

時間情報を考慮した収穫時品質予測手法の検討

中西豪太[†] 峰野博史^{††}

[†]静岡大学情報学部 ^{††}静岡大学大学院情報学領域 [‡]JST さきがけ

1. はじめに

近年，農業従事者の高齢化に伴い，熟練農家をもつ高度な栽培技術が新規就農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．例えば，熟練農家は栽培期間中に長年の経験と勘に基づいて収穫時品質を見通すことで栽培の安定化を図っているが，経験の浅い新規就農者が見通しを立てることは困難である．

本研究では，収穫時品質に与える影響が生育状態ごとに異なることに着目し，生育状態と時間情報を考慮した高次元な特徴量から収穫時品質予測に重要な特徴量選択を行うことで，高精度な植物収穫時品質を予測する手法を提案する．

2. 関連研究

植物の収穫時品質を予測する研究として，積算温度を用いた収穫時品質の予測手法[1]や，特定の環境データに着目した品質との関係性の分析[2]による予測手法などがある．[1]は温度が植物の成長を促進させることに着目し，閾値を越えた温度の積算値を用いた統計処理により収穫時の糖度予測を行っている．[2]は植物の総乾物生産量が積算受光量に比例することに着目し，光利用効率の向上要因を分析することによる収量の増加，および収量予測の可能性を提案している．しかし，植物は生育に伴い栽培環境やその時間帯が与える影響が異なると考える．つまり，生育ステージや時間帯といった植物の成長における時間情報を考慮した機械学習による収穫時品質を予測する手法の検討が必要である．

3. 提案手法

提案手法の概要を図 1 に示す．提案手法は，図 1 に示すように A. 特徴量選択部と B. 収穫時品質予測部に大別される．A. 特徴量選択部では，まず①経時特徴量データを生育ステージごとに分割した後，時間帯情報ごとに分割する．次に，②分割された各経時特徴量データから基本統計量を算出し，生育状態と時間情報を考慮した特徴量の算出を行う．基本統計量は，一般に農業従事者は特定の 1 日の変化ではなく，植物の長期的な変化を考慮した栽培技術を用いるため，大域的な変化を表現可能な積算値，平均値，最大値，最小値，標準

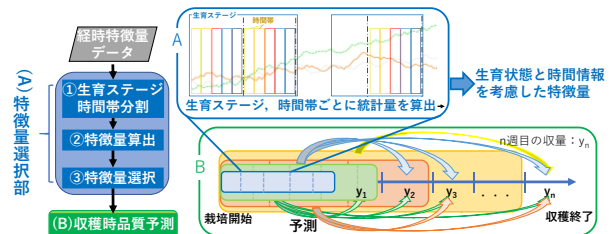


図 1 提案手法の概要

偏差とした．また，栽培を通した特徴量として温度や日射量などの環境データの積算値，栽培開始からの経過日数など栽培した季節の考慮が可能な特徴量を算出する．③算出された特徴量に対し RFE (Recursive Feature Elimination[3]) によって特徴量選択を行い，収穫時品質に対する重要な特徴量の選択を行う．RFE は，モデル作成時の特徴量の重み付けが可能な機械学習手法を用いてモデル作成を行い，重要度の低い特徴量の削除を再帰的に繰り返す手法である．RFE を用いた特徴量選択によって，機械的に収穫時品質に対して重要度の高い特徴量の選択が可能であると考えられる．

B. 収穫時品質予測部では，A. 特徴量選択部によって選択された特徴量を用いて回帰モデルを構築し，収穫時品質の予測を行う．A. 特徴量選択部において選択された特徴量を用いて構築した回帰モデルによる予測時の誤差と選択する前の特徴量を用いた予測時の誤差を比較することで，特徴量選択の有意性を評価する．

4. 基礎評価

4.1. 検証内容

提案手法の基礎評価として，収穫時品質に対して重要な特徴量の選択と選択された特徴量を用いた収量の予測を行った．使用したデータは 2016 年 7 月から 2017 年 12 月にかけて複数のトマト栽培農場において低段密植栽培された 38 回の栽培データを使用した．いずれの栽培もトマトの品種は CF 桃太郎ヨークを使用し，栽培データは温度(°C)，飽差(g/m³)，日射量(J/m²/s)，CO₂濃度(ppm)を 1 分周期で測定したデータを使用した．また，収量は選果機により 1 日毎に計測された収量値を用いた．

農学者の知見を参考に生育ステージの区分は，定植日から予測対象週の収穫開始日までを 5 等分し生育ステージ I ～生育ステージ V とした．時間帯区分は，栽培した農場の日の出時刻から日没時刻を 3 等分し，それぞれ時間帯 1，時間帯 2，時間帯

Investigation of harvest quality prediction method considering time information

Gota Nakanishi[†], Hiroshi Mineno^{††}

[†]Faculty of Informatics, Shizuoka University

^{††}College of Informatics, Shizuoka University

[‡]JST PRESTO

3, 日没から翌日の日の出時刻を 2 等分しそれぞれ時間帯 4, 時間帯 5 と区分した. 日の出と日没時刻を時間帯の区分指標とすることで, 植物の成長要因である光合成の考慮が可能となる. また, A. 特徴量選択部の RFE で用いる外部学習器は最終的に XGBoost[5]を使用することとした.

予測する収量は一般的に低段密植栽培トマトの収穫が終了する 4 週間後までの 1 週間ごとの総収量を対象とし, 使用する学習器は代表的な非線形学習器であり, 良好な性能を得られた XGBoost を用いることとした. 汎化性能の評価には交差検証を使用し, 予測結果の誤差指標として平均絶対誤差(MAE:Mean Absolute Error)と平均二乗誤差平方根(RMSE:Root Mean Squared Error)を使用した.

4.2. 検証結果

図 2 に RFE によって選択された, 収量に対して重要度の高い特徴量を示す. 生育ステージ V の時間帯 1, 時間帯 2 の日射量が重要度の高い特徴量として選択された事がわかる. 時間帯 1 と時間帯 2 は日の出時刻から昼過ぎの時刻であるため, 生育ステージ V において日の出から光強度が高まっていくことで光合成速度が大きくなり[2], 光合成反応が促進されたことが収量の向上に繋がったと考える. また, 同様に生育ステージ V における日の出後の温度の積算値の重要度が高いという結果となった. 植物の光合成速度を決定する要因である温度, 光強度, CO₂ 濃度のそれぞれが限定要因となるため, 温度と光強度の両方が高まったことが収量の増加に繋がったと考える. また, 植物は一般的に光合成速度が正午頃に最大となり, その後は徐々に低下するという推移を示す[4]ことから, 生育ステージ V の日の出後から昼過ぎの光合成速度の増加が収量に対して重要であり, 選択結果は妥当であると言える.

以上の重要度の高い特徴量を用いて機械学習を行い, 1 週間ごとの総収量を予測した際の誤差を図 3 に示す. 全特徴量を用いた予測(図 3(a))と比較して選択された特徴量を用いた予測(図 3(b))では全体として MAE が 76.9kg, RMSE が 111.79kg となり, 最大で MAE が 48.9%, RMSE が 41.0%の誤差が削減された. つまり, 選択された特徴量は収量予測に対して重要な特徴量であると言える. また, 予測対象週が 1 週目と 2 週目の時, MAE と RMSE の両方が大きかった. RMSE は外れ値の影響を受けやすい誤差指標であり, 予測対象週が 1 週目と 2 週目における収量の例外的な値を予測できていないため, 誤差が大きくなったと考える. 以上より, 生育ステージや時間帯といった植物の成長における時間帯情報を考慮した機械学習を用いることで植物の収穫時の品質予測に重要な特徴量の選択を実現し, 収穫時の品質を高精度に予測できる見通しを得た.

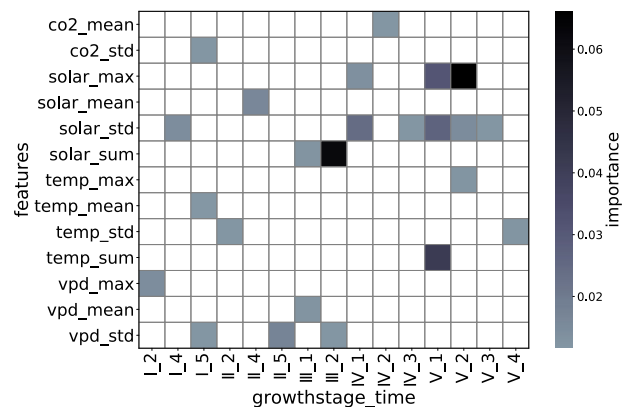


図 2 特徴量選択結果

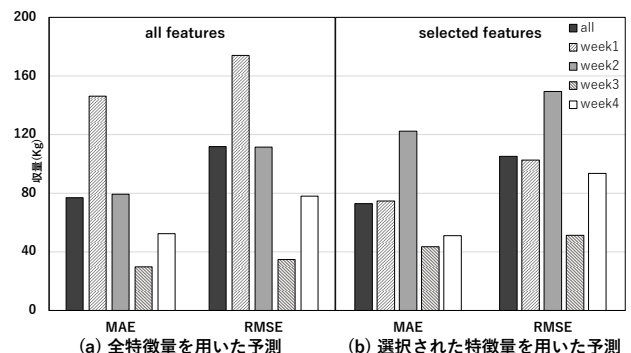


図 3 予測時の誤差

5. おわりに

本研究では, 生育状態と時間帯を考慮した特徴量による収量を予測する手法の検討を行った. 提案手法により, 機械的に収量予測に対して重要度の高い特徴量の選択を実現した. 今後, データセットの拡充だけではなく, 例外的な収量値の前処理方法の工夫や例外的な値と関係のある環境要因の分析, 灌水情報などを考慮することで, 精度向上を目指し農業技術の形式知化に貢献していく.

謝辞 本研究の一部は JST さきがけ(JPMJPR1505), テラスマイル株式会社の支援を受け実施された.

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, J.ASEV Jpn., Vol.17, No.1, pp.7-13(2006).
- [2] 東出忠桐:施設トマトの収量増加を目的とした受光と物質生産の関係の利用, 園芸学研究, Vol.17, No.2, pp.133-146(2018).
- [3] Guyon, I., et al.: Gene selection for Cancer classification using support vector machines, Mach. Learn., Vol.46(1-3), pp.389-422(2002).
- [4] 藤澤弘幸, 他: JM1, JM7, JM8 および M.9 台木がリンゴの葉の光合成速度に及ぼす影響, 園芸学研究, Vol.9, No.2, pp.171-176(2010).
- [5] Tianqi Chen, et al.: XGBoost: A Scalable Tree Boosting System, ACM SIGKDD, pp.785-794(2016).