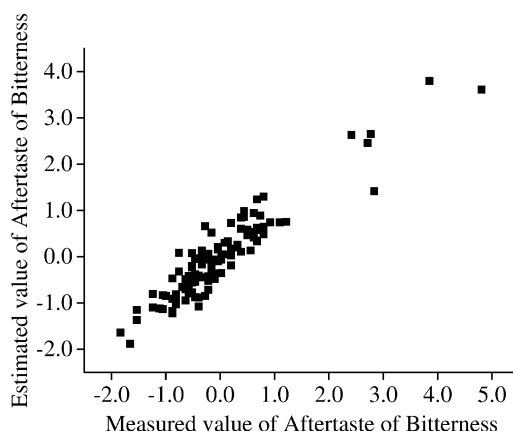


Fig. 3. Estimated values to measured values of Aftertaste of Bitterness (Algorithm 3).

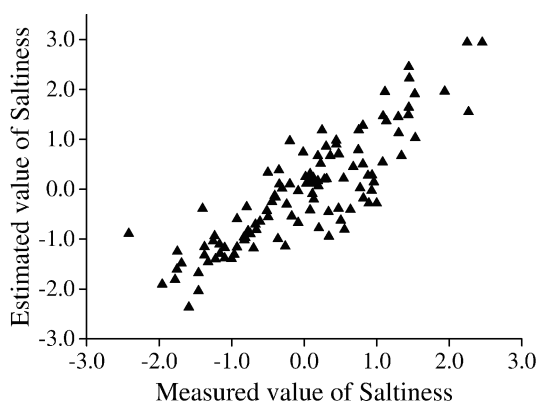


Fig. 4. Estimated values to measured values of Saltiness (MRA).

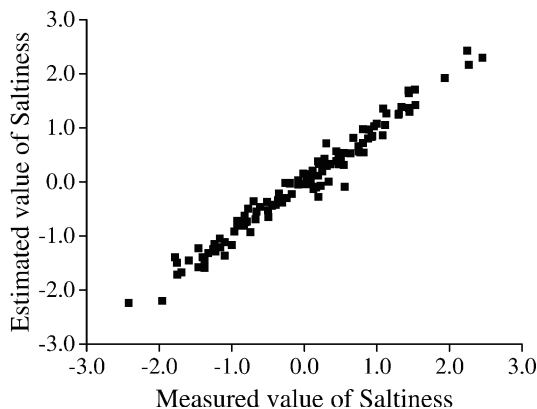


Fig. 5. Estimated values to measured values of Saltiness (Algorithm 3).

の初期値により左右されるため、染色体 g_i の初期値を変えて 4 回計算を行い、最も良い (小さい) $E[\bar{e}_\ell]$ を Table 3 に記載した。比較のため、線形重回帰分析 (MRA) による平均正規化誤差も Table 3 に示す。Table 3 より、MRA に比べて Algorithm 3 は平均正規化誤差が約 10 分の 1 に減少していることがわかる。

MRA により得られた平均正規化誤差が最も大きい苦味 (Aftertaste of Bitterness) の実測値と推定値の関係 (散布図) を Fig. 2 に、Algorithm 3 により得られた苦味の実測値と推定値の関係を Fig. 3 に示す。また、MRA により得られた平均正規化誤差が最も小さい塩味 (Saltiness) の実

測値と推定値の関係を Fig. 4 に、Algorithm 3 により得られた塩味の実測値と推定値の関係を Fig. 5 に示す。これらの図から Algorithm 3 は、清酒の成分から味データを精度良く再現していることがわかる。

〈3・2〉 味データと清酒の成分の関係の応用 本章では、特定の成分を清酒に添加した際の味データの変化量の推定値を味覚センサの応答の確認や清酒の醸造プロセスの改良に応用する可能性を探る。

Table 4. Measured and estimated values of the variations in Sourness.

Additive component	Malic acid	Citric acid	Glucose
Measured value	1.307	-0.310	-0.370
Estimated value	0.319	-0.237	-0.039

まず、味との関係が明確であり、かつ清酒に十分な分量が含まれているリンゴ酸とクエン酸に対する酸味の変化量を、以下の手順に従い味覚センサにより計測した。

- (1) 成分の添加量を決定する。添加量は、添加後の成分の分量が〈3・1〉節のデータ集合 (113 サンプルの清酒) の成分の分布を逸脱しない範囲に収めるために、データ集合の成分の最大値と基準酒[†]の成分の差の $1/3$ とする。
- (2) 所定の添加量の成分を基準酒に添加し、酸味を計測する。
- (3) 酸味の変化量 (添加後の酸味と添加前の酸味の差分) を求める。

同様な手順に従い、清酒の甘味を決める最も重要な成分であるグルコースに対する酸味の変化量も計測した。甘味は酸味を抑える働きがあるため、これにより、酸味センサに対するグルコースのマスキング効果が調べられる⁽⁴⁾。

次に、上記酸味の変化量の推定値を計算した。尚、目的変数は酸味、説明変数はリンゴ酸、クエン酸、およびグルコース、回帰係数の算出には〈3・1〉節のデータ集合を用いた。説明変数を上記3変数に限定した理由は以下の通りである。

- (1) 38次元の清酒の成分と3次元の統計量からなるデータを16次元に圧縮し、それらを説明変数とした場合、ある特定の清酒の成分を変化させた場合でも、その成分と相関が強い成分も同様に変化させていることになる。そのため、特定の成分のみを添加した実測値を正確に再現できない。
- (2) 38次元の清酒の成分と3次元の統計量を説明変数とした場合、説明変数間の相関に起因する多重共線性により、ある特定の説明変数と目的変数の関係を正確に再現できない。また、説明変数の次元に対して標本数が少ないため、分析の汎用性が低い。

上記酸味の変化量の推定値と実測値を、Table 4 に示す。Table 4 より、推定値と実測値の両者で、リンゴ酸 (Malic acid) の添加により酸味が増加し、グルコース (Glucose) の添加により酸味が減少している。これらの結果は、リンゴ酸とグルコースに対する酸味センサ出力および官能評価と一致する。一方、クエン酸 (Citric acid) の添加では、推定値と実測値の両者とも、酸味が減少している。味覚センサは、類似した味に類似した応答をする広域選択性を持っているため⁽⁴⁾⁽⁵⁾、酸味の強いクエン酸の添加により酸味センサの出力は強くなるはずである。しかし、この実験結果

は、クエン酸の添加に対して想定される酸味センサの出力および官能評価とは異なる結果となっている。また、酸味の推定値が減少していることから、この傾向は、〈3・1〉節のデータ集合の清酒全体に共通していることが分かる。そのため、これらの結果は、クエン酸と何らかの成分 (清酒に含まれる他の成分) との相互作用により酸味センサがクエン酸単体に対する反応と異なる反応を示す現象、すなわち酸味センサに対するマスキング効果の存在を示唆している。

リンゴ酸およびグルコースの添加に対する酸味の変化量の推定値は、実測値ほど大きくない。すなわち、推定精度が低い。これは、回帰係数の算出に用いられた〈3・1〉節のデータ集合の清酒には、基準酒からリンゴ酸のみまたはグルコースのみが添加量だけ増加したサンプルが存在しないためであると考えられる。

これらの分析から、以下の知見が得られる。

- (1) 特定の成分を添加することによる味データの変化量の推定値の増減が、想定される味の変化量の増減 (官能評価または味データの実測値等) と異なる場合 (クエン酸の場合):
 - 添加された成分と何らかの成分 (清酒に含まれる他の成分) との相互作用により、添加された成分に対する味覚センサの応答がマスキングされている可能性がある。
- (2) 味データの変化の推定値の精度が悪い場合 (リンゴ酸およびグルコースの場合):
 - 分析に用いられた清酒のサンプルには、当該成分のみが添加量だけ増加したサンプルが存在しないことから、分析に用いられた清酒のサンプルの醸造法では、当該成分のみが増加した清酒が醸造されていない。
 - そのため、仮に、当該成分のみが増加した清酒の味の評価が高い場合でも、そのような清酒の醸造は困難である可能性がある。

4. まとめ

清酒の成分と味覚センサの出力 (味データ) の関係を、非線形重回帰分析により Legendre 関数を基底関数とする多項式で表し、成分から味データを高い精度で推定できることを示した。また、特定の成分が添加された基準酒の味データの変化量の推定値と想定される味の変化量 (官能評価または味データの実測値) を比較することにより、味覚センサの応答の精度を推定できる。さらに、成分に着目した醸造プロセスの改良の可能性を判断するためのデータを提供できることを示した。今後、分析に用いられる清酒のサンプルを増やし、さらに、様々な成分の添加実験を実施することにより、味覚センサおよび化学分析による分析結果を、醸造プロセスにフィードバックしていく予定である。

[†] データ集合の中の標準的な清酒であり、味覚センサの調整に用いられている。

文 献

- (1) K. Toko: "Taste Sensor", Trans. JASTS, Vol.4, No.1, pp.21-32 (1997) (in Japanese)
都甲 潔:「味覚センサ」, 日本味と匂学会誌, Vol.4, No.1, pp.21-32 (1997)
- (2) Y. Kobayashi, M. Habara, H. Ikezaki, R. Chen, Y. Naito, and K. Toko: "Advanced Taste Sensors Based on Artificial Lipids with Global Selectivity to Basic Taste Qualities and High Correlation to Sensory Scores", Sensors, Vol.10, No.4, pp.3411-3443 (2010)
- (3) 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)
- (4) N. Cristianini and J.S. Taylor: "An Introduction to Support Vector Machines and Other Kernel-based Learning Methods", Cambridge Univ. Press, UK (2000)
- (5) D.E. Goldberg: "Genetic Algorithms in Search Optimization and Machine Learning", Addison-Wesley, USA (1989)
- (6) 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 SM, 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)
- (7) R.A. Johnson and D.W. Wichern: "Applied Multivariate Statistical Analysis 5th ed.", Pearson Education (Prentice Hall), USA (2001)
- (8) 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)
- (9) 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)
- (10) M. Satoh, H. Satoh, and H. Ikezaki: "Optimization of Food Ingredients and their Blend Ratios Based on Taste Sensor Output", IEEJ Trans. on SM, Vol.135, No.2, pp.57-64 (2015) (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)
- (11) H. Fukui, T. Ishida, T. Nishimura, and H. Matsuda: "Correlation between the Results of a Sensory Test and an Instrumental Analysis of the Effect of Mirin for Suppressing Saltiness and Sourness", J. Cookery Science of Japan, Vol.39, No.1, pp.49-56 (2006) (in Japanese)
福井 裕・石田丈博・西村敏英・松田秀喜:「本みりんの塩味・酸味抑制効果における官能評価と味覚センサーの分析結果との相関性」, 日本調理科学会誌, Vol.39, No.1, pp.49-56 (2006)
- (12) I.N. Bronshtein and K.A. Semendiyayev: "Handbook of Mathematics", Springer-Verlag, UK (1997)
- (13) 石川雄章編:「増補改訂清酒製造技術 新版」, 日本醸造協会 (2009)
- (14) 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」, 日本醸造協会誌, Vol.69, No.11, pp.774-777 (1974)
- (15) H. Utsunomiya, A. Isogai, and H. Iwata: "Amakara Categories for Type Designation", J. of Brewing Society of Japan, 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

$\mathbf{x} \stackrel{\text{def}}{=} (x_1, \dots, x_{d_x})^T$, $\mathcal{D}_x \stackrel{\text{def}}{=} \{\mathbf{x} | x_{\min d} \leq x_d \leq x_{\max d}, 1 \leq d \leq d_x\}$ とする。本章では, $\mathbf{x} \in \mathcal{D}_x$ 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⁽¹²⁾。

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

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

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。

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

$\{K_d(x_d, k_d)\}$ 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。

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

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。

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

基底関数 $\phi_i(\cdot)$ を次式で定義する。

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

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, $\mathcal{Z}_d \stackrel{\text{def}}{=} \{0, 1, \dots, N_d\}$, \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。

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

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

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

app. Table 2. 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	ginjo-shu or daiginjo-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 3. Classification of samples of sake.

Taste of sake	Sweet	28
	Dry	85
Brewing method	Alten	53
	Junmai	42
	Tousan	18
Rice-polishing ratio	H-RPR	83
	L-LPR	30

2. 清酒データ

清酒は、製法品質表示基準に示された精米歩合や麴米の使用割合、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⁽¹³⁾。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⁽¹⁴⁾、甘辛度⁽¹⁴⁾、新甘辛度⁽¹⁵⁾などが提案されてきた。本論文で用いられた清酒113サンプルを、精米歩合、添加物、新甘辛度に基づく区分に従い分類し、それらの味データと成分の統計情報を app. Table 1 に示す。

また、精米歩合、添加物、新甘辛度に基づく区分の詳細を app. Table 2 に示す。ここで app. Table 2 において、AV は新甘辛度、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。また、本論文で用いられた清酒113サンプルを新甘辛度、添加物、および精米歩合で分類した際のサンプル数を app. Table 3 に示す。

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(株)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(株)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。情報ノ宮路の下工房代表。



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(株)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。



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(株)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(株)ワイ・アール・ピ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。

