Machine Learning in Analyses of the Relationship between Japanese Sake Physicochemical Features and Comprehensive Evaluations

Satoru SHIMOFUJI1,2,†, Motoko MATSUI2, Yukari MURAMOTO3, Hironori MORIYAMA1, Reina KATO1, Yoshiro HOKI1, Haruhiko UEHIGASHI1

1 Kochi Prefecture Industrial Technology Center, 3992-3, Nunoshida, Kochi-shi, Kochi 781-5101, Japan, 2 Department of Food Sciences and Nutritional Health, Graduate School of Kyoto Prefectural University, 1-5 Shimogamohangi-cho, Sakyo-ku, Kyoto 606-8522, Japan, 3 Kyoto Prefectural University, 1-5 Shimogamohangi-cho, Sakyo-ku, Kyoto 606-8522, Japan

We investigated the contributions of physicochemical features to a comprehensive evaluation of the Japanese sake known as ‘Junmai Ginjo’ by applying machine learning. We used 173 samples of the commercial Japanese sake. The sensory evaluation was conducted by 35 panelists. The panel conducted the evaluation of each sample using five statements for the comprehensive evaluation of the sample. General analysis, substance-related nucleic acid, volatile components and simplified analyses were measured as physicochemical analyses. We performed regression analyses using a multiple regression analysis (MRA), partial least squares regression (PLS) and machine learning employing a support vector machine (SVM), an artificial neural network (ANN), and random forest (RF). The results of these five analysis methods have demonstrated that machine learning (especially RF) provides comparable or higher prediction accuracy and better fitting than MRA. We also discuss the contribution of each physicochemical feature to the evaluation scores based on the regression coefficients obtained by MRA and the features’ importance obtained in RF. The analysis of the individual scores indicated that ethyl caproate and isoamyl acetate make large contributions to influence the sake evaluation.

Keywords: machine learning, random forest, Junmai Ginjo, sensory evaluation, comprehensive evaluation

1. Introduction

Evaluations of the qualities of food (e.g., the palatability, preferences, quality, and grade) are comprehensive evaluations based on various factors. Shimada [1] proposed that there are three types of factors that humans use to evaluate the palatability of food: (1) direct factors, which concern the physicochemical features of food; (2) indirect factors, which are dependent on the humans’ physiological and psychological states; and (3) background factors, which are dependent on human knowledge and experience (e.g., food culture, information about the food). In the field of food science, researchers try to clarify the contribution of direct factors to a comprehensive evaluation directly, by controlling the indirect factors as much as possible. This control is achieved by actions such as setting the conditions of the evaluation environment, training the evaluators, and collecting and considering subjects’ background factors by questionnaire.

Sensory evaluation data and physicochemical data are used for analyses of comprehensive evaluations of food. In many cases, a multiple regression analysis (MRA) is performed with physicochemical data used as the explanation variables and average scores from the evaluation panel used as objective variables. Multiple regression coefficients are used to compare the contribution of food attributes to palatability as rated by a panel.

MRA makes it easy to interpret the contribution of variables, but also has problems such as multicollinearity. [2] Therefore, partial least squares regression (PLS) is often used like MRA. [3] However, MRA and PLS take into account the complex relationship among the many components of food, but it is a linear regression analysis. The criteria for a comprehensive evaluation of food are often nonlinear of the food components’ concentrations. For example, the most preferred food sample is not the sample that is the most salty; palatability is nonlinear of the salt concentration. Food evaluation results also vary.
widely among individuals. Sagara [4] thus proposed that various analytical methods could be combined with an MRA for palatability analyses.

Machine learning has been shown to be useful for the prediction of nonlinear models, as a data mining method. It has high flexibility, high prediction accuracy, and provides a nonlinear analysis. Machine learning has been applied in a wide range of fields such as bioinformatics [5], the automatic classification of documents [6], and image recognition [7]. Examples of the machine learning used in these fields are support vector machine (SVM), artificial neural network (ANN), and random forest (RF) machine learning.

The SVM algorithm was proposed by Vapnik [8]. This algorithm works to find an optimal separating hyperplane between classes [9]. An ANN is an analytical method that models the human brain’s neurons as a computer-based attempted replication of the brain’s information processing [10]. The RF method is based on a classification algorithm developed by Leo Breiman [11] that uses an ensemble of classification and regression trees. There are many other methods of machine learning, and a study comparing machine learning methods have also been reported [12].

Machine learning has also been used in the field of food science. Cortez et al. [13] applied machine learning to predict evaluation values from analysis values of wine, and they compared the prediction accuracy of MRA, ANN and SVM. Qiu et al. [14] applied to multivariate statistical methods (linear discriminant analysis and partial least squares regression (PLS)) and machine learning (RF and SVM) to qualitative classification and quantitative regression of five types of strawberry juices based on processing approaches. Sugimoto et al. [15] reported the application of machine learning in the analysis of the relationship between the metabolite profiles and sensory evaluation of zatsumi in refined Japanese sake. This study compares the results of analysis with MRA and SVM.

Thus, even in the field of food science, there are many reports that use machine learning to predict sensory evaluation scores. However, in situations such as product development and quality improvement, it is necessary not only to model more accurately relationship between the comprehensive evaluation and the physicochemical analysis values, but also to understand the evaluation trends of the panel more clearly.

Therefore, we conducted the present study to determine the contributions of physicochemical features to the comprehensive evaluation of food by applying machine learning as a nonlinear analysis method. We thought that a more accurate estimation model could be obtained by comparing with a general statistical method (MRA and PLS) and a typical machine learning method (ANN, SVM and RF). Furthermore, in order to quantify the evaluation tendency, we considered the importance of variables obtained from machine learning and the regression coefficients obtained from MRA. This is because the analysis process is black-box-like simply by applying machine learning, and it is difficult for a person to understand the contribution of variables.

We used a commercially available brand of sake classified as ‘Junmai Ginjo,’ in which strong quality defects (bad taste, offensive odor, etc.) have been removed by preliminary examination of sensory analysis, for our analysis of the comprehensive evaluation by several machine learning methods. Junmai Ginjo is made only with rice (its polished rice ratio is <60%), rice koji (i.e., rice malt or yeast), and water (without the addition of brewers alcohol).

Our reasons for using this dataset (data of physicochemical analyses and sensory evaluation) are as follows. (1) Sake is a palatable drink, and it is easy to analyze the panelist’s trends in a comprehensive evaluation of sake. (2) The 35 panelists we used were sake makers, retailers, or researchers close to the expert panel. (3) One panelist can evaluate many sake samples on the same basis. (samples: \( n=173 \)) (4) The variables representing the characteristics of sake are commonly understood.

We also attempted to quantify the tendency of individual deviations in palatability by comparing the analysis results by both the average score and the individual evaluation score.

2. Materials and Methods

2.1 Samples

We used 173 samples of the commercial Japanese sake Junmai Ginjo. These samples were produced in the 2017 brewing year. We received these samples from Liquor shop (Hasegawa Saketen, Inc.). These samples were stored at \(-1^\circ C\) for quality preservation until they were analyzed and evaluated. The samples were opened and immediately subjected to evaluation and analysis.

2.2 Sensory evaluation

The sensory evaluation was conducted by 35 panelists.
who were engaged in the research, manufacture, and/or sale of sake. Each sake sample was placed in a ceramic sake cup and served at room temperature (around 20°C). The panelists were blinded to the bland names to eliminate prejudicial bias. The panel conducted the evaluation of each sample using five statements for the comprehensive evaluation of the sample. The panelists were asked to rate the sample by choosing one of the following statements to describe the sake: (1) The expression of harmony and the characteristics of flavor are excellent as drinking characteristics of sake. (2) The harmony and the characteristics of flavor are good as drinking characteristics of sake. (3) The harmony and characteristics of flavor are ordinary (average) as drinking characteristics of sake. (4) One or more aspects of the sake sample other than statements (1–3) are somewhat defective. (5) One or more aspects other than statements (1–4) are defective. These five statements corresponded to 1 to 5 points, respectively. The average score for each sample was calculated from the 35 individual panelists’ score.

2.3 Physicochemical Analysis

2.3.1 General analyses

We measured the sake samples’ acidity, amino acid content, and glucose content as indicators of the taste of the sake. The acidity and amino acid content were analyzed by the methods authorized by Japan’s National Tax Agency [16]. The glucose content was determined by the mutarotase glucose oxidase method [17] using the Wako Glucose CII-test (Wako Pure Chemicals, Osaka, Japan).

2.3.2 Substance-related nucleic acid

The optical density (OD) values 260 and 280 were measured as described by Yamanaka et al. [18]. These are indicators showing the degree of leakage of nucleic acid components due to the death of yeast during the manufacturing process, and they are thought to affect the quality evaluation.

2.3.3 Volatile components

Volatile components were measured as aspects of the flavor of the sake. Acetaldehyde, ethylacetate, propan-1-ol (nProOH), isobutanol (iBuOH), isoamylacetate, 3-methyl-1-butanol (iAmiOH), ethyl caproate, ethyl caprylate, and caproic acid were measured by head-space gas chromatography (GC) (model 7890B, Agilent, Santa Clara, CA) according to the method of Yoshizawa et al. [19]. First, 2 mL of sample and 200 μL of internal standard solution (2000 ppm n-amyl alcohol, 200 ppm methylcaproate) were placed in a vial and sealed with an aluminum cap. Head space GC was then used to analyze these components.

2.3.4 Simplified analyses

As a simplified analysis, the samples’ Brix was measured by Brix meter (PAL-J, Atago, Tokyo, Japan); the pH was measured by a pH meter (LAQUAtwin-pH22, Horiba, Fukuoka, Japan), and electrical conductivity was measured by an electrical conductivity meter (LAQUAtwin-EC-33, Horiba).

2.4 Regression analyses

We performed separate regression analyses using MRA, PLS and machine learning: ANN, SVM, and RF. Physicochemical analysis values were used as explanatory variables. These analysis values were standardized to a zero mean and one standard deviation (SD) for the regression analyses. In MRA, there is a problem of multicollinearity by using one with high correlation between explanatory variables. Therefore, before regression analysis, Pearson’s correlation coefficient and variance inflation factor (VIF) [20] between physicochemical analysis values were confirmed. Suspected multicollinearity when variables with VIF greater than 10 are included.

The average score and the 35 individual scores of sensory evaluation data were used as objective variables. The regression analyses were carried out by R software [21]. The R development environment is a multi-platform open-source environment that can be operated by Windows and Linux. R is a high-level matrix programming language for statistical processing and data analysis.

All of the regression methods were implemented using the R package caret [21], which provides a good interface to access many machine-learning algorithms in R. The caret package tunes the parameters of statistical analyses and machine learning, and selects the values which maximize the accuracy according to the validation selected (leave-one-out, k-fold, etc.) [22]. The method parameters of caret are as follows: for MRA, lm (stats package) ii); for PLS, pls (pls package) [23]; for ANN, nnet (nnet package) [24]; for SVM, svmRadial (kernlab package) [25]; and for RF, rf (random forest package) [26]. In each of the machine learning methods, we performed 10-fold cross-validation to generalize the model, and a prediction model was created. Other analysis condition settings (hyperparameters, etc.) used the default ones in the caret package.

2.5 Accuracy of the regression analyses

The accuracy of the regression analysis was verified by

© 2020 Japan Society for Food Engineering
dividing it into train data and test data to avoid over-learning. First, 158 products (90% of the total: 173 products) were randomly divided into training data, and the remaining 15 products were used into test data for accuracy verification. Next, a regression analysis was performed using the training data to obtain a prediction model. From the obtained prediction model, we calculated the prediction value of the comprehensive evaluation of test data (data not used to create the prediction model), compared it with the actual measurement value, and examined the accuracy of each analysis method. In addition, we also examined the fitting of the prediction model by comparing the prediction accuracy of the training data in the same way.

For the performance of each regression analysis, we used an error metric, i.e., the mean absolute error (MAE) with reference to Cortez et al. [13]. For each sample, the absolute value of the difference between the predicted values in the regression model and the actual measured values was obtained. The percentage of samples included within error tolerance 'T' was taken as the prediction accuracy of the analysis model. We set the error tolerance as 0.25 in the average score analysis and 0.5 in the individual score analysis based on the range of scores. The individual-score results for single panelists were as follows (mean ± SD): Score 1, 13 ± 10; score 2, 50 ± 16; score 3, 75 ± 19; score 4, 32 ± 17; score 5, 3 ± 6. The individual scores are integer data, so we set the tolerance for the individual scores as 0.5 based on the correct answer rate with a rounding off of the predicted value.

2.6 Contribution of each feature to the sensory evaluation

The contribution of each feature of the sake to the sensory evaluation was estimated from the standardized regression coefficient obtained by the MRA and the importance obtained by the RF method. The importance in the RF method is measured by the average increase in squared 'out-of-bag' (OOB) residuals when the variable is permuted [26].

In order to capture the panel’s evaluation trends for this examination, the contribution of sake to each feature was verified using a 5-fold cross-validated regression analysis model for all sample data.

3. Results and Discussions

3.1 Statistics and the distribution of average scores and individual scores

The 35 panelists’ average scores for the Junmai Ginjo sake samples were as follows: minimum, 2.11; maximum, 3.86; 1st quantile (25%), 2.51; and 3rd quantile (75%), 3.02. The average scores thus ranged from 2.11 to 3.86. In this analysis, we set the tolerance for the accuracy of the regression analyses of the average score to 0.25 based on the range of values and the distribution. (To approximate 5 steps: 2.0 – 2.5 – 3.0 – 3.5 – 4.0) The individual-score results for single panelists were as follows (mean ± SD): Score 1, 13 ± 10; score 2, 50 ± 16; score 3, 75 ± 19; score 4, 32 ± 17; score 5, 3 ± 6. The individual scores are integer data, so we set the tolerance for the individual scores as 0.5 based on the correct answer rate with a rounding off of the predicted value.

3.2 Physicochemical analyses

Table 1 provides the results of the physicochemical analysis and VIF. The correlation coefficient between each physicochemical analysis data was from -0.69 to 0.93. All variables were used for regression analysis, as VIF was less than 10 (1-7) across all analysis data.

3.3 Accuracy of the regression analyses

The data in Table 2 explain the prediction accuracy of each of the five regression analyses for test data of the average scores. The accuracy T-value (0.25) of test data was in 73.3 for MRA, ANN, SVM and RF, 60.0 for PLS. The prediction accuracy of PLS was lower than other methods, but there was no difference between MRA and machine learning.

The MAE and RMSE are indicators of the degree of error, and the smaller the MAE and RMSE values are, the higher the prediction accuracy is. In the SVM and RF, MAE and RMSE values were slightly smaller than those of MRA, as follows. MAE: MRA 0.26, PLS 0.28, ANN 0.27, SVM 0.24, rf 0.25; RMSE: MRA 0.37, PLS 0.38, ANN 0.39, SVM 0.34, RF 0.35. The data in Table 3 explain the fitting of regression model for train data of the average score. The accuracy T-value (0.25) of training data was 64.6 for MRA and 60.1 for PLS, whereas the corresponding values for the machine learning methods were 67.1 for ANN, 80.4 for SVM, and 92.4 for RF. Since the error in the prediction accuracy of the test data is small, it is considered that the analysis using machine learning...
(especially SVM and RF) captures the evaluation tendency more clearly than MRA. Figure 1 illustrates the relationship between the predicted values and the measured values in each regression analysis with respect to the average score. The reason why the prediction accuracy of the test data did not exceed 73.3 is that the prediction accuracy of the sample with a high average score is low, as can be seen from the prediction of test data of each regression analysis. In order to further improve the prediction accuracy, comprehensive analysis data such as metabolome analysis, electronic nose and an electronic tongue analysis are necessary. The prediction of training data with the RF method is a sharp plot showing that the fitting is better than the other method. The MAE and RMSE values of RF are smaller than those obtained with the other four analysis methods.

Table 4 shows the prediction accuracy of each regression analyses for test data with respect to the individual scores. The accuracy T-value (0.50) of training data was 41.3 for MRA and 43.2 for PLS, whereas the corresponding values for the machine learning methods were 42.3 for ANN, 43.4 for AVM, and 44.6 for RF. The prediction accuracy of SVM and RF was higher than MRA and PLS, but no significant differences were found. In the ANN,
Fig. 1 The accuracy of the prediction score by the five regression analyses and the measured scores in the average score of the 35 panelists rating samples of *Junmai Ginzō* sake (*n*=173).

○: training data. ◆: test data.
A: MRA (lm). B: PLS (pls). C: ANN (nnet). D: SVM (svmRadial). E: RF (rf).
MAE and RMSE values were slightly higher than those of other methods, but did not show any significant differences as in the prediction accuracy. The reason that the accuracy of prediction in individual scores was lower than the average score is that the data of the objective variable used in the analysis was continuous data in the average score and discontinuous data (5-level score) in the individual score.

The data in Table 5 explain the fitting of regression model for train data of the individual score. The accuracy T-value (0.50) of training data was 51.4 for MRA and 48.7 for PLS, whereas the corresponding values for the machine learning methods were 53.5 for ANN, 55.8 for SVM, and 83.0 for RF, demonstrating that the fitting of prediction model of the machine learning was better than that of the MRA and PLS.

These results demonstrated that the RF method provides significantly better fitting of prediction model than other regression analyses. An example of the relationship between the predicted values and the measured values in each regression analysis for the individual scores is shown in Figure 2. As observed for the regression analyses of the average scores, the prediction of individual scores by the RF is shown as a sharp plot in training data because the MAE and RMSE values with this method are small.

Together these results indicate that the random forest prediction model can grasp the characteristics of the panellists more accurately than the other analysis methods with respect to the current data set. The RF method is said to have the following characteristics:

(a) It can be used when there are many more variables than observations.
(b) It can be used both for two-class and multi-class problems of >2 classes.
(c) It provides a good predictive performance even when most of the predictive variables are noise, and it thus does not require a pre-selection of variables.
(d) It does not overfit.
(e) It can handle a mixture of categorical and continuous predictors.
(f) It incorporates interactions among predictor variables.
(g) There are high-quality and free implementations, i.e., the original Fortran code from Breiman and Cutler iii), and an R package from Liaw and Wiener [26].
(h) It returns measures of variable importance.
(i) There is little need to fine-tune parameters to achieve excellent performance. The most important parameter to choose is mtry (the number of input variables tried at each split), but Liaw and Wiener [26] reported that the default value is often a good choice. In addition, the user must decide how many trees to grow for each forest (ntree) as well as the minimum size of the terminal nodes (nodesize).

These characteristics of RF are speculated to be the factors that make it possible to accurately capture the characteristics of the evaluator. In the classification problem, the prediction accuracy of the RF is high [12]. Our present findings demonstrate that the RF method was also effective for analyzing the relationship between the results of a comprehensive evaluation of sake and the sake's physicochemical features.

### Table 4  Prediction accuracy of regression analyses for test data of the individual score

| Regression analyses | MRA | PLS | ANN | SVM | RF |
|---------------------|-----|-----|-----|-----|----|
| Caret method        | lm  | pls | nnet| svmRadial | rf |
| Accuracy $T=0.50$   | $41.3 \pm 14.1$ | $43.2 \pm 13.5$ | $42.3 \pm 11.7$ | $43.4 \pm 15.3$ | $44.6 \pm 15.0$ |
| MAE                 | $0.69 \pm 0.15$ | $0.68 \pm 0.15$ | $0.72 \pm 0.16$ | $0.66 \pm 0.16$ | $0.69 \pm 0.16$ |
| RSME                | $0.84 \pm 0.16$ | $0.84 \pm 0.16$ | $0.89 \pm 0.18$ | $0.83 \pm 0.16$ | $0.84 \pm 0.17$ |

$n=35$.

### Table 5  Fitting of regression model for train data of the individual score

| Regression analyses | MRA | PLS | ANN | SVM | RF |
|---------------------|-----|-----|-----|-----|----|
| Caret method        | lm  | pls | nnet| svmRadial | rf |
| Accuracy $T=0.50$   | $51.4 \pm 9.1^{ab}$ | $48.7 \pm 9.6^{a}$ | $53.5 \pm 8.2^{ab}$ | $55.8 \pm 10.0^{b}$ | $83.0 \pm 7.7^{c}$ |
| MAE                 | $0.58 \pm 0.11^{ab}$ | $0.61 \pm 0.13^{a}$ | $0.56 \pm 0.11^{ab}$ | $0.51 \pm 0.13^{b}$ | $0.29 \pm 0.05^{c}$ |
| RSME                | $0.73 \pm 0.13^{a}$ | $0.76 \pm 0.14^{a}$ | $0.71 \pm 0.13^{a}$ | $0.69 \pm 0.14^{a}$ | $0.36 \pm 0.06^{b}$ |

$n=35$. Tukey–Kramer test, $p<0.01$. Value with different superscript letters are significantly different.
Fig. 2  Accuracy of the prediction score by the regression analyses and the measured scores in the individual ratings of Junmai Ginjo. One panelist’s data are shown (n=173 sake samples).

○: training data. ◆: test data.
A: MRA (lm). B: PLS (pls). C: ANN (nnet). D: SVM (svmRadial). E: RF (rf).
3.4 The contribution of each feature to the sensory evaluation

Since the prediction accuracy provided by the RF method was the highest among the machine learning methods, we used the importance values of the RF method for our investigation of the contribution of each feature to the evaluation score. In addition, comparing the prediction accuracy of machine learning with MRA and PLS for test data, there was no decrease in prediction accuracy due to over-learning, so we verified the contribution of each feature to the sensory evaluation using 5-fold cross-validation of all sample data. We based the contribution on the regression coefficients obtained by the MRA and the importance values obtained by the RF method. Table 6 shows the contribution of each feature to the evaluation score.

The coefficient of the MRA represents the magnitude of the contribution of the explanation variable to the objective variable. That is, a variable with a large absolute value of the regression coefficient can be considered a factor with a large influence on the evaluation. Regarding the average scores, isoamyl acetate (|0.39|), OD 260 (|0.34|), and OD 280 (|0.31|) were in descending order of the absolute value of the regression coefficients. For the importance values of the RF method, the larger the value, the more important the factor is for prediction. Ethyl caprylate (100), ethyl caproate (80), and ethyl acetate (63) were in descending order of importance. In other words, the tendency of the contribution of the features differs between the regression coefficient and the importance.

Among volatile components, those having a large absolute value of the regression coefficient and those having a high degree of importance were scattered in order to confirm the relationship with the average score. (Fig. 3.) Components with relatively large regression coefficients, such as Isoamyl acetate and Ethyl caproate, can be read for their tendency from a scatter plot. Samples with higher concentrations of these components have lower average scores (good evaluation). On the other hand, even if the importance is high, such as Ethyl caprylate or Ethyl acetate, the components whose absolute value of the regression coefficient is small are difficult to visually recognize the relationship. That is, it is suggested that

| Regression analysis: | Average score | Individual score |
|---------------------|--------------|-----------------|
| Variable            | Mean         | mean±SD         |
| (Intercept)         | 2.86±0.27    |                 |
| General analyses    |              |                 |
| Acidity             | 0.07±0.14    | 0.00±0.10       |
| Amino acid content  | 0.00±0.10    | 0.06±0.14       |
| Glucose             | 0.10±0.14    | 0.06±0.14       |
| Substrate-related nucleic acid | | |
| OD260               | 0.34±0.24    | 0.07±0.20       |
| OD280               | 0.31±0.20    | 0.01±0.20       |
| Volatile components:|              |                 |
| Acetaldheyde        | 0.03±0.08    | 0.05±0.08       |
| Ethyl acetate       | 0.02±0.15    | 0.03±0.15       |
| nProOH              | 0.13±0.09    | 0.13±0.09       |
| iBuOH               | 0.08±0.13    | 0.15±0.14       |
| Isoamyl acetate     | 0.39±0.24    | 0.15±0.14       |
| pAmiOH              | 0.14±0.17    | 0.15±0.17       |
| Ethyl caproate      | 0.23±0.25    | 0.12±0.25       |
| Ethyl caprylate     | 0.06±0.10    | 0.06±0.10       |
| Caproic acid        | 0.05±0.17    | 0.05±0.17       |
| Simplified analyses |              |                 |
| pH                  | 0.00±0.10    | 0.07±0.10       |
| Electrical Conductivity | 0.16±0.09 | 0.11±0.09       |
| Brix                | 0.09±0.08    | 0.08±0.08       |

Values in parentheses are the top three (in descending rank) orders of absolute values.
the relationship between these components and the evaluation is a nonlinear relationship. The non-linear relationship may be due to the following: (a) Just the right concentration exists, (b) Balance with other ingredients. 

In product development, it is suggested that the target values of these component concentrations need to be determined by regression analysis after narrowing down the target panels and product categories.

Regarding the individual scores, isoamyl acetate (|0.15|), iAmiOH (|0.15|), and ethyl caproate (|0.12|) had the regression coefficients with the largest absolute value. The order of importance values of RF was ethyl caproate (66), isoamyl acetate (56), and ethyl caprylate (53). Because the SD was large, the individual difference in the evaluation tendency was large. However, the same trend occurred in the regression coefficient and importance values for the individual scores. The analyses of the individual scores indicated that volatile components such as ethyl caproate and isoamyl acetate make a large contribution to the evaluation. These volatile components provide valuable aroma characteristics in Japanese sake (like flowers and/or fruits), especially in Junmai Ginjo, that are called ginjoko. For this reason, we suspect that our panelists first classified the sake samples with aroma.

Fig. 3 Relationship between volatile components and average score
as a clue for the evaluation of Junmai Ginjo.

We then plotted the coefficients and the importance values of isoamyl acetate and ethyl caproate, which contributed greatly to the individual evaluation, in order to visualize individual differences (Figs. 4 and 5). From the coefficient graphs it can be seen that the data of both (or either) of these factors gathered at minus areas, and there were many panelists who preferred sake containing ginjoko. The fragrance of ginjoko thus led to high evaluation scores. On the other hand, as can be seen in the graph of the importance values, it is possible to discriminate which volatile component is more important (or not). Since the prediction accuracy of the RF regression analysis was higher than that of the MRA, the importance seems to express the evaluation trend more accurately. The possibility of grouping panels was suggested by the use of importance values.

3.5 Evaluation of trends by combining the regression coefficients and importance values

Importance values cannot be used to indicate whether the contributions of variables are positive or negative. We therefore examined the evaluation tendency by combining the use of the regression coefficients and the use of importance values (Table 4). For the average scores, OD 260 had the highest importance (51) and the absolute value of the regression coefficient is also large (0.34), suggesting that there is a linear relationship with the evaluation score. Ethyl caprylate had the highest importance value (100), but its regression coefficient (−0.09) was around 0. In other words, ethyl caprylate appears to be an important component and there is a non-linear relationship between this concentration and the evaluation score. The same tendency was observed with ethyl acetate (regression coefficient 0.02, importance value 63). These results demonstrated that the relationship between the sensory evaluation result and the component concentration was nonlinear. It is thus apparent that the nonlinear analyses (machine learning) provided higher prediction accuracy compared to the linear analysis (MRA). As we noted in the Introduction, new knowledge that could not be obtained by using only an MRA or the RF method alone can be obtained by considering importance values and regression coefficients in combination.

3.6 Individual differences in the contribution of each variable

We next visualized the individual differences in the contribution of each variable by plotting the relationship between the regression coefficients and the importance values for the individual scores and the average scores. Regarding typical features that contribute to taste and flavor, the distribution of regression coefficients and importance values for the average scores and individual scores are shown in Figure 6. The panelists who indicated that isoamyl acetate is important (importance > 75) tended to provide negative regression coefficients (−0.0 to approx. −0.5). In other words, isoamyl acetate was shown to be a fragrance component that easily leads to a high evaluation.

In contrast, although many of our panelists regarded
ethyl caproate as important (importance > 75), its regression coefficients were dispersed from plus to minus overall (0.4 to -0.6). That is, the individual differences in the evaluation of ethyl caproate tended to be large. The results for ethyl caprylate suggested that the relationship between the concentration and its evaluation is not a linear relationship because the coefficients gather around 0 irrespective of the importance values. A similar tendency was observed for the amino acid content.

As noted earlier, by visualizing the combination of the regression coefficient and the importance value for each variable, it is possible to estimate the panelists' tendencies. Such information can be applied to consumer marketing research, which is one of the purposes of palatability analyses.

Although consumer marketing research often classifies and targets consumers by using demographic data such as age, gender, and lifestyle, consumers' preferences also show diversity, and it is becoming necessary to classify consumers independently of their demographic data. It would be difficult to analyze the evaluation tendency of each individual because of the enor-
mous amount of data and analytical processing that would be involved, but the development of new algorithms and analysis methods will contribute to this endeavor. If it is possible to classify a consumer panel's results according to the preference tendency of each panelist, a more accurate exploration of consumers' expectations could be achieved. In order to obtain a more sophisticated model, it is therefore necessary to select the appropriate variables and analytical methods, set appropriate analysis parameters, perform data preprocessing, and obtain more accurate cross-validations.

Then, as issues of this analysis for improving the prediction accuracy of unknown samples, there are a comprehensive analysis that captures features including defects and a collection of sensory evaluation scores using a ratio scale. Our present study's results demonstrate that by combining the use of an MRA and the RF method, it is possible to determine trends among evaluators conducting a comprehensive evaluation of sake.

Acknowledgments

We thank Mr. Hasegawa and Mr. Takagi of Hasegawa Saketen, Inc. for providing the sake samples and evaluation data.

References

1) J. Shimada; "Tabemono no Oishisa wo Ketteisuru Youin." (in Japanese) Jpn. J. Clin. Nutr. 77, 367 (1990).
2) Belsley, D. A., Kuh, E., Welsch, R. E. Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. New York: Wiley., 1980.
3) D. Cozzolino, G. Cowey, K. A. Lattey, P. Godden, W. U. Cynkar, R. G. Dambergs, L. Janik, M. Gishen; Relationship between wine scores and visible–near-infrared spectra of Australian red wines. Anal. Bioanal. Chem. 391, 975–981 (2008).
4) Y. Sagara; "Evaluation methodology for palatability based on Food Kansei Model." (in Japanese) J. Cookery Sci. Jpn., 56, 317–325 (2009).
5) R. Díaz–Uriarte, S. A. De Andres; Gene selection and classification of microarray data using random forest. BMC Bioinformatics. 7, 3 (2006).
6) F. Sebastiani; Machine learning in automated text categorization. ACM. Comput. Surv. 34, 1–47 (2002).
7) A. Krizhevsky, I. Sutskever, G. E. Hinton; Imagenet classification with deep convolutional neural networks. Adv. Neural. Inform. Process. Syst. 25, 1097–1105 (2012).
8) V. Vapnik; Pattern recognition using generalized portrait method. Automat Rem. Contr., 24, 774–780 (1963).
9) C. Cortes, V. Vapnik; Support–vector networks. Mach. Learn., 20, 273–297 (1995).
10) H. Ni, S. Gunasekaran; Food quality prediction with neural networks. Food. Technol., 52, 60–65 (1998).
11) L. Breiman; Random forests. Mach. Learn., 45, 5–32 (2001).
12) M. Fernández–Delgado, E. Cernadas, S. Barro, D. Amorim; Do we need hundreds of classifiers to solve real world classification problems? J. Mach. Learn. Res. 15, 3133–3181 (2014).
13) P. Cortez, A. Cerdeira, F. Almeida, T. Matos, J. Reis; Modeling wine preferences by data mining from physicochemical properties. Decis. Support. Syst. 47, 547–553 (2009).
14) S. Qiu, J. Wang, L. Gao; Discrimination and characterization of strawberry juice based on electronic nose and tongue: Comparison of different juice processing approaches by LDA, PLSR, RF, and SVM. J. Agr. Food Chem., 62, 6426–6434 (2014).
15) M. Sugimoto, T. Koseki, A. Hirayama, S. Abe, T. Sano, M. Tomita, T. Soga; Correlation between sensory evaluation scores of Japanese sake and metabolome profiles. J. Agr. Food Chem., 58, 374–383 (2009).
16) T. Nishiya (Ed.); Commentary for standard method established by National Tax Agency, 4th edition. Brewing Society of Japan, Tokyo (1993).
17) I. Miwa, J. Okuda, K. Maeda, G. Okuda; Mutarotase effect on colorimetric determination of blood glucose with –D-glucose oxidase. Clin Chem Acta. 37, 538–540 (1972).
18) S. Yamanaka, S. Kawanishi, K. Okui; "Changes of the Nucleic Acid Related Substances in Sake Making Process." (in Japanese) J. Brew. Soc. Japan, 77, 817–821 (1982).
19) K. Yoshizawa; "The rapid assay for aroma compounds in sake by headspace method." (in Japanese) J. Brew. Soc. Jpn., 68, 59–61 (1973).
20) R. M. O’Brien; A caution regarding rules of thumb for variance inflation factors. Qual. Quant., 41, 673–690 (2007).
21) M. Kuhn; Building predictive models in R using the caret package. J. Stat. Softw. 28, 1–26 (2008).
22) M. Kuhn; A short introduction to the caret package. R Found. Stat. Comput. 1–10 (2015).
23) B. Mevik, R. Wehrens; The pls package: Principal component and partial least squares regression in R. J. Stat. Softw., 18, 1–24 (2007).
24) W. N. Venables, B. D. Ripley; "Modern Applied Statistics with S, 4th edition", Springer (2002).
25) A. Karatzoglou, A. Smola, K. Hornik, A. Zeileis; kernlab – An S4 package for kernel methods in R. J. Stat. Softw. 11, 1–20
26) A. Liaw, M. Wiener; Classification and regression by randomForest. R. News, 2, 18-22 (2002).

URLs cited
i) https://www.R-project.org/ (June 3, 2019).
ii) https://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html (June 3, 2019).
iii) https://www.stat.berkeley.edu/~breiman/RandomForests/cc_software.htm (June 3, 2019).
日本酒の総合評価と物理化学的特徴との関係性の解析における機械学習の適用

下藤悟1,2†, 松井元子2, 村元由佳利3, 森山洋憲1, 加藤麗奈1, 甫木嘉朗1, 上東治彦1

1 高知県工業技術センター, 2 京都府立大学大学院, 3 京都府立大学

食品の品質の総合評価を解析するには, 官能評価データと物理化学的なデータを用いるのが一般的である.
従来の解析手法としては, 線形解析である重回帰分析 (MRA) や部分的最小二乗回帰 (PLS) を行ってるものが多い. しかしながら, 食品の味の総合評価は, 食品の成分などの特徴に対して非線形な関係があることは経験的にもよく知られている. 一方, 近年では非線形的な解析を行う手法として, データマイニングの分野において, 機械学習が採用されており, 柔軟性があり, 予測精度が高い解析ができるといわれている.

そこで本研究では, 官能評価による日本酒の品質の総合評価に対する物理化学的特徴の寄与をより明確にすることを目的とし, その関係性の解析に機械学習を適用した. 一般的な統計手法である MRA, PLS と代表的な機械学習手法である人工ニューラルネットワーク (ANN), サポートベクターマシン (SVM) およびランダムフォレスト (RF) で比較を行うことで, より正確な予測モデルを得ることができると考えた. さらに, プラグリスの特性化のために機械学習から得られる変数の重要度を MRA から得られる回帰係数を組み合わせて考察を行った.

試料には日本酒（純米吟醸）173 品を用い, 官能評価は 35 名の熟練されたパネリストによって行い, 品質は 5 段階で評価した.
物理化学的特徴を得るために, 酸度, アミノ酸度, グルコース含量といった一般的な分析, Brix, 導電率, pH のいった簡単な分析を行った.

官能評価スコアへの物理化学的特徴の寄与を, 回帰分析によって検討した. 説明変数に物理化学的特徴の分析値を, 目的変数に個々のパネリストの個々の評価スコアと平均スコアを用いた. 解析には R を用いた i.

回帰分析は, MRA と PLS, 機械学習 (SVM, ANN および RF) により行った. 各解析には caret パッケージを使用し, 解析条件の最適化を行った.
回帰分析の精度の検証を, 過学習を避けるために train データと test データに分割して行った. まず, 全体の 90% に当たる 158 品をトレーニングデータ, 残りの 10% に当たる 15 品を精度検証用データにランダムに分割した. さらに, トレーニングデータを用いて回帰分析を行い, 予測モデルを得た. 得られた予測モデルから, テストデータ（予測モデルの作成に使用していないデータ）の総合評価の予測値を計算し, 実測値と比較し, 各分析手法の精度を調べた. さらに, トレーニングデータについても同様に予測精度を比較することで, 予測モデルのフィッティングについて調べた. 予測精度は, 許容範囲内の誤差に含まれる試料の割合, 平均絶対誤差 (MAE), 二乗平均平方根誤差 (RMSE) で評価した. これからの解析方法の結果から, MRA よりも機械学習（とくに RF）の方が回帰回帰時のフィッティングがよく, 日本酒の品質の総合評価を高い精度で解析できる可能性が示唆された. また, MRA で得られた回帰係数と RF で得られた重要度から, 評価スコアに対する各物理化学的特徴の寄与についても検討した. MRA で得られた回帰係数は, 符号により評価への影響の良し悪しが判別できる. また, 評価スコアの寄与も大きいほど評価への寄与も大きいと考えられる.
一方, RF で得られた重要度は, 0〜100 の値のため, 評価への影響の良し悪しは判別できないが, 値の大きいものほど予測精度に大きく影響することを示す指標である.
個々のパネリストのスコアの解析から, 日本酒の品質評価にカプロン酸エチルと酢酸イソアミルといった香気成分大きく寄与していることが示唆された. さらに回帰係数と重要度の値を組み合わせて評価傾向を確認したところ, 総合評価と成分濃度には非線形関係のものがあることが示唆された.
以上の結果から, 日本酒の品質の総合評価における傾向について, MRA と RF を組み合わせることでより明確に捉えることができた.