

酒造原料米の消化性Brixの機械学習による予測に関する研究

齋藤健太* 樋口誠一* 和田健太郎* 横堀正敏*

Study on Machine Learning Prediction of Enzyme Digestibility of Rice for Sake Making

SAITO Kenta*, HIGUCHI Seiichi*, WADA Kentaro*, YOKOBORI Masatoshi*

抄録

酒造原料米の溶解性は日本酒の品質に影響を及ぼす。その指標となる消化性Brixを迅速に予測するモデルの構築を目指した。精米歩合35~80%かつ水分7.2~14.2%と対象範囲を拡大し、説明変数として吸水率及びメッシュ農業気象データを用いた機械学習による回帰分析を行い、最適な解析手法を検討した。さらに予測精度の高い解析手法における説明変数の重要度を明らかにした。本法を活用することで、仕込みに使用する酒造原料米の溶解性が迅速に把握でき、最適な原料処理が可能となると期待される。

キーワード：米，日本酒，溶解性，消化性，機械学習

1 はじめに

酒造原料米（以下、原料米という）の溶解性は、生育時の気象条件¹⁾、品種、精米歩合²⁾及び水分²⁾によって変化し、日本酒の醸造工程や品質に影響を及ぼす。この溶解性の指標は酒造用原料米全国統一分析法²⁾による消化性 Brix で表される。

白米の浸漬あるいは蒸米の晒しなどの原料処理の調整により、原料米の溶解性を制御できる³⁾が、その条件を決めるには仕込み前に溶解性を把握する必要がある。しかし、原料米の溶解性分析には時間と手間がかかる⁴⁾ため、多くの酒造現場では行われていない。

酒米研究会では、全国の原料米を調査している。品種特性と年毎の傾向を把握するため、精米歩合70%かつ水分13.5%と限定された条件の原料米の消化性Brixの分析結果を酒造期後に提供している。仕込みに使用する原料米がそれ以外の精米歩合や水分である場合は直接評価できないため、溶解性は把握されないまま仕込まれ、造り手の経験則で

管理されており、十分な対応ができていない。

既報³⁾では部分的最小二乗（以下、PLSという）回帰分析による消化性Brixの予測を行い、迅速に把握できるようになったが、精米歩合70%かつ水分13.5%と限定された条件であった。そこで本研究では、精米歩合と水分の対象範囲を拡大した消化性Brix予測モデルの構築を目的とした。説明変数は迅速分析が可能な白米水分、10分吸水率、20分吸水率、近赤外分析（以下、NIRという）及びメッシュ農業気象データ⁴⁾とした。PLS、L1正則化線形（以下、LASSOという）、線形サポートベクターマシン（以下、SVMという）、人工ニューラルネットワーク（以下、ANNという）、ランダムフォレスト（以下、RFという）及び勾配ブースティング（以下、GBという）といった機械学習による予測モデルを構築し、予測精度を比較し、最適な解析手法を明らかにする。さらに、最も予測精度の高かった解析手法における説明変数の重要度を明らかにする。

* 食品・バイオ技術担当

表1 試料の品種		表2 試料の精米歩合		表3 試料の白米水分	
品種	調査点数	精米歩合	調査点数	白米水分	調査点数
さけ武蔵	104	30-39%	8	7.0-9.9%	29
山田錦	42	40-49%	11	10.0-11.9%	32
五百万石	12	50-59%	61	12.0-12.9%	102
その他の酒米	53	60-69%	14	13.0%以上	62
一般米	14	70-79%	128	全合計	225
全合計	225	80%	3		
		全合計	225		

2 実験方法

2.1 試料

品種、生産地、生産年、精米歩合、白米水分の異なる原料米 225 点を用いた（表 1～表 3 参照）。生産年は 2017～2023 年であった。

2.2 原料米分析

20 分吸水率及び消化性 Brix の測定は酒造用原料米全国統一分析法²⁾に準じた。10 分吸水率は、20 分吸水率と同様の方法で、時間を 10 分間とした。白米水分は酒米・麴水分計（ライスタ f7、ケツト科学研究所）の酒米精米モードで測定した。

2.3 NIR

近赤外スペクトルの測定には、フーリエ変換赤外分光分析装置（Frontier MIR/NIR plus NIRA、PerkinElmer Japan）を用いた。近赤外反射アクセサリ（NIRA）の窓板を白米（米粒のまま）で覆い、セラミックディスクをリファレンスとして反射モードで測定した。637～2700 nm の範囲を 0.3～0.4 nm 間隔で、積算回数は 32 回とした。3 回繰り返し測定し、反射率の平均値を求めた。

データ処理は、最小二乗法に基づいた平滑化と数値微分を行い、915, 1200, 1355, 1460, 1580, 1708, 1776, 1825, 1936, 2058, 2106, 2174, 2282, 2320, 2430 nm の各波長スペクトルにおける 2 次微分ピーク面積値を求めた。

2.4 メッシュ農業気象データ

国立研究開発法人農業・食品産業技術総合研究機構のメッシュ農業気象データシステム⁵⁾の日最高気温、日平均気温、日最低気温、日積算降水量、日照時間、全天日射量より、出穂後 10 日間ごとに 40 日間までの平均値を求めた。また平均気温は出穂後 20、30、40 日間の平均気温を求めた。圃場位置が特定できない米については、県単位のデータとした。

2.5 回帰分析

消化性 Brix への単回帰分析での相関係数及び決定係数（以下、 R^2 という）を比較した。

機械学習では R (version 4.3.1)⁶⁾のパッケージ tidymodels⁷⁾を用いた。説明変数には白米水分、10 分吸水率、20 分吸水率、気象データ（27 変数）、NIR データ（95 変数）の全 125 変数を用いた。変数増減法により絞り込み、LASSO、SVM、ANN、RF、GB による回帰分析を行った。動作エンジンには、LASSO で glmnet パッケージ、SVM で kernlab パッケージ、ANN で nnet パッケージ、RF で ranger パッケージ、GB で XGboost を指定した。PLS 回帰分析では、pls パッケージを用い、潜在変数は 3 とした。試料はランダムに区分し、全体の 80% に当たる 179 点を学習データ、残りの 20% に当たる 46 点を検証データとした。10 分割交差検証法を用いた。目的変数となる消化性 Brix には対数変換を行い、全ての変数を平均 0、分散 1 となるように標準化した。最適なハイパーパラメータの選択

には、二乗平均平方根誤差（以下、RMSE という）を指標とした。ハイパーパラメータには、LASSO では penalty 及び mixture を、SVM では cost 及び margin を、ANN では hidden_units、penalty 及び epochs を、RF 及び GB では mtry 及び trees を用いた。

2.6 回帰分析の予測精度

予測精度は、許容誤差 0.5 以内に含まれる試料の割合（以下、正解率という）、平均絶対誤差（以下、MAE という）、RMSE、 R^2 で評価した。酒造用原料米全国統一分析法に定められるばらつき許容範囲は 5% であることから、許容誤差を 0.5 とした。

2.7 各説明変数の重要度

各説明変数の重要度は、相関係数及び 2 乗した Shapley 値⁸⁾により評価した。2 乗した Shapley 値には、Variance Shapley（以下、VS という）、Baseline Shapley（以下、BS という）、All Baseline Shapley（以下、ABS という）、Cohort Shapley（以下、CS という）を用いた。CS の算出の際に各説明変数を 10 区分に分類した。

3 結果及び考察

3.1 回帰分析の予測精度

各種分析値の平均値、最大値、最小値の一部を表 4 に示す。説明変数は変数増減法により 125 変数のうち 17 変数に絞り込んだ。回帰分析に用いた 17 変数は、20 分吸水率、1~10 日平均気温、11~20 日平均気温、21~30 日平均気温、31~40 日

平均気温、1~10 日最低気温、11~20 日最低気温、21~30 日最低気温、31~40 日最低気温、1~10 日日照時間、11~20 日日照時間、21~30 日日照時間、31~40 日日照時間、1~10 日全天日射量、11~20 日全天日射量、21~30 日全天日射量、31~40 日全天日射量であった。タンパク質に影響を及ぼす気温及び日射量を説明変数に加えることで、既報³⁾の PLS 回帰分析の説明変数に用いていた NIR が不要となったと考えられる。

6 つの機械学習による回帰分析の予測精度を表 5 に示す。正解率は GB で最も高く、MAE 及び RMSE は GB で最も低く、 R^2 は GB で最も高かった。GB による消化性 Brix 予測値及び実測値を図 1 に示す。GB による予測値は 8.4~11.6 と幅広く分布していた。他の機械学習モデルによる予測範囲は狭く誤差も大きかったため、広範囲の予測が可能で正解率が高かった GB による予測が適していると考えられる。また PLS の R^2 が既報³⁾より低下したことは、精米歩合及び白米水分の対象範囲を拡大したためと考えられる。

表 4 各種分析値の平均値、最大値、最小値

変数	平均値±SD	最大値	最小値
消化性Brix (%)	10.2±0.9	12.1	7.7
精米歩合 (%)	62.1±10.8	80.0	35.0
白米水分 (%)	12.0±1.5	14.2	7.2
10分吸水率 (%)	21.7±4.5	34.2	12.1
20分吸水率 (%)	31.2±4.3	49.8	20.6

n=225

表5 機械学習による回帰分析の予測精度

回帰分析	PLS	LASSO	SVM	ANN	RF	GB
正解率	54.3	54.3	58.7	67.4	69.6	78.3
MAE	0.53	0.50	0.52	0.48	0.38	0.35
RMSE	0.64	0.62	0.65	0.80	0.46	0.43
R^2	0.53	0.57	0.51	0.48	0.76	0.80

n=46

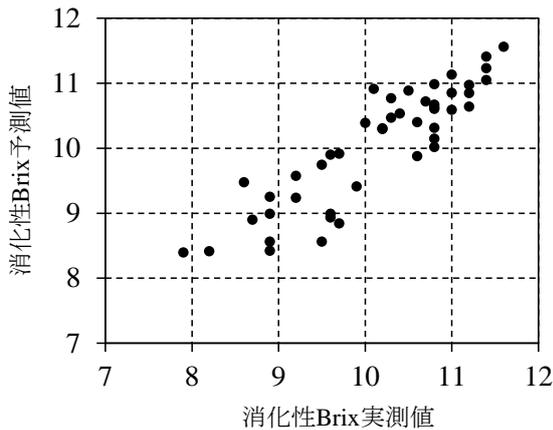


図1 GBによる消化性 Brix 予測値及び実測値

3.2 消化性 Brix へ影響を与える要因

消化性 Brix に対する各説明変数の単回帰分析による相関係数及び R^2 の一部を表6に示す。タンパク質由来と考えられる NIR の 2174nm の 2 次微分ピーク面積値が正の相関を示したことは、既報⁹⁾と一致した。また日射量はタンパク質組成に影響を及ぼすことが報告され¹⁰⁾、出穂後 11~20 日日照時間及び全天日射量の R^2 が高かったことから、消化性 Brix へタンパク質が影響を及ぼしていると考えられる。

最も予測精度の高かった GB における各説明変数の重要度を表す Shapley 値を図2に示す。基準と比較した重要度を表す BS 及び ABS では、出穂後 11~20 日平均気温の重要度が高く、これは先行研究¹¹⁾と一致した。区分別の影響の強さを表す CS では 17 変数すべてが関わっており、特に出穂後 11~20 日日照時間及び全天日射量の重要度が

高かった。重要度の差の原因を表す VS では出穂後 31~40 日の平均気温及び最低気温の重要度が高かった。合成データを生成しない CS を用いることで、平均気温だけでなく、日照時間、日射量及び吸水率が消化性 Brix へ与える影響が明らかになった。

表6 各説明変数の相関係数及び R^2

変数	相関係数	R^2
白米水分 (%)	-0.488	0.238
10分吸水率 (%)	0.133	0.018
20分吸水率 (%)	0.332	0.110
1~10日平均気温 (°C)	-0.354	0.125
11~20日平均気温 (°C)	-0.434	0.188
21~30日平均気温 (°C)	-0.455	0.207
31~40日平均気温 (°C)	-0.486	0.236
1~10日最低気温 (°C)	-0.324	0.105
11~20日最低気温 (°C)	-0.375	0.140
21~30日最低気温 (°C)	-0.429	0.184
31~40日最低気温 (°C)	-0.477	0.228
1~10日日照時間 (h/day)	-0.456	0.208
11~20日日照時間 (h/day)	-0.529	0.280
21~30日日照時間 (h/day)	-0.377	0.142
31~40日日照時間 (h/day)	-0.348	0.121
1~10日全天日射量 (MJ/m ² /day)	-0.471	0.222
11~20日全天日射量 (MJ/m ² /day)	-0.539	0.290
21~30日全天日射量 (MJ/m ² /day)	-0.434	0.188
31~40日全天日射量 (MJ/m ² /day)	-0.425	0.181
2174nmの2次微分ピーク面積	0.529	0.280

n=225

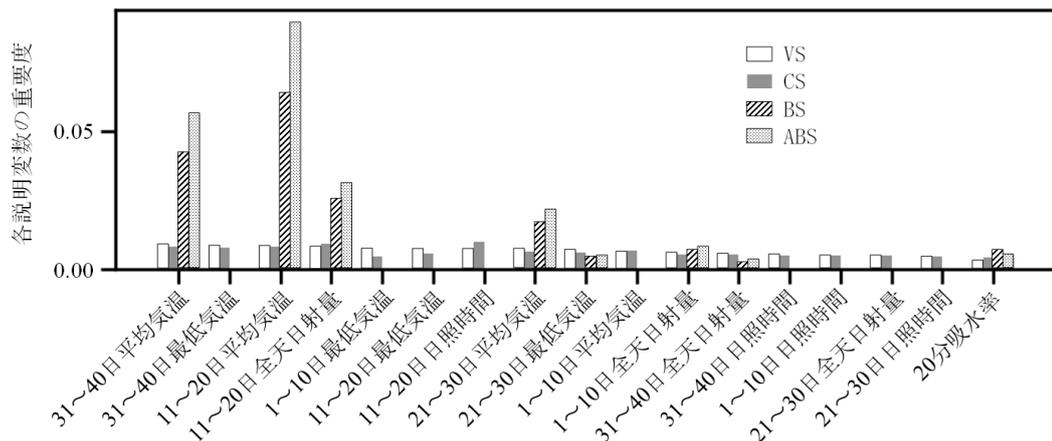


図2 GBにおける各説明変数の重要度を表す Shapley 値

4 まとめ

迅速に得ることができる 20 分吸水率及びメッシュ農業気象データを説明変数として、原料米の対象範囲を精米歩合 35~80%かつ水分 7.2~14.2%と拡大した機械学習による消化性 Brix 予測モデルを構築した。PLS、LASSO、SVM、ANN、RF 及び GB により消化性 Brix を予測したところ、GB の正解率 78.3%と最も高かった。

さらに、Shapley 値を用いて GB における各説明変数の重要度が明らかになった。

今後は、実際に現場で使用される原料米の消化性 Brix 予測値を情報提供し、求められる予測精度の確認を行いたい。実際に現場で使用される白米水分 7~12%の試料のデータを蓄積することで、予測精度の向上が期待される。

謝 辞

メッシュ農業気象データシステムを御提供いただきました国立研究開発法人農業・食品産業技術総合研究機構、原料米を御提供いただきました酒造会社及び農家の皆様に感謝します。

参考文献

- 1) 奥田将生, “米デンプン構造と醸造特性・気象条件との関係”, 生物工学会誌, vol. 90, no. 5, pp. 227-230 (2012).
- 2) 酒米研究会編, “酒造用原料米全国統一分析法”, 酒米研究会, (2023).
- 3) 齋藤健太, 樋口誠一, 和田健太郎, 横堀正敏, “酒造原料米の溶解性予測システムの確立に関する研究—酒造原料米の消化性 Brix 予測値の活用—”, 埼玉県産業技術総合センター研究報告, vol. 18, pp. 35-39 (2020).
- 4) 菅原雅通, 鍋倉義仁, 金桶光起, “米溶解性簡易推定法の比較”, 日本醸造協会誌, vol. 118, no. 7, pp. 529-534 (2023).
- 5) 国立研究開発法人農業・食品産業技術総合研究機構, “メッシュ農業気象データシステム”, https://amu.rd.naro.go.jp/wiki_open/doku.php?id=s tart, 2025.10.20

- 6) “The R Project for Statistical Computing”, <https://www.r-project.org>, 2025.1.22
- 7) “tidymodels”, <https://www.tidymodels.org/>, 2025.11.8
- 8) Masayoshi Mase, Art B.Owen, Benjamin Seiler, “Explaining black box decisions by Shapley cohort refinement”, arXiv:1911.00467, vol. 2, pp. 1-26 (2020).
- 9) 齋藤健太, 横堀正敏, 樋口誠一, 和田健太郎, “酒造原料米の溶解性予測システムの確立に関する研究—酒造原料米の吸水性及び消化性の重回帰分析による予測—”, 埼玉県産業技術総合センター研究報告, vol. 17, pp. 30-34 (2019).
- 10) 大平陽一, “水稻種子貯蔵タンパク質の組成、含有率に及ぼす栽培・環境条件の影響ならびに米粒内分布特性に関する研究”, 近畿中国四国農業研究センター研究報告, vol. 14, pp. 1-49 (2015).
- 11) 池上勝, 磯野幸治, 藤本啓之, 加藤雅宣, 杉本琢真, 高橋圭, 奥田将生, “兵庫県における酒米品種の消化性と登熟期間の気温との関係”, 日本作物学会紀事”, vol. 90, no. 4, pp. 451-456 (2021).