

# 機械学習を用いた糖度予測モデルの開発

吉田研祐<sup>†1</sup> 大場みち子<sup>†2</sup> 高森満<sup>†3</sup>

**概要:** 現在、食料ロスについて問題視されており、そのうちの1つに農家ででの選果作業で大きさや形の規格に漏れた規格外野菜が含まれる。食料ロスの削減の取り組みとして、規格外野菜を低価格もしくは無償で消費者へ提供する事例は見られる。しかし、規格外野菜に、糖度の規格で選果し、糖度別に販売する例は見られない。その理由として、選果作業に適した糖度計が見られないことが考えられる。本研究は、規格外野菜の廃棄量削減を目的とし、機械学習を用いた糖度の予測モデルを開発する。開発した予測モデルを用いて、大きさや形の規格外野菜を糖度別に分類する。糖度が高い野菜を高糖度な商品として販売できれば、規格外野菜の廃棄量は削減でき、食料ロスの問題解決につながると思われる。本研究は、ミニトマトを対象とする。ミニトマトのデータを記録するアプリケーションの開発や、環境要因の調査、複数の機械学習の予測精度の結果を報告する。

**キーワード:** ミニトマト, 糖度, 機械学習

## Development of Prediction Models for Sugar Concentration Using Machine Learning

KENSUKE YOSHIDA<sup>†1</sup> MICHIKO OBA<sup>†2</sup>  
MITSURU TAKAMORI<sup>†3</sup>

**Abstract:** Today, food loss is brought into question and contains disposal of imperfect vegetables which are not meeting quality standards of size or shape during sorting vegetables at each farm. There are many cases to provide these imperfect vegetables with consumers cheaply or free of charge as an approach to reduce food loss. However, imperfect vegetables are not sorted based on a standard of sugar concentration, and there is no case of selling them by sugar concentration. As a reason, sugar refractometers are considered to be not suitable for this sorting vegetables. In this study, we aim to reduce the amount of disposal of imperfect vegetables and develop a prediction model of sugar concentration using machine learning. We classify imperfect vegetables which are not meeting standards of size or shape based on a standard of sugar concentration using the developed prediction models. If vegetables with high sugar concentration can be sold as a product with high sugar concentration, it is possible to reduce the amount of disposal of imperfect vegetables and to solve the problem of food loss. This study targets cherry tomatoes. Here, we report the development of applications that record data of cherry tomatoes, investigation of environmental factors, and results of prediction accuracy of multiple machine learning.

**Keywords:** Cherry Tomatoes, Sugar Concentration, Machine Learning

### 1. はじめに

世界的な食料問題の1つに、可食部分（食べることのできる部分）が残った状態で廃棄する食料ロスが挙げられる[1]。この食料ロスの問題は、国内でも問題視されており、年間で約640万トンもの食料が廃棄されていると報告されている[2]。食料ロスは、家庭系と事業系の2つに大別され、それぞれの年間の廃棄量は、約310~330万トンと言われている[2]。家庭系とは、家庭での食べ残しを指し、事業系には、レストラン・スーパーでの売れ残りや食べ残し、農家での選果作業で規格漏れになった規格外野菜が含まれる。規格とは、農家での選果（野菜を等級別に分別する作業）の1つの基準である。各野菜の規格は様々であるが、大きさ・重量・色が主な規格である。野菜が規格に適合するか

どうかは、農家の目視で行われることがほとんどであり、独自の規格を設けている場合もある。食料ロスの削減の取り組みとして、大きさや形で規格外となった野菜を、規格に適合した野菜よりも低価格もしくは無償で消費者へ提供する企業は多く見られる[3][4]。しかし、大きさや形の規格に外れた野菜に、糖度のような別の規格を設けて選果し、糖度別に販売する例はあまり見られない。糖度は品質評価の主な指標であり、糖度の高さと消費者の嗜好には高い相関がある[5]。糖度の規格を設けて選果する例が見られない理由として、このような選果作業に適した糖度計が見当たらないことが考えられる。糖度計は、破壊式と非破壊式の2種類に大きく分けられる。破壊式糖度計は、1台あたり安価である。しかし、野菜1個の糖度を計測する手間が非常に大きく、また野菜に傷をつけて果汁を採取しなくてはならない[6]。一度傷をつけた野菜は、売り物として扱うことができない。非破壊式糖度計は、作物に傷をつけずに光センサーで糖度を短時間で計測することができる。しかし、1台あたり約30~100万円と非常に高額である[7]。また、選果では多量の野菜を扱うため、複数人で同時に行うことが多

<sup>†1</sup> 公立はこだて未来大学大学院  
Future University Hakodate Graduate School

<sup>†2</sup> 公立はこだて未来大学  
Future University Hakodate

<sup>†3</sup> 株式会社アプレ  
Apure Inc.

い。非破壊式糖度計を用いて糖度別に選果を行う場合、複数台の非破壊式糖度計が必要となり、多額の出費を伴ってしまう。そのため、農家では非破壊式糖度計による糖度別の選果を一般的に実現できていない。

このように、現在広く使用されている2種類の糖度計では、「機能」と「価格」のどちらか一方に偏っており、規格外野菜に糖度という新たな規格を設けて選果する場合に適していない。そこで、本研究では、糖度別の選果に適した糖度計測システムを開発する。従業員が、各自のスマートフォンから糖度計測システムにデータを入力すると、機械学習で糖度を予測し、予測結果を表示することで、野菜を傷つけることなく糖度を計測でき、複数人で糖度の選果作業を行うことができると考えた。糖度別に分類し、高精度な商品として販売もしくは、ある糖度カテゴリに分類されたものを加工品の材料として扱うことができれば、規格外野菜の廃棄量は削減でき、食料ロスの問題の改善につながる。

本研究の目的は、規格外野菜の廃棄量削減である。そのために、機械学習を用いた糖度予測モデルの開発を研究目標とする。

## 2. 関連研究と課題

### 2.1 関連研究

関連研究として、野菜の表面画像から、野菜の品質の1つである「熟度」をランダムフォレストとサポートベクターマシンを使用して分類している研究がある[8]。この研究では、色、見た目、熟度などの指標は、消費者の反応に大きな影響をもたらすため、非常に重要な役割を担っている。しかし、これらは現在も生産者の主観的な判断や経験値に依存しており、高品質の野菜を消費者に届けるためには、客観的で正確に野菜を評価する必要がある。また、熟度の客観的評価を実現することで、未熟の野菜の収穫を防ぐことができる。トマトと赤ピーマンを対象とし、それぞれ250枚、175枚の画像(3,664×2,748ピクセル)のデータセットである。関連研究の解決アプローチは、前処理・特徴抽出・分類の3つのプロセスで構成されている。

前処理では、撮影した画像を250×250ピクセルにリサイズと、背景除去、RGB表色系からHSV表色系への変換を行っている。

特徴抽出では、主成分分析を用いて次元を削減している。主成分分析とは、できるだけ情報を失うことなく多くの次元の測定値を少ない個数の総合的指標で集約して記述する方法であり、次元削減を目的として用いられている。

分類では、交差検証を使用し、線形カーネルSVM、多項式カーネルSVM、ランダムフォレストの正解率を求め、精度を比較している。入力データであるトマトと赤ピーマンの画像を5つの熟度に分割している。結果として、線形カーネルSVMでの分類精度が最も良く、トマトと赤ピーマンに対して正解率90%以上の精度で分類している。

### 2.2 課題

関連研究[8]は、農作物の外観情報を学習データとして、高精度に熟度を分類することができたと述べているが、対象を「糖度」とした場合、学習データに外観情報のみを用いるのは不十分である。その理由として、農作物の品質は、

次に示す通り、環境要因の影響に左右されるからである。

栽培環境の二酸化炭素の濃度と相対湿度を変化させ、キュウリの幼植物の生育を調査した関連研究[9]によると、二酸化炭素濃度を500ppm程度で相対湿度86%の場合より、二酸化炭素濃度1,000ppm程度で相対湿度86%の場合は、葉面積・光合成速度が有意に高い結果となったと報告している。葉面積と糖度には関係があり、葉面積が高くなることで、果実の内部に転移する光合成産物が増加し、糖度が高くなるという報告がある[10]。

昼夜温を制御した環境下で果実発育期のイチゴを育て、果実の生育日数や品質に与える気温の影響を調査した研究がある[11]。この研究では、昼夜温を15°C、20°C、25°Cに統一した環境をそれぞれ作り、イチゴの生育日数(頂花房の第一番花の開花～完熟までの期間)と糖度(Brix)を調査したところ、25°Cに統一した場合は生育日数が23日と最も早かったのに対し、15°Cに統一した場合は、40.4日と最も遅く、生育日数に20日近くの違いがあると述べている。また、25°Cに統一した場合の糖度は、5.2%であったのに対し、15°Cに統一した場合の糖度は10.6%と約2倍高くなったと述べている。

このように、環境要因は野菜の品質に非常に大きな影響を与えていることがわかる。関連研究[8]のように、外観情報のみを用いた場合、正確に糖度を捉えるのは困難である。

## 3. 提案手法

2.2の課題より、糖度を正確に捉えるために、農作物の外観データと環境データに着目し、機械学習で糖度を予測する。本研究における課題は、次に示す通りである。

- (1)農作物の外観・環境データの正確かつ効率的な記録
- (2)精度の高い精度予測モデルを開発

2つの課題に対する解決アプローチは、次に示す通りである。

- (1)データ記録するためのアプリケーションを開発
- (2)機械学習の精度の高い予測モデルを複数開発して比較

3.1に対象とする実験施設と農作物について、3.2にデータ記録用のアプリケーションの開発について、3.3に機械学習の候補について述べる。

### 3.1 対象とする実験施設と農作物

本研究の対象となる実験施設は、水耕栽培を行っている植物工場Aである。植物工場とは、年間を通じて作物を安定に栽培できる施設であり、主な特徴は3つある。

- (1)気温・湿度・二酸化炭素濃度などの環境データのセンシング
- (2)二酸化炭素付与装置、温度調節機、作物への人工光(LED)照射装置のコントロール
- (3)圃場やビニールハウスよりも、外気からの影響が小さい

対象とする農作物は、ミニトマトの品種Mである。これを選択した理由は、次のとおりである。

- (1)植物工場Aでは、果菜類(1つの苗で成実・着色を繰り返す、果肉が存在する作物)のミニトマトを主に栽培している

(2)ミニトマトは生育期間が長く、環境要因の変化を受けやすい

植物工場 A では、ミニトマトを図 1 のように多段栽培しており、「2~6 段」の間に成実している。



図 1 ミニトマトの多段栽培

植物工場 A でセンシングしている環境データは、気温・湿度・二酸化炭素濃度・人工光の照度・養液濃度など多岐にわたる。これらの環境データは 10 分ごとにセンシングされ、クラウドに保存されており、保存された環境データを利用する。

対象とする環境データは、気温・湿度・二酸化炭素とし、照度と養液濃度は、機械学習の学習データの候補から除いた。2つの環境データを除いた理由を次に示す。

(1)植物工場 A は、白色・淡紅色の人工光を農作物に照射しているが、照射時間を朝 4 時~夜 7 時までの 15 時間と定めている。時間ごとの照度の強さもコントロールしており、日ごとに違いが見られない。

(2)水耕栽培の養液の濃度も常に一定であり、照度と同じようにデータの差が表れない。

このように、変化が表れないデータを機械学習で用いることは不適切であり、2つの環境データの重要度は低いと判断したためである。

### 3.2 データ記録アプリケーションの開発

農作物の播種日(種を蒔いた日)、糖度、段数(成実した位置)、大きさを入力し、クラウドにデータが記録される Android アプリケーションを開発した。図 2 は、開発し

たアプリケーションの画面例である。非破壊式糖度計[7]を使用して糖度を計測し、画面に入力する。段数を「2~6 段」から選択、色を「出荷できる色、出荷には至らないが着色、着色していない」から選択する。



図 2 データ記録アプリケーション画面

### 3.3 機械学習の候補

目的変数の糖度を多クラスに分類する。分類の手法として、k 近傍法、ランダムフォレスト、XGBoost、ニューラルネットワークの代表的な 4 つの機械学習アルゴリズムを用いる。機械学習のモデルを開発する環境は、Ubuntu 18.04 LTS, Anaconda 5.3.1, Python 3.6.6, keras 2.2.4, scikit-learn 0.20.2, xgboost とする。scikit-learn や xgboost, keras は、Python の機械学習のオープンソースライブラリである。これらのライブラリを使用して、k 近傍法、ランダムフォレスト、XGBoost、ニューラルネットワークの学習器を生成する。学習器に訓練データとテストデータをあてはめて結果を予測する。4 つの学習器の予測結果の正解率を求めて精度を比較する。正解率が最も高かった学習器が糖度の予測に最適であると判断する。

k 近傍法は、データセットの中から対象のデータと類似度が最も高い k 個の標本を抽出し、それらの標本が属するクラスを基に多数決でクラスラベルを求める方法である [12]。

ランダムフォレストは、アンサンブル学習の一種であり、複数の決定木を構築し、各決定木の予測値の多数決から最終的な目的変数を予測する手法である [13]。アンサンブル学習とは、複数の弱学習器を組み合わせることで強学習器を構築する方法である。

XGBoost は、C++ で実装された高速の勾配ブースティングアルゴリズムのことであり、機械学習の競技会である Kaggle で優れた結果を残していると報告がある [14]。ブースティングとは、直列的に弱学習器を結合し、最終的な予測精度を高めるアルゴリズムで、アンサンブル学習の一種である。XGBoost では、パラメータ推定での最適化手法に

勾配降下法を用いている。

ニューラルネットワークとは、人間の神経回路網の構成や動作をモデル化したものである[15]。複数のニューロンとシナプスで構成されており、ニューロンは、シナプスで繋がっている。前段のニューロンから重みをかけ合わせた入力信号を受け取ると、ニューロン内の膜電位が上昇する。この膜電位が閾値を超えると、ニューロンは発火し、後段のニューロンへと刺激を送る。これをネットワーク全体で繰り返すことで予測・分類の機能を果たす。

#### 4. 予備実験

予備実験の目的は、糖度を的確に予測するために、ミニトマトの糖度に関する要因を調査することである。使用するデータは、ミニトマトの糖度データと外観データである。実施する予備実験は、以下に示す通りである。

- (1)糖度と色の関係
- (2)糖度と段数の関係
- (3)糖度と大きさの関係
- (4)着色開始日から収穫日までの日数

##### 4.1 糖度と色の関係

「着色していない、出荷には至らないが着色、出荷できる色」をそれぞれ「not\_red, a\_little\_red, red」と表記し、糖度の範囲を「2以上3未満, 3以上4未満, 4以上5未満, 5以上6未満」の4つに区分する。対象とするデータは、2018年10月27日から11月13日までに生育中のミニトマトの275件である。図3は、糖度と色の関係性を入れ線グラフで表示したものである。x軸に糖度, y軸に各色の糖度占有割合を示す。

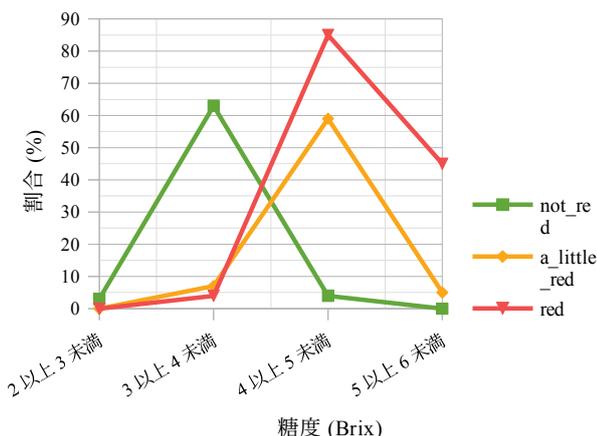


図3 糖度と色の関係

not\_red (緑線) は、糖度3以上4未満に集中しており、a\_little\_red, red (黄線, 赤線) は似たような傾向を示し、糖度4以上5未満に集中していることがわかった。糖度5以上6未満は、redの割合が高かった。この結果より、色が赤色に近づくほど糖度が高くなる傾向にあるといえる。not\_redとa\_little\_red, redの間で糖度データの差が大きいため、予備実験(2), (3)では、not\_redを対象外とした。またa\_little\_redとredの糖度が2以上3未満にほとんど存在

しないため、予備実験(2), (3)では、糖度2以上3未満を対象外とした。

##### 4.2 糖度と段数の関係

予備実験(1)の結果を踏まえ、糖度「2以上3未満」を除き、「3以上4未満, 4以上5未満, 5以上6未満」の3つに糖度の範囲を区分する。段数は、「2段, 3段, 4段, 5段, 6段」の5つに区分する。対象とするデータは、2018年10月27日から11月13日に生育中のミニトマト205件であり、予備実験(1)の結果を踏まえ、not\_redは対象外としている。図4は、段数ごとの糖度の割合を帯グラフで表示したものであり、x軸に各段数の糖度占有割合を示し、y軸に段数を示す。

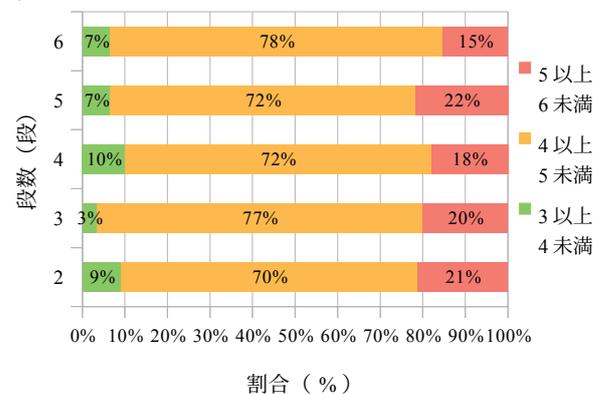


図4 糖度と段数の関係

図4より、糖度4以上5未満(黄色)が各段数で7割ほど占めていることがわかった。また、各段数の糖度分布に明確な違いは見られないため、段数と糖度の関係性は低いと判断できる。

##### 4.3 糖度と大きさの関係

ミニトマトの大きさを縦の長さ×横の長さで求め、糖度との相関を算出し、無相関検定より得られた相関は有意であるかどうかを調査する。対象とするデータは、予備実験(2)と同様の2018年10月27日から11月13日に生育中のミニトマト205件である。図5は、糖度とミニトマトの大きさの相関分布であり、x軸に糖度, y軸にミニトマトの大きさを示している。

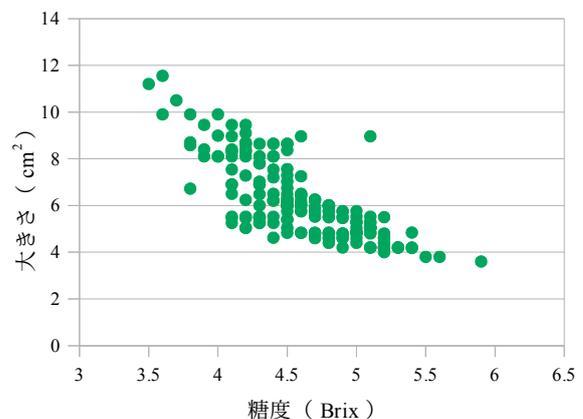


図5 糖度と大きさの関係

図5より、糖度と大きさに強い負の相関が見られた(相関係数=-0.77)。この結果より、ミニトマトの大きさが小さいほど、糖度が高い傾向にあることがわかった。また、表1は、無相関検定の結果である。p値が有意水準1%よりも小さいため、帰無仮説「母集団の相関係数が0である」を棄却でき、大きさと糖度には有意な相関があると言える。これらの結果より、ミニトマトの糖度予測モデルのパラメータに大きさを入れることで、糖度を正確に捉えることができる考えた。

表1 無相関検定

|          |          |
|----------|----------|
| 相関係数     | -0.77    |
| t 値      | 17.03    |
| df (自由度) | 204      |
| p 値      | 5.65E-41 |

#### 4.4 着色開始日から収穫日までの日数

ミニトマトが成実し、着色を開始してから収穫日までの平均日数を調査した。スマートフォンで定点撮影した複数の画像を並べ、タイムラプス動画を作成し、着色開始日から収穫日までの日数を動画より確認する調査方法を採用した。スマートフォンは、植物工場内のプランターを支える柱に1台固定し、常時電気を補給する状態にした。予備実験(4)で扱ったスマートフォンは、植物工場Aと株式会社Nの共同研究用端末である。このスマートフォンには、株式会社Nが開発した、一定時間の間隔で写真を自動撮影するアプリケーションが備わっている。このアプリケーションを使用して、ミニトマトの苗全体を数十分間隔で撮影した複数の写真を並べてタイムラプス動画を作成した。対象期間は、2019年1月1日から2月18日である。図6は、撮影したミニトマトの写真の例である。



図6 スマートフォンで自動撮影したミニトマトの例

タイムラプス動画から目視でミニトマトを確認できたのは、47個であった。ミニトマトを定点カメラで撮影したが、実が重なったり、生い茂った葉により途中で確認が困難になったミニトマトも複数あった。目視で確認できた47個のミニトマトの着色開始日から収穫日までの日数を調査した結果を図7に示した。着色開始日～収穫日までに要した日数は、5～9日間であり、最も多かった日数は7日間であった。

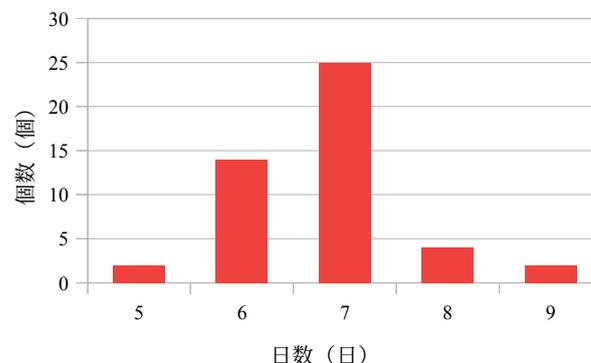


図7 ミニトマトの着色開始日から収穫日までの日数

## 5. 本実験

本章では、k近傍法、ランダムフォレスト、XGBoost、ニューラルネットワークの4手法の中から糖度の予測に最も適した手法を調査するために行った実験の内容と、その結果を述べる。

### 5.1 仮説

本実験を実施するにあたり、仮説を2つ立てる。

仮説(1): 7日間の環境データを考慮することで糖度を正確に捉えることができる

予備実験(1)で色と糖度の関係を明らかにしたように、実が赤くなるにつれて糖度が高くなる。予備実験(4)では、ミニトマトが着色開始してから収穫するまでの期間は7日間が最も多いことがわかった。これらの結果より、着色開始日から収穫日までの期間の環境データが糖度に影響を及ぼしており、この7日間の環境データを考慮することで糖度を正確に捉えることができるのではないかという仮説を立てる。

仮説(2): 機械学習を用いることで、季節ごとの環境データや糖度データの特徴を掴むことができる

植物工場では、環境を制御しているため、年間を通して安定的に栽培できるが、施設の環境は外気象の影響をある程度受ける。環境データと糖度データを学習していくことで、季節の環境変化による糖度の変化も柔軟に捉えることができ、年間を通して高い精度で糖度を予測できるのではないかという仮説を立てる。

## 5.2 対象データ

対象とするデータは、2018年11月28日～2019年6月15日に生育中のミニトマト2,176件分の糖度データ、大きさデータ、収穫日データ、環境データ（気温、湿度、二酸化炭素濃度）である。これらのデータを用いて、新たに特徴量の生成やスケーリングの前処理や、グリッドサーチでハイパーパラメータを最適化した後に、正解率を基準として評価を行う。

## 5.3 前処理

前処理の段階では、学習器が精度高く予測できるように、対象データから特徴量を新たに生成する作業を実施する。前処理で用いるデータは、収穫日データと環境データであり、表2のように、新しく生成した特徴量とする。以下にそれぞれの特徴量について詳しく説明する。

### ・四半期について

実験先の植物工場では、外気象の影響をわずかに受け、環境が変化する。そのため、季節ごとの環境の変化でミニトマトの糖度も変化すると考えられる。季節ごとの糖度変化も考慮する上で、特徴量に季節的要因が必要であると判断し、「四半期」という特徴量を生成する。これは、収穫日データから「月」のデータを抽出し、「4～6月」、「7～9月」、「10～12月」、「1～3月」の4つのグループに分ける。

### ・収穫日前7日間の環境データについて

実験を実施する上で用意した環境データは、1日の平均気温・湿度・二酸化炭素濃度である。5.1の仮説で述べたとおり、収穫日前7日間のデータを学習器に当てはめることにしたため、1日の平均データから収穫日前7日間のデータを新たに生成する必要がある。

本実験の目的変数であるミニトマトの糖度は、3.5～5.8%の範囲に分散しており、表3のように4つのカテゴリに分ける。

表2 特徴量の生成

| 収集したデータ    | 新しく生成した特徴量  | 備考   |
|------------|---|--|
| 収穫日        | 四半期   | 1st: 4～6月<br>2st: 7～9月<br>3rd: 10～12月<br>4th: 1～3月 |
| 日平均気温      | 7日間の平均気温<br>7日間の最高気温の平均<br>7日間の最低気温の平均<br>7日間の差分気温の平均 | ※差分気温とは、最高気温と最低気温の差                                |
| 日平均湿度      | 7日間の平均湿度  |  |
| 日平均二酸化炭素濃度 | 7日間の平均二酸化炭素濃度   |  |

表3 糖度カテゴリ化

| 糖度      | 糖度カテゴリ  |
|---------|---------|
| ～4.0    | sweet_1 |
| 4.1～4.5 | sweet_2 |
| 4.5～5.0 | sweet_3 |
| 5.0～    | sweet_4 |

## 5.4 ハイパーパラメータの最適化

グリッドサーチを用いて4つの手法の最適なハイパーパラメータを探し、そのモデルの性能を評価する。グリッドサーチは、指定されたハイパーパラメータのリストから値を抽出する。抽出された値の組み合わせごとに、モデルを評価し、最適な値の組み合わせを探索するツールである。scikit-learnのパッケージには、K分割交差検証を行いながら、グリッドサーチを実施するGridSearchCVというツールが備わっており、本実験ではこのツールを用いた。K分割交差検証とは、標本をK群に分割し、K-1群の標本群を学習に用い、その学習モデルに残りの1群を当てはめて評価するというサイクルをK回繰り返す手法である。K回分の結果の平均が最終的な評価となる。本研究では、K=5とした。元のデータセットの70%を訓練データに、30%をテストデータにランダムに分割した後、GridSearchCVクラスをインスタンス化し、訓練データを当てはめる。ここでGridSearchCV内では、交差検証を行いながらハイパーパラメータの探索を行う。グリッドサーチの実施後、最も最適なハイパーパラメータが出力される。最適ハイパーパラメータを適用したモデルに、訓練データから独立したテストデータを用いて、モデルの性能を評価する。モデルの評価軸は、正解率とする。ハイパーパラメータのうち、ランダムフォレストとXGBoostのn\_estimatorsは、それぞれ100, 1,000に固定する。ニューラルネットワークのepochsは、200に固定する。表4は、最適化するハイパーパラメータのリストである。

表4 ハイパーパラメータのリスト

| 機械学習        | ハイパーパラメータ              | 値                                      |
|-------------|------------------------|--|
| k近傍法        | 重み<br>k近傍オブジェクト数       | uniform, distance<br>5, 7, 8, 10, 12   |
| ランダムフォレスト   | max_depth              | 2, 4, 6, 8                             |
| XGBoost     | max_depth              | 2, 4, 6, 8                             |
| ニューラルネットワーク | 2層目のユニット数<br>3層目のユニット数 | 5, 6, 7, 8, 9, 10<br>5, 6, 7, 8, 9, 10 |

## 5.4 実験結果

表5は、各手法で最も正解率が高かったケースの特徴量の組み合わせと最適なハイパーパラメータ、訓練データの予測精度、テストデータの予測精度をまとめた結果である。4つの手法のうち、テストデータに対する予測精度が最も高かったのは、XGBoostであった。現段階の学習器の予測精度は、非常に良いとは言えないが、概ね分類できる精度である。また、それぞれの手法において、特徴量の重要度は、異なっていることが表5よりわかる。今回の結果より、XGBoostが最も精度が高かったが、今後も4つの手法で予測精度を比較し、糖度予測に最適なモデルを選択していく必要があると考えられる。

表5 実験結果

| 手法          | 特徴量                       | 最適なハイパーパラメータ              | 訓練データ予測精度 | テストデータ予測精度 |
|-------------|---------------------------|---------------------------|-----------|------------|
| k近傍法        | 最高気温<br>平均湿度              | 重み：<br>distance<br>k：10   | 73.0%     | 76.9%      |
| ランダムフォレスト   | 平均気温<br>平均湿度              | max_depth：<br>8           | 75.8%     | 76.4%      |
| XGBoost     | 最高気温<br>最低気温              | max_depth：<br>6           | 76.8%     | 78.0%      |
| ニューラルネットワーク | 最高気温<br>最低気温<br>平均二酸化炭素濃度 | 2層目のユニット：10<br>3層目のユニット：8 | 74.6%     | 76.8%      |

## 6. 糖度予測システムの開発

### 6.1 システム概要

機械学習の4つの手法から選択した1つの予測モデルをもとに、Webシステム（以下、提案システムと記す）を開発する予定である。ターゲットユーザは、植物工場で選果作業を行う従業員とする。ミニトマトは、大きさや形の規格から外れ、廃棄扱いされる予定の数十個を対象とする。従業員がスマートフォンからミニトマトの収穫日、大きさを入力すると、予測モデルを用いた糖度分類の計算が行われる。そして、スマートフォンの画面に糖度カテゴリが表示され、ミニトマトを糖度カテゴリごとに分類することで選果作業をする。

### 6.2 システムの評価と廃棄量削減の評価

提案システムの評価には次の2つを予定している。

- (1) ユーザテストとアンケートによる定性的なシステム評価
- (2) 提案システムの予測結果と実測値の誤差を比べる予測精度の定量的評価

ユーザテストでは、植物工場の選果作業を行う従業員数名に提案システムを使用し、糖度別にミニトマトを分類してもらう。使いやすさ、わかりやすさなどについて4段

階で評価してもらう。自由記述も用意し、提案システムに対する感想を広く収集する。予測精度の評価では、提案システムで予測した結果と、非破壊式糖度計を用いて実際に計測した糖度を比較し、どれほど正しく分類できたかを評価する。

また、本研究の目的は、規格外野菜の廃棄量削減であるため、糖度規格と提案システムにより、どれくらいのミニトマトの廃棄量削減を見込めるかの定量的評価を予定している。廃棄予定のミニトマトから数十個を取り出し、廃棄量削減のための実験に使用する。糖度の規格と実験提案システムを用いてこれらのミニトマトの糖度の分類をした結果、何%のミニトマトを廃棄ではなく高糖度な商品として扱えそうかを調査する。

## 7. おわりに

### 7.1 まとめ

本研究の目的は、規格外野菜の廃棄量削減である。目的を達成するための目標は、機械学習を用いた糖度予測モデルの開発である。(1)農作物の外観・環境データの正確かつ効率的な記録、(2)精度の高い糖度予測モデルを開発の2点を研究課題として、2つの課題に対する解決アプローチは、(1)データ記録するためのアプリケーションを開発、(2)機械学習の精度の高い予測モデルを複数開発して比較である。実験では、取得したデータから新たに特徴量を生成、最適なハイパーパラメータの探索を実施した。実験結果として、機械学習の各手法で特徴量の重要度は、異なることがわかった。また、機械学習の4つの手法のうち、XGBoostを用いた予測モデルが最も精度が高く、糖度を概ね分類できる精度であり、研究目的を達成できる見通しを得た。

### 7.2 今後の課題

現段階では、ミニトマトの糖度を概ね分類できる精度であるが、十分とはいえない。今回は、半年分のデータで学習を行ったが、データ数を増やしたり、モデルの改善することで、精度が向上すると考えられる。そのため、今後も引き続きデータ収集、前処理、ハイパーパラメータの見直しを行い、精度の比較をする。また、最も精度のよい学習器を用いて、提案システムを開発する予定である。11月までに提案システムを開発し、12月より植物工場の従業員を対象として実験をし、システムの評価と廃棄量削減の評価をする。

### 謝辞

本研究を進めるにあたり、予測モデルの開発に用いた植物工場内の環境データやミニトマトのデータは、株式会社アプレの協力によるものである。ここに深く感謝の意を表す。

## 参考文献

- [1] “国際農林業協働協会 世界の食料ロスと食料廃棄”.  
[http://www.jaicaf.or.jp/fao/publication/shoseki\\_2011\\_1.pdf](http://www.jaicaf.or.jp/fao/publication/shoseki_2011_1.pdf), (参照 2019-01-06).
- [2] “農林水産省 食品ロスの削減に向けて ~NO-FOODLOSS PROJECT~”.
- [3] “株式会社日本野菜 タダヤサイ 販売商品”.  
<https://www.tadayasai.com/>, (参照 2019-01-30).
- [4] “オイシックス・ラ・大地株式会社 Oisix もったいない野菜”.  
[https://www.oisix.com/Category.mottainai\\_vegetable.htm](https://www.oisix.com/Category.mottainai_vegetable.htm), (参照 2019-01-30).
- [5] 河野俊夫. 外観情報に基づくトマトのゼリー量推定. 植物工場学会誌, 2003, vol. 15, no. 4, p. 195-204.
- [6] “株式会社アタゴ 誰でも解る!糖度計ガイド”.  
[http://www.atago.net/japanese/g\\_fruits.html](http://www.atago.net/japanese/g_fruits.html), (参照 2019-01-30).
- [7] “千代田電子工業 製品の特徴”.  
<https://www.chiyoda-denshi.co.jp/oishika/feature/index.html>, (参照 2019-01-30).
- [8] Elhariri, E. El-Bendary, N. Hassanien, A.E. Badr, A. Hussein, A.M.M. and Snášel, V. Random Forests Based Classification for Crops Ripeness Stages. In: Kömer P., Abraham A., Snášel V. (eds) Proceedings of the Fifth International Conference on Innovations in Bio-Inspired Computing and Applications IBICA 2014, Advances in Intelligent Systems and Computing, vol 303, pp 205-215, Springer, Cham.
- [9] 鈴木真美, 松尾誠治, 梅田大樹, 岩崎泰永. CO<sub>2</sub>施用時の高い相対湿度がキュウリの生育, 光合成速度, 窒素含量に及ぼす影響. 日本冷凍空調学会論文集, 2014, vol. 31, no. 3, p. 331-337.
- [10] 福地信彦, 本居聡子, 宇田川雄二. 摘果及び整枝がトマトの果実糖度と収量に及ぼす影響. 園芸学研究, 2004, vol. 3, no. 3, p. 277-281.
- [11] 熊倉裕史, 穴戸良洋. イチゴの果実肥大に及ぼす温度の影響. 園芸学会雑誌, 1994, vol. 62, no. 4, p. 827-832.
- [12] 吉田圭吾, 高山泰一, 福原弘太郎, 内田敦, 関根秀真, 鹿志村修. スパース正則化を用いたハイパースペクトルデータの解析: インドネシア水田地帯における水稻生育状況診断および収量予測への適用. 日本リモートセンシング学会誌, 2012, vol. 32, no. 5, p. 287-299.
- [13] 寺谷諒, 守屋和幸. 機械学習の手法を用いた自己保全管理農地の発生に関する要因分析と予測モデルの構築 - 京都府綾部市を対象地域として -. システム農学, 2018, vol. 33, no. 4, p. 137-147.
- [14] Yan, W. and Xuelei, S. N.. A XGBOOST RISK MODEL VIA FEATURE SELECTION AND BAYESIAN HYPER-PARAMETER OPTIMIZATION. International Journal of Database Management System, 2019, vol. 11, no. 1, p. 1-17.
- [15] 岡崎篤也. ニューロモーフィックコンピューティングを支えるハードウェア技術の現状と研究動向. 日本ロボット学会誌, 2017, vol. 35, no. 3, p. 209-214.