

植物収穫時品質に関与する経時特徴量の検討

中西豪太^{†1} 水野涼介^{†2} 今原淳吾^{†3} 前島慎一郎^{†3} 峰野博史^{†4}

概要: 日本の高度な栽培技術の喪失危機を解決し、熟練農業従事者の栽培技術を新規農業従事者への継承可能とするために、栽培技術を形式知化する研究が行われている。特に高品質な農産物の栽培には長年の経験と勘に基づく栽培技術が必要であり、新規農業従事者には敷居が高く、継承に時間を要する。植物の栽培は生育状況と1日の時間帯を長期的に考慮して行う必要があるが、これらを考慮した形式知化の研究は少ない。そこで本研究では、膨大な特徴量の中から関連性の高い経時特徴量を選択するために、特徴量選択手法として用いられる正則化回帰モデルを適用し、生育状況と時間帯を重畳させた経時特徴量を用いることで、植物収穫時品質に関与する特徴量を選択する手法を提案する。選択された特徴量を用いて植物収穫時品質の推定を行ったところ、特徴量選択を行う前より MAE と RMSE が小さい値であることを確認した。そのため、正則化回帰モデルによる特徴量選択手法を植物栽培環境に適用することで、生育状況と時間帯を考慮した経時特徴量の分析が可能であると考えられる。

Analysis of temporal environmental conditions for plant growth related to harvesting quality

GOTA NAKANISHI^{†1} RYOSUKE MIZUNO^{†2} JUNGO IMAHARA^{†3}
SHINICHIRO MAEJIMA^{†3} HIROSHI MINENO^{†4}

1. はじめに

近年、農業従事者の高齢化が進み、世界的に高水準な日本の農業技術が新規農業従事者に継承されずに失われてしまうことが懸念されている。特に、ブランド価値のある高糖度なトマトなど高品質な農産物を栽培する技術は熟練農業従事者の長年の経験と勘によって培われた栽培技術であり、継承には長い年月を必要とする。そのため、高品質な農作物を栽培する技術は新規農業従事者に継承されずに喪失してしまうという課題がある。

この課題を解決するために、熟練農業従事者の栽培技術を形式知化する研究として、経験則と勘に基づいていた判断していた環境条件と植物の因果関係を明らかにする研究が行われている。例えば、環境条件が与える影響の分析として、積算温度と収穫時品質との分析[1]、気象情報が植物の品質に与える影響[2]、光強度条件など栽培環境条件の違い[3][4][5]が収穫時の品質に与える影響の分析が行われている。しかし、栽培環境と収穫時品質の分析において、植物の生育過程などの生育状況や1日の時間帯の考慮が行われていることは少なく、栽培技術の形式知化を十分に実現できているとは言えない。そのため、新規農業従事者が熟練農業従事者の栽培技術を継承することは依然として難しくなっている。

本研究では、収穫時の品質に関与する特徴量を検討する

ため、生育状況と時間帯を重畳した経時特徴量データを抽出し、正則化回帰モデルを適用することで、収穫時品質との関連性の高い特徴量選択手法を提案する。選択された特徴量を用いて収穫時品質を推定し、推定時の誤差を評価することで、植物栽培環境において生育状況と時間帯を考慮した分析が可能であることを示す。

本論文は全5章から構成される。第2章で関連研究について述べ、第3章で正則化回帰モデルを用いた収穫時品質に関与する特徴量選択手法について説明する。第4章で提案手法に関する評価について述べ、第5章で結論と今後の課題についてまとめる。

2. 関連研究

2.1 逐次変数選択法 (Stepwise 法) [6]

膨大な変数群から重要な変数を選択する代表的な手法として、逐次変数選択法がある。逐次変数選択手法とは、変数増加法 (Forward stepwise selection) と変数減少法 (Backward stepwise selection) を組み合わせた変数選択手法である。変数増加法は最初全ての変数に対して、モデルに加えた場合の p 値や Wald の χ^2 値などのモデルの評価指標となる統計量を算出する。p 値を変数選択の評価指標と仮定すると、事前に決定された閾値を満たす指標のうち最小の値を持つ変数をモデルに加えて、モデルを構築する。これらの処理を繰り返してモデルに含まれる変数を増加させていく。閾値を満たす指標が算出されなくなった時点で変数の追加を終了する。このときの変数の組み合わせを最良の組み合わせとする。一方、変数減少法は、変数増加法とは逆に、最初に全ての変数をモデルに取り込んだモデルを作成する。その後、モデルに含まれる変数のうち、決め

†1 静岡大学情報学部

Faculty of Informatics, Shizuoka University

†2 静岡大学大学院総合科学技術研究科

Graduate School of Integrated Science and Technology, Shizuoka University

†3 静岡県農林技術研究所

Shizuoka Prefectural Research Institute of Agriculture and Forestry

†4 静岡大学学術院情報学領域/JST さきがけ

College of Informatics, Shizuoka University / JST PRESTO

られた閾値を満たさない指標を持つ変数をモデルから除去していき、閾値を満たさない指標が算出されなくなった時点で変数の除去を終了する。逐次変数選択法は、最初に変数増加法と同様に変数の追加を行う。変数の追加のたびに、モデルに既に含まれている変数の中で変数減少法と同様に除去が可能な変数があれば除去する。その後、追加及び除去する変数がなくなるまで変数の選択を行い、最終的に残った変数を最良の変数の組み合わせとする。しかし、逐次選択法は、より少ない変数でより効率的に予測することが目的であり、目的変数との因果関係を考慮した変数選択を行わないため、選択された変数と目的変数の因果関係の分析には向いていない。また、変数の取捨選択を繰り返すため、重要な変数の最適な組み合わせを得ることは困難である。

2.2 PLS[7][8] を用いた変数選択

PLS は、データから潜在変数を算出し、その潜在変数への回帰を行う回帰手法である。潜在変数は、潜在変数と目的変数の共分散が最大かつ、潜在変数同士が相関関係を持たないように求める。PLS 回帰アルゴリズムは、複数提案されており、最も代表的な手法が、NIPALS(Nonlinear Iterative Partial Least Squares)[7]アルゴリズムである。NIPALS アルゴリズムは、十分な精度が得られるまで、回帰残差を用いて回帰係数と PLS 回帰式の更新を行い、回帰係数の算出を繰り返す手法である。このときの回帰係数の絶対値が大きい変数のみを選択する PLS- β 法、VIP (Variable Importance in Projection) の値が大きい変数のみを選択する PLS-VIP 法が PLS 回帰を用いた変数選択手法として代表的である。しかし、PLS- β 法、PLS-VIP 法のいずれにおいても、人間の決めた閾値に基づいて変数の選別を行うため、最適な変数選択を行うためには、閾値の決定が困難であると考ええる。

2.3 Random Forest[9]の変数の重要度に基づく変数選択

Random Forest は異なる決定木を多数作成し、その結果の平均を求めることで、決定木の欠点である過剰適合を抑制するアンサンブル学習手法である。Random Forest は推定時に使用した特徴量の重要度を算出可能であり、その値に基づいて変数を選択可能である。Random Forest の変数の重要度は各決定木における変数の重要度の平均値によって算出され、その算出方法は、以下の通りである。1.無作為にデータを選択して決定木を作成する。2.作成した決定木で使用されている1つの変数に関して、データの並び順をランダムに変更する。3.並び順の変更前後で、決定木の精度を比較し、大幅な精度の変化が観測された場合、重要な変数とする。4.多数の決定木にて上記の1から3の手法を用いてその結果の平均値を取り、Random Forest の変数の重要度とする。Random Forest の各変数の重要度は変数全体から見た相対的な値であるため、変数の重要度を0にするスパースな推定を行わない。したがって、変数選択を行うには、閾

値など人手を介して重要な変数を選別する必要があると考える。

2.4 正則化項を用いた変数選択[10][11][12]

遺伝子解析など、変数の次元数がサンプルサイズに比べて遥かに大きいという課題に対処できる統計手法として正則化項によるスパース推定を用いた変数選択手法がある。正則化項を用いたスパースな推定を行う代表的な手法として、Lasso (Least Absolute Shrinkage and Selection Operator[13]) 回帰や Elastic Net[14]回帰などがある。Lasso 回帰モデルは、互いに相関の高い変数群が含まれている場合、推定時に相関の高い変数の中の1つだけが選択され、他の変数は回帰係数が0と推定され、選択されないという欠点を持つため、相関の高い変数群の中で選択される変数は、推定を行うごとに変化する。そのため、Lasso 回帰による変数選択は不安定とされている。この Lasso の問題点を解決するため、考案された回帰手法が Elastic Net 回帰である。Elastic Net 回帰は、Ridge[15]回帰と Lasso 回帰を混合した回帰手法である。Ridge 回帰は相関の高い変数群を考慮することができるため、Lasso 回帰の正則化項と Ridge 回帰の正則化項の強さのバランスをとることで、Lasso 回帰の問題点を解決する。また、Lasso 回帰と Elastic Net 回帰においては、自動で変数の選別を行うため、閾値の決定など人間の介入を行わずに変数選択を実現できる。

2.5 関連研究のまとめ

以上から、変数選択においては、p 値などの統計的な指標を算出した変数に対して、事前に設定した閾値を用いて選別することで変数選択を行う手法が多い。しかし、重要な変数を余すことなく、選別することが可能な閾値を決定することは難しい。そのため、閾値による変数選別手法ではなく、Lasso 回帰や Elastic Net 回帰など自動で重要な変数を選別する手法を用いるべきであると考ええる。

3. 提案手法

3.1 概要

本研究では、生育状況と時間帯情報を考慮した収穫時品質に関与する特徴量を選択する手法を実現するため、正則化回帰モデルによる特徴量選択を植物栽培環境に適用する手法を提案する。正則化回帰モデルのうち、特に Lasso 回帰は L_1 正則化項を、Elastic Net 回帰は L_1 正則化項と L_2 正則化項の両方を回帰モデルの損失関数に加えて最小化することでパラメータ推定と特徴量選択を同時に行うため、正規化回帰モデルを用いて選択された特徴量は、収穫時品質関与する特徴量の選択を実現すると考える。提案手法の流れを図1に示す。提案手法は大別して、特徴量抽出部分と特徴量選択部分から構成される。特徴量選択部分では、経時特徴量データを生育状況と時間帯ごとに分割した後、経時特徴量データのセンサ値ごとの単位の違いを正規化によ

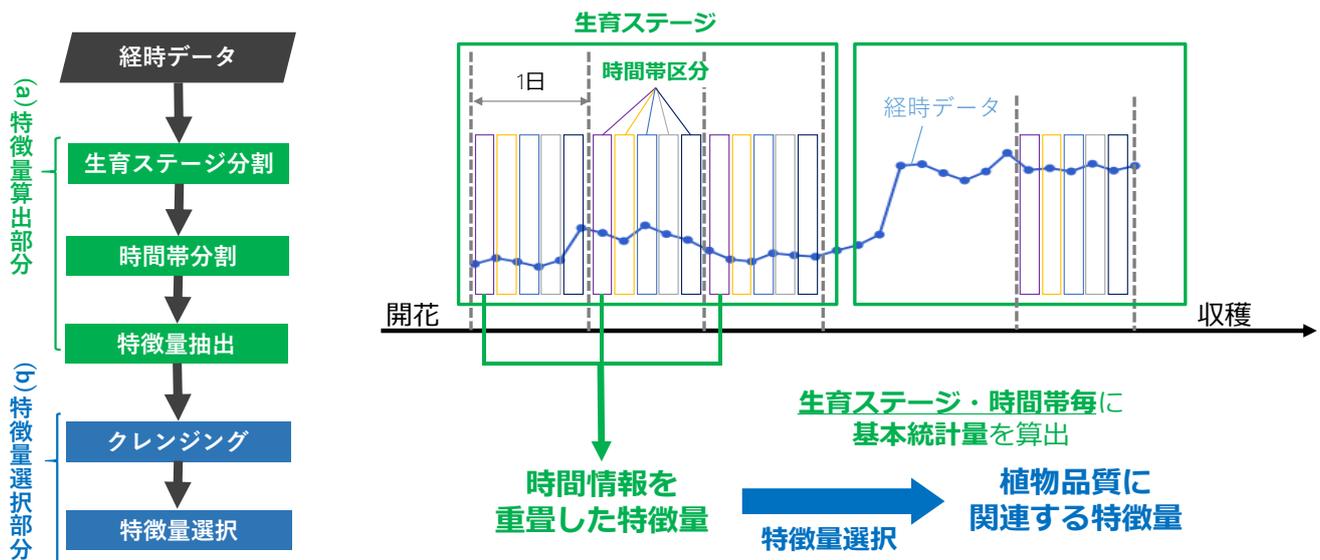


図 1 提案手法の概要

り揃える。その後、分割した各経時特微量データの基本統計量を特微量として抽出することで生育状況と時間帯情報を重畳した特微量の算出を行う。

次に、特微量選択部分では、特微量抽出部分で抽出した特微量と収穫時品質との相関分析を基に特微量をヒューリスティックに精査した後、Lasso 回帰, Elastic Net 回帰を用いて、収穫時の品質と関連が深い特微量を選別する。

3.2 特微量抽出

特微量抽出部分では、最初に経時特微量データと生育状況、時間帯情報を重畳するために、経時特微量データを生育状況と時間帯情報に分割する。生育状況と時間帯情報の分割はイベントに基づいて行う。生育状況は、植物の開花後から収穫までの果実の生育ステージを区切りとして分割を行う。例としてトマトの生育ステージを挙げると、肥大期、緑熟期、白熟期、催色期のように果実の生育に合わせた分割を行うことで、開花からの日数による分割に比べ、果実の個体差を考慮した生育状況の分割が可能であると考えられる。また、時間帯情報については、日の出、日の入りは光量子量との関係性が高い重要なイベントであるため、これらのイベントは時間帯分割の指標に含める。一般に植物は光合成が生長に重要であるため、日の出、日の入りを時間帯分割の指標に含めることは、光合成との関係の分析時に非常に役に立つと考える。次に、各センサーデータに対して生育状況ごとに z-score normalization による正規化を行った。z-score normalization の式を(1)式に示す。

$$z_i = \frac{X_i - \mu}{\sigma} \quad (1)$$

X_i は時点 i におけるデータを示し、 μ 、 σ はそれぞれデータの平均値、データの標準偏差を示す。正規化によってセンサーごとの単位の違いをなくし、経時特微量選択時に特定のセンサーの単位の違いによる影響を抑える。最後に、分割され

た特微量データそれぞれに対し、基本統計量を算出することで、生育状況、時間帯情報を重畳した経時特微量の抽出を行う。基本統計量に関しては、一般に熟練農業従事者は特定の一日の変化を考慮しているのではなく、長期的な変化を考慮した栽培技術を用いているため、大域的な変化を表現できる積算値、平均値、最大値、最小値を基本統計量として使用する。

3.3 特微量選択

特微量選択部分では、クレンジングによる事前の特微量の精査と、クレンジングした特微量を用いた Lasso 回帰と Elastic Net 回帰による特微量選択から構成される。クレンジングには相関分析結果を用いた。抽出された経時特微量と目的変数の相関係数を求めて事前に特微量を精査することで、抽出された全ての特微量を用いて回帰モデルを構築するよりも、特微量選択時の誤った解釈を抑制できると考える。その後、Lasso 回帰, Elastic Net 回帰を用いて、クレンジング後の経時特微量から収穫時の糖度に関連する特微量を選別する。Lasso 回帰と Elastic Net 回帰を式 2 に示す。

$$y = \sum_{j=1}^p \beta_j x_j + \varepsilon \quad (2)$$

ここで x_j , y , β_j , ε はそれぞれ説明変数、目的変数、パラメータ、バイアス項を示す。パラメータ β_j は、線形回帰モデルの損失関数に過学習を防ぐ正則化項を加えた下記の損失関数を最小化することで算出される。Lasso 回帰の損失関数を式 3 に、Elastic Net 回帰の損失関数を式 4 に示す。

$$\begin{aligned} \hat{\beta}_{Lasso} = \operatorname{argmin} & \|y - \sum_{j=1}^p \beta_j x_j\|_2^2 \\ & + \lambda \sum_{j=1}^p |\beta_j| \end{aligned} \quad (3)$$

$$\hat{\beta}_{ElasticNet} = \underset{\beta}{\operatorname{argmin}} \|y - \sum_{j=1}^p \beta_j x_j\|_2^2 + \lambda \sum_{j=1}^p \{\alpha |\beta_j| + (1 - \alpha) \beta_j^2\} \quad (4)$$

λ は正規化項の影響の強さを表すハイパーパラメータであり、 α は L_1 正規化項と L_2 正規化項の相対的な強さを調整するハイパーパラメータである。正規化項のうち特に、 L_1 正規化項は目的変数と関係の強い説明変数のパラメータを非ゼロとし、目的変数と関係が弱い説明変数のパラメータをゼロにするため、説明変数に経時特徴量、目的変数に収穫時品質を用いて Lasso 回帰や Elastic Net 回帰を用いることで、収穫時品質に関与する経時特徴量を選別できると考える。なお、ハイパーパラメータ λ と α の値は、交差検証によって決定した。

3.4 収穫時品質の推定

特徴量選択の効果の検証のため、収穫時品質の推定を行う。Lasso 回帰と Elastic Net 回帰それぞれにおいて選択された特徴量を用いて回帰モデルを構築し、各推定精度を比較した。使用した回帰手法としては SVR (Support Vector Regression) [16] と Random Forest を用いる。評価指標としては、平均絶対誤差 (MAE: Mean Absolute Error (式 5))、平均二乗誤差平方根 (RMSE: Root Mean Squared Error (式 6)) を用いた。

$$MAE = \frac{\sum_{i=0}^N |y_i - \hat{y}_i|}{N} \quad (5)$$

$$RMSE = \sqrt{\frac{\sum_{i=0}^N (y_i - \hat{y}_i)^2}{N}} \quad (6)$$

式 5、式 6 において N はデータ数を示し、 y_i と \hat{y}_i は時点 i における予測値、真値を示す。Lasso 回帰または Elastic Net 回帰によって選択された経時特徴量を用いた推定時の誤差と選択前の経時特徴量を用いた推定時の誤差を比較することで、収穫時品質に関与する経時特徴量の選択の有意性の評価を行う。

4. 評価

4.1 実験方法

植物収穫時品質に関与する経時特徴量の分析の評価として、提案手法を用いて算出された経時特徴量と収穫時糖度を用いて、2 種類の評価実験を行った。まず、特徴量抽出部分の有効性を検証するために、特徴量抽出部分で抽出した 319 次元の特徴量に対して、収穫時の糖度との相関分析を行った。次に、特徴量選択部分の効果を検証するために、特徴量選択部分で抽出された特徴量をクレンジング後の経時特徴量を用いて、Lasso 回帰と Elastic Net 回帰モデルによる特徴量選択を行った。最後に、Lasso 回帰と Elastic Net 回帰それぞれによって選択された経時特徴量と選択前

の特徴量を用いて、SVR と Random Forest による回帰モデルを構築し、収穫時糖度を推定することで、特徴量選択の効果の評価した。

実験は、静岡県農林技術研究所内の 3 ヶ所の試験区にて栽培したトマトを対象として、2016 年 12 月 19 日から 2017 年 5 月 9 日にかけて行った。栽培概要としては、次の通りである。品種: '桃太郎ヨーク' を 2016 年 11 月 21 日播種、ヤシガラ培地を詰めた 4 号 (黒ポリ) 鉢に 2016 年 12 月 19 日に定植、栽培密度 3300 株/10a、4 段階摘心の普通トマト栽培。EC: 大塚 A 処方 0.8~1.5mS/cm。給液制御システム: 処理区ごとに体重計センサを用いた重量を指標とした給液制御とし、設定値は処理区間で同じにした。各処理区では、温度と湿度、散乱光を無線散乱光センサセンサノード [17]、CO₂ 濃度は CO₂ センサ (NMA-VRC-II) を用いて計測し、2016 年 12 月 19 日から 2017 年 5 月 9 日の間、1 分間隔で収集したデータを用いて実施した。なお、以後 temp, humi, photon, co2 はそれぞれ、温度、湿度、光量子量、CO₂ 濃度を指す。

生育ステージは、1. 第一花開花期区: 開花から開花 15 日 2. 果実肥大期区: 開花後 15 日から開花後 29 日 3. 緑熟期区: 開花後 29 日から開花後 37 日 4. 白熟期区: 開花後 37 日から開花後 44 日 5. 催色期区: 開花後 44 日から収穫という各熟期の日数 [18][19] を目安 (図 2) として同じ日数割合と仮定し、各経時特徴量データを動的に区分した。時間帯区分は日の出から日没間の時間を 3 等分し、それぞれ、時間帯 I、時間帯 II、時間帯 III、日没から翌日の日の出までの時間を 2 等分し、それぞれ、時間帯 IV、時間帯 V とした (図 3)。日の出、日没の時間は散乱光センサのセンサデータをもとに 1 日ごとに動的に決定し、時間帯を区分する。抽出された経時特徴量のうち、時間帯 V は夜間の時間帯であるため、光合成が行われることはなく、明らかにトマト

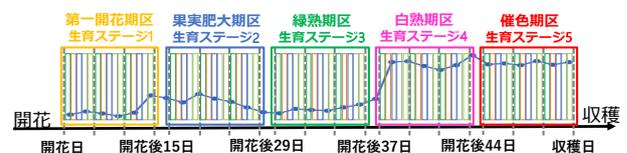


図 2 生育ステージ区分目安

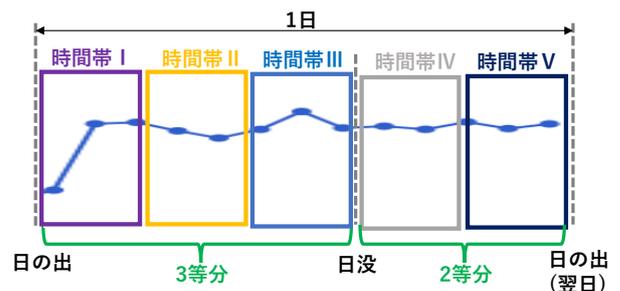


図 3 時間帯区分

の収穫時糖度とは関係性がないと判断したため、事前に除去した。また、基本統計量は大域的な変化を表現することができる。積算値、最大値、最小値、平均値を用いた。なお、以後は積算値、最大値、最小値、平均値をそれぞれ、sum, max, min, ave と表現する。

4.2 相関分析結果

相関分析を行う理由の一つは、相関の高い経時特徴量を精査することで、正規化回帰モデルに用いる経時特徴量を事前にクレンジングすることである。クレンジングを行う前の収穫時糖度と経時特徴量との相関分析結果のうち相関係数の絶対値が大きい上位 10 個の特徴量を図 4 に示す。図 4 より、複数の生育ステージにおいて、時間帯IVの光量の積算値が収穫時糖度との相関が高いことが分かる。しかし、時間帯IVは日没後であるため、光量子量と相関が高いという結果は異常である。そのため、正規化回帰モデルを用いた特徴量選択においてノイズとなり、誤った解釈をしてしまう恐れがあるため、事前に除去する。異常と判断できる相関の高い経時特徴量の除去を一般に相関が高いとされている、相関係数の絶対値が 0.7 以上の特徴量に対して行うことで、学習器の誤った解釈を抑制する。クレンジング後の収穫時糖度との相関係数の絶対値が大きい経時特徴量の上位 10 個の特徴量を図 5 に示す。図 5 より、最も相関の高い経時特徴量は生育ステージ 1 の時間帯IIにおける光量子量の積算値であった。生育ステージに着目すると、生育ステージIIは時間帯IIだけではなく、時間帯Iにおいても光量子量の積算値との相関が高いことがわかる。時間帯Iと時間帯IIは、日の出からの時間帯であることから、生育ステージ 1 において光強度が増加したことで、光合成速度が光飽和点付近まで上昇し、光合成を促進させた[20]といえる。また生育ステージ 3 においても時間帯Iと時間帯IIにおける光量子量の積算値との相関が高く、同様のことがいえる。そのため、収穫時糖度の高い栽培処理区ほど、生育ステージ 1 と生育ステージ 3 における光量子量が多く、光合成を介して光合成産物の濃度勾配が高くなったことによって果実

への転流速度が高まり[21][22]、果実の糖度が高まったと考える。時間帯に着目すると、時間帯Iの経時特徴量との相関が高いことがわかる。時間帯IIは日の出を分割の指標とした時間帯であり、光合成が開始される時間帯と解釈することができる。光合成速度は、光条件が良い正午頃に最大となり、その後徐々に低下するという推移を示す[23][24]ため、光量子量の増加による光条件の向上が光合成速度を高め、収穫時の果実糖度を高めたと考えられる。

全体の相関係数の分布を図 6 に示す。図 6 より、生育ステージ 1 と生育ステージ 2 における湿度との相関が高く、収穫日に近づくに連れて、相関が低く推移していることがわかる。植物は、乾燥時に気孔と呼ばれる体内の水分量を調節するための口を閉じることで、乾燥時の水分を保持する。湿度が高くなると気孔を開くことで体内の水分を放出し、CO₂を吸収する。湿度は時間帯Iが特に相関が高い。これは、日の出によって光量子量が増加したことで栽培環境の気温が上昇し、蒸発散が行われ湿度が高まったことで植物の気孔が開き、CO₂吸収効率を向上したことで、光合成速度が上昇したことで光合成が促進され、糖度の上昇につ

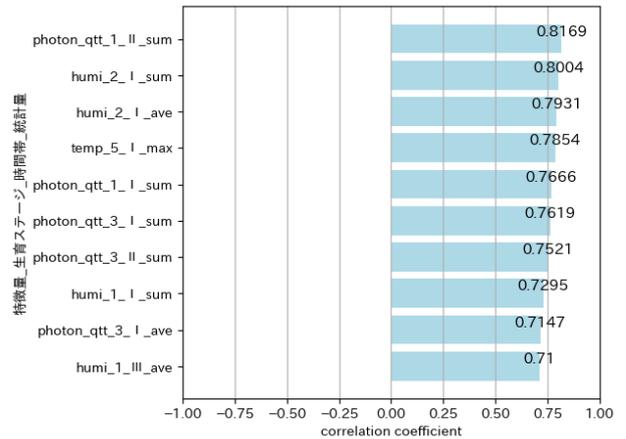


図 5 クレンジング後の各特徴量と収穫時糖度との相関係数

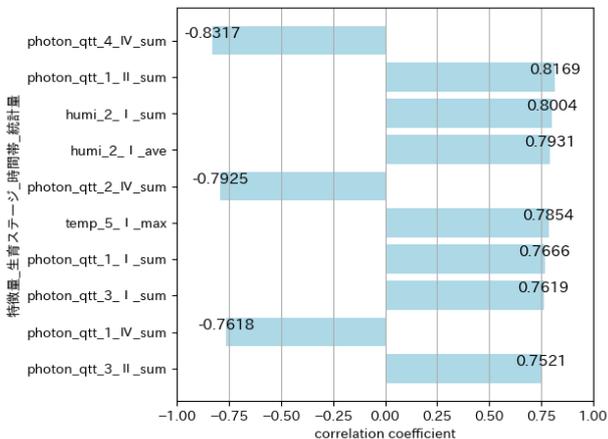


図 4 クレンジング前の各特徴量と収穫時糖度との相関係数

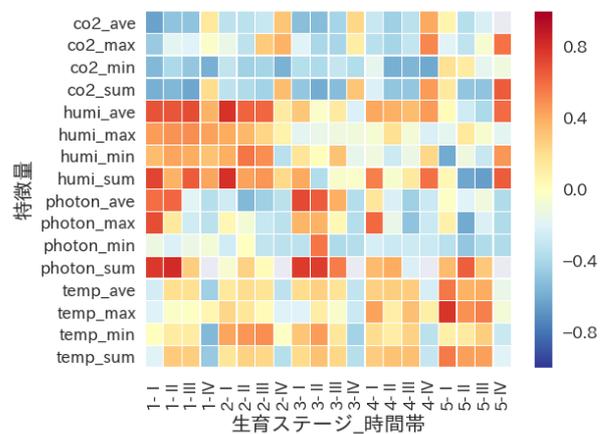


図 6 クレンジング後の各特徴量と収穫時糖度との相関係数分布

ながったと考える。生育ステージが推移していくと共に相関が低くなるため、植物の生育が進むにつれて、湿度の収穫時糖度への影響は小さくなると考えられる。

4.3 正則化回帰モデルによる特徴量選択

クレンジング後の経時特徴量に対して、正則化回帰モデルのうち、Lasso 回帰と Elastic Net 回帰により選択された経時特徴量とその回帰係数をそれぞれ図 7 と図 8 に示す。Lasso 回帰と Elastic Net 回帰ともに、特に回帰係数の大きな特徴量は共通しており、生育ステージ 3 と生育ステージ 4 の CO₂ 濃度が収穫時糖度に関与する経時特徴量として選択された。CO₂ 濃度は生育ステージ 4 の時間帯IIにおける最小値が収穫時糖度に関与する経時特徴量として選択されている。しかし、光合成を促進させるためには CO₂ は必須であり、最小値が選択されていることは、特徴量選択において誤った解釈が行われてしまったといえる。この問題はクレンジングの手法を改良し、更にノイズとなる特徴量を事前に除去することで解決可能であると考えられる。また、光量子量は生育ステージ 3 の時間帯IIの最大値が特に重要な経時特徴量として選択された。時間帯IIは 1 日の中で最も

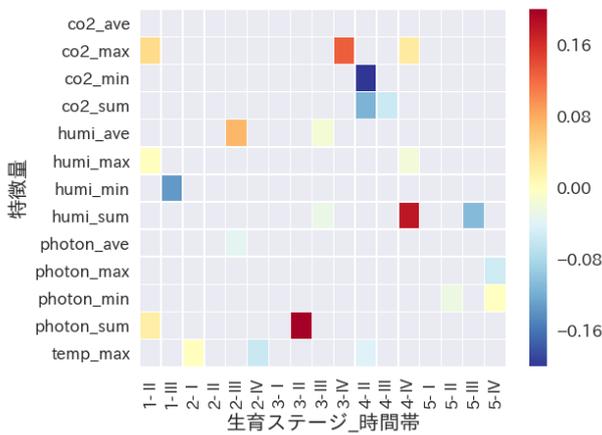


図 7 Lasso 回帰により選択された特徴量とその回帰係数

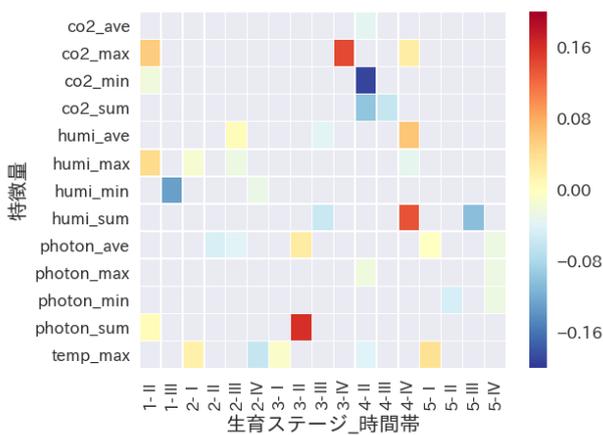


図 8 Elastic Net 回帰により選択された特徴量とその回帰係数

光が強く照射される時間帯であることより、光量子量の増加によって光合成が促進され、収穫時の果実糖度が高まったと考える。この結果は相関分析の結果とも一致しており、トマトの収穫時糖度は生育ステージ 3 の光条件を向上させ、光合成を促進することが重要であると考えられる。今回の実験では、生育ステージ 3 は緑熟期を想定している。トマトは緑熟期に収穫時の 7 割程度の糖度を含有し、緑熟期以降に更に糖度が上昇する[25]。そのため、緑熟期の光量子量を増加させることが、収穫時糖度を向上する要因になると期待できる。

4.4 選択された特徴量による推定

特徴量選択の効果を示すために、Lasso 回帰と Elastic Net 回帰によって選択された経時特徴量を用いて SVR と Random Forest による収穫時糖度の推定を行った。推定結果をそれぞれ図 9 と図 10 に示す。図 9 と図 10 より、SVR、Random Forest のいずれにおいても、Elastic Net 回帰によって選択された経時特徴量を用いた場合が最も精度が良いことが分かる。SVR に比べて、Random Forest の推定誤差が大きかった理由について考察する。Random Forest は決定木を大量に作成し、その平均値を推定結果として用いる回帰手法である。Lasso 回帰では 24 次元、Elastic Net 回帰では 37 次元と選択された特徴量の数が少なく、大量に異なる決定木が作成できなかったため、推定精度が SVR に比べて低くなったと考えられる。特徴量選択前の経時特徴量で構築した回帰モデルより、選択された経時特徴量のみを用いて構築した回帰モデルで高い推定精度を実現できたことから、選択された経時特徴量は収穫時糖度への関与の大きい経時特徴量である可能性が高いといえる。また、Lasso 回帰で選択された経時特徴量を用いて構築した回帰モデルが選択前の経時特徴量で構築した回帰モデルよりも推定精度が低く

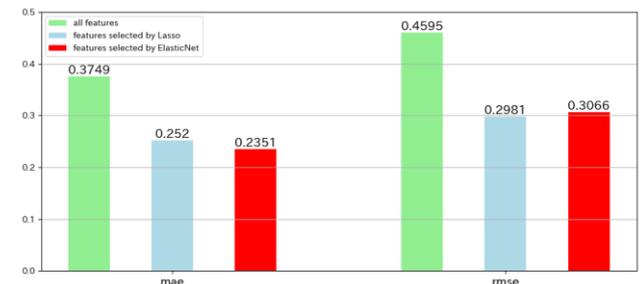


図 9 SVR による推定誤差

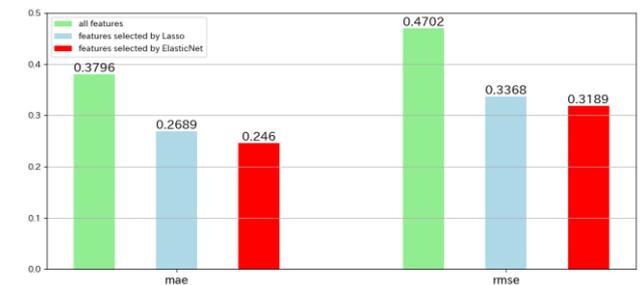


図 10 Random Forest による推定誤差

なった原因については、Lasso 回帰の問題点である相関の高い変数群の中から1つしか変数選択を行わないという問題によって選択されなかった経時特徴量の推定精度への影響や、選択された経時特徴量の数が極端に少なかったことが原因であると考えられる。Elastic Net 回帰と Lasso 回帰における変数選択結果の違いが推定精度にもたらした影響の分析に関しては、今後の課題としたい。

5. おわりに

本研究では、植物の生育状況と時間帯を考慮した栽培環境と収穫時品質の関連性を分析する手法の提案として、正則化回帰モデルを用いた特徴量選択による分析手法を提案した。提案手法により、植物収穫時品質に関与する経時特徴量の選択を行い、選択された経時特徴量のみを用いた推定精度と選択前の経時特徴量を用いた時の推定精度の比較を行ったところ、選択前の経時特徴量を用いた時の推定精度よりも高精度で推定できた。そのため、正則化回帰モデルによる収穫時品質に関与する特徴量の選択は、植物栽培環境への適用が可能であると考えられる。

今後、他の栽培環境に提案手法を適用することによる提案手法の汎用性の評価と、クレンジングの手法のさらなる改良を検討する。また、植物生理学からの観点や、気象情報など別の環境条件を考慮することで特徴量選択の精度を高め、農業技術の形式知化の確立に貢献していきたいと考える。

謝辞

本研究の一部は、JST さきがけ (JPMJPR15O5) と静岡県新成長戦略研究の支援を受けて実施されました。

参考文献

- [1] 横塚弘毅."山梨県における積算温度に基づいたブドウ糖度の予測,"ASEV 日本ブドウ・ワイン学会誌 17.1, pp.7-13, 2006.
- [2] 山下勇輝,他."トマト栽培実験による環境データ収集と環境要因の評価," 第77回全国大会講演論文集 2015.1, pp.813-814, 2015.
- [3] 吉田佳子,他 "保水シート耕のトマト一段密植栽培における夏季遮光の影響と果実糖度及び収量向上," 大分県農林水産研究指導センター研究報告 2, pp.1-14, 2011.
- [4] 浜本浩,他."3段取りトマト栽培における群落内補光の時間帯が収量に及ぼす効果と補光の経済性," 植物環境工学 22.2, pp.95-99, 2010.
- [5] 望月龍也,他."トマト果実における糖含量およびその栽培・環境条件に対する安定性の品種間差異," 園芸学会雑誌 68.5, pp.1000-1006, 1999.
- [6] 峰岸達也,他."ロジスティック分析でのステップワイズ法と決定木による属性選択法の実データをもちいた比較," フェジシステムシンポジウム講演論文集 25, A2-02, 2009.
- [7] Wold, H. "Soft Modelling by Latent Variables: The Non-Linear Iterative Partial Least Squares (NIPALS) Approach," Journal of Applied Probability 12.51, pp.117-142, 1975.
- [8] 橋本淳樹,他."PLS 回帰におけるモデル選択," アカデミア情報理工学編 10, pp.39-49, 2010.
- [9] Liaw, A., et al., "Classification and regression by randomForest," R news 2.3, pp.18-22, 2002.
- [10] 廣瀬慧."Lasso タイプの正則化法に基づくスパース推定法を用いた超高次元データ解析," 数理解析研究所講究録 1908, pp.57-77, 2014.
- [11] 阪本亘,他."正則化法を用いたロジスティック回帰モデルによる多次元データでの変数選択手法に関する研究," 数理解析研究所講究録 1703, pp.32-52, 2010.
- [12] 廣瀬慧."スパースモデリングとモデル選択," 電子情報通信学会誌 99.5, pp.392-399, 2016.
- [13] Tibshirani, R. "Regression shrinkage and selection via the lasso," Journal of the Royal Statistical Society 58.1, pp.267-288, 1996.
- [14] Hui Zou., et al., "Regularization and Variable Selection via the Elastic Net," Journal of the Royal Statistical Society 67.2, pp.301-320, 2005.
- [15] Hoerl, A., et al., "Ridge regression: Biased estimation for nonorthogonal problems" Technometrics, 12.1, pp.55-67, 1970.
- [16] Basak, D., et al., "Support vector regression," Neural Information Processing – Letters and Reviews 11.10, pp.203-224, 2007.
- [17] Ibayashi, H., et al., "A Reliable Wireless Control System for Tomato Hydroponics," Sensors 16.5, pp.664, 2016.
- [18] 野口有里紗,他."一段密植栽培トマトの果実品質に及ぼす濃度培養液処理開始時期の影響," 東京農大農学集報 57.1, pp.9-13, 2012.
- [19] 城島十三夫,他."露地・ハウス栽培の桃色および赤色系トマト品種の果実の肥大・着色特性と高温期における色素の形成," 園芸学会雑誌 63.3, pp.581-588, 1994.
- [20] 稲垣昇,他."アスパラガスの光合成に及ぼす光強度,CO₂濃度及び温度の影響," 園芸学会雑誌 58.2, pp.369-376, 1989.
- [21] 吉岡宏,他."果菜類における光合成産物の動態に関する研究(5)," 野菜試験場報告 9, pp.63-81, 1981.
- [22] 荒木卓哉,他."トマトにおける果実生長および光合成産物の転流の動態に対する環境作用(第3報)," 生物環境調節 39.1, pp.53-58, 2001.
- [23] 藤澤弘幸,他."JM1, JM7, JM8 および M.9 台木がリンゴ葉の光合成速度に及ぼす影響," 園芸学研究 9.2, pp.171-176, 2010.
- [24] 石原邦,他."湛水状態の水田に生育する水稻の個葉光合成速度の日変化に影響する要因について,"日本作物学会紀事 56.1, pp.8-17, 1987.
- [25] 石井孝典,他."トマト品種の果実成分と熟度の関係," 東北農業研究 47.1, pp.275-276, 1994.