

機械学習を用いたミニトマトの糖度予測システムの開発

吉田研祐^{†1} 大場みち子^{†2} 高森満^{†3}

概要: 国内では、食の安心・安全についての消費者の関心が高まってきており、無農薬で野菜を栽培可能な植物工場が注目を集めている。植物工場で栽培された野菜は、小売店で「無農薬」という点をアピールして販売されることが多い。しかし、今後、個々の植物工場が利益を伸ばし、固定客を確保するためには、「無農薬」だけでなく「消費者に好まれる要素・ニーズ」を野菜に付加価値を与える必要がある。消費者に好まれる要素・ニーズは多様であり、そのうちのひとつは糖度である。植物工場で栽培された野菜の中から糖度が高い野菜を選び出し、「無農薬かつ高糖度な野菜」として販売することで、消費者の目に留まりやすくなり、また、固定客を増やすことにつながる。糖度別に選果する際に糖度計を用いるが、糖度計には複数の問題がある。本研究では、糖度別の選果を目的とする。研究の対象の野菜をミニトマトとして、糖度予測モデルの作成と糖度予測システムを開発し、予測精度を評価した。

キーワード: 植物工場, 機械学習, ミニトマト, 糖度, 予測

Development of Prediction System for Sugar Content of Cherry Tomatoes Using Machine Learning

KENSUKE YOSHIDA^{†1} MICHIKO OBA^{†2}
MITSURU TAKAMORI^{†3}

Abstract: In Japan, consumers are increasingly interested in food safety, and plant factories that can grow vegetables without using agricultural chemicals are attracting their attentions. Vegetables cultivated in plant factories are often sold at retail stores with an emphasis of “organic”. However, in order for individual plant factories to increase profits and secure regular customers in the future, it is necessary to add not only “organic” but also “needs and elements preferred by customers” to vegetables. Elements and needs favored by customers are diverse, and one of them is sugar content. By selecting vegetables with high sugar content from vegetables cultivated in the plant factory and selling them as “organic and high sugar content vegetables”, customers will be more noticeable. Also, it leads to securing regular customers. When sugar content meter(SCM) is used in order to weigh sugar content, there are several problems by that. The purpose of this research is to classify vegetables by sugar content. Target vegetables of this research are cherry tomatoes. I created sugar content prediction models, developed the sugar content prediction system, and evaluated the prediction accuracy.

Keywords: Plant Factory, Machine Learning, Cherry Tomatoes, Sugar Content, Prediction

1. はじめに

国内では、消費者の食の安心・安全についての関心が高まっている[1]。農薬を使用せずに、安心・安全な野菜を栽培できる点において植物工場が注目を集めており、徐々にその数を増やしている[2]。植物工場は、高度に環境を制御した気密性の高い施設であり、「年間を通して安定的な栽培が可能、無農薬で安心・安全な野菜の生産が可能」という特徴がある[3]。ここで、国内の植物工場での栽培や販売の例を取り上げる。福島県の東西しらかわ農業協同組合では、植物工場に農薬を使用せずに、サンチュやリーフレタスなどを栽培している。栽培した野菜は、地元の小売店と取引されており、「無農薬、柔らかく味も良い」という特徴をアピールして販売している[4]。兵庫県の株式会社兵庫ネクス

トファームでは、希少な品種のミニトマトを植物工場に栽培しており、「果皮の張りの良さ、弾けるような新食感」という特徴をアピールして小売店で販売している[4]。このように、植物工場に栽培された野菜は、業務提携を結んだ店舗と取引され、「無農薬で安心・安全な野菜」であることをキャッチフレーズとして販売されることが一般的である。また、「無農薬」の他に、「味が良い、果皮の張りが良い」など消費者から好まれる要素・ニーズを、アピールポイントとして販売する植物工場も現れている。今後、個々の植物工場が利益を伸ばし、固定客を確保するためには、「無農薬栽培」だけでなく「消費者に好まれる要素・ニーズ」を野菜に付加価値として与えて販売するなど、差別化を図ることが求められる。野菜に対する消費者のニーズは多様であり、甘み、風味、栄養、見た目などである[5]。トマトを例としてあげると、消費者に好まれる要素は、高糖度であり、トマトの品質評価の主指標とも言われている[6]。植物工場に栽培された野菜のうち、高糖度な野菜を「無農薬栽培された安心・安全な野菜、かつ高糖度な野菜」として販売することで、消費者の目に留まりやすくなり、固定客の確保

^{†1} 公立はこだて未来大学大学院
Graduate School of Future University Hakodate

^{†2} 公立はこだて未来大学
Future University Hakodate

^{†3} (株)アプレ
Apure Inc.

につながると考えられる。糖度別に選果する際に糖度計を用いるが、ここで複数の課題がある。一般的に使用されている糖度計は破壊式である[7]。破壊式糖度計は、1台あたり安価であるが、計測する手間が大きく、また果皮に傷を付けて果汁を採取する必要がある。一度果皮に傷を付けた野菜を販売することはできない。破壊式糖度計の弱点を補ったのが非破壊式糖度計である[8]。果皮に傷をつけることなく計測できるが、非常に高額である。また、複数人で野菜を選果することが一般的である。非破壊式糖度計を用いて糖度別に選果する場合、複数台必要となり、多額の出費を伴ってしまう。このように、2種類の糖度計は、糖度別の選果に適していない。

そこで、本研究では、糖度を高精度に予測し、野菜を糖度別に選果できるようにすることを研究目的とする。精度の高い糖度の予測モデルを作成し、作成したモデルを用いて、糖度別の選果に適した糖度予測システムを開発する。

2. 関連研究と課題

2.1 熟度の予測の関連研究

トマトと赤ピーマンの外観画像から、品質の1つである熟度を機械学習を使用して分類する研究がある[9]。色、熟度の品質は、消費者の反応に影響をもたらすため、重要な役割を担っている。この研究では、トマトと赤ピーマンそれぞれ250、175枚の画像を使用している。提案アプローチは、前処理、特徴抽出、分類の3つのフェーズで構成されている。前処理では、画像のリサイズ、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。対象の分類アルゴリズムは、サポートベクターマシンとランダムフォレスト（以下、RF）であり、正解率を基準として、分類精度を評価している。結果として、トマトと赤ピーマンの熟度を最も精度よく分類したアルゴリズムは、線形カーネルサポートベクターマシンであり、正解率90%以上であった。

2.2 環境要素の関連研究

農業の分野では、野菜の品質と環境要素の関係性を調査するための研究が数多く行われている。高い湿度の状態であると、気孔開度が大きくなり、葉内へのCO₂の供給が容易になると言われている。関連研究[10]では、CO₂濃度・湿度の条件が異なる環境を複数用意し、各環境で育てたキュウリの生育や光合成速度、窒素含量に及ぼす影響について調査している。CO₂濃度500ppm程度で湿度86.0%の場合より、CO₂濃度1,000ppm程度で湿度86.0%の場合のほうが、葉面積、光合成速度、葉の窒素含量などが有意に高くなった。葉面積は、湿度が74.5%の場合より、86.0%の場合のほうが有意に増加した。葉面積とトマトの糖度の関連性の研究[11]では、苗全体の葉面積が多いトマトは、摘葉や摘心によって苗全体の葉面積を減らしたトマトより糖度は明らかに高いと述べている。葉面積が増加すると果実内に

転移する光合成産物が増え、糖度が向上すると述べている。

2.3 課題

野菜の外観的要素は、「色情報、大きさ、着色日数」など様々なものが考えられる。関連研究[9]は、野菜の外観画像の色情報を外観的要素として機械学習することで、高精度に熟度を分類することができたと述べている。本研究では、野菜の品質の対象を「糖度」とするため、予測モデルを作成する前に、「糖度」を捉える上で有効な野菜の外観的要素を新たに調査・検証する必要がある。また、環境について述べた研究の結果[10][11]より、「糖度」は様々な環境要素から影響を受けることがわかる。そのため、「糖度」を対象とした場合、外観的要素のみを学習させてもよい精度が期待できないという課題がある。本研究では、外観的要素だけでなく、環境要素も学習させることで課題の解決を図り、より着実に糖度を予測する。

3. 対象施設と対象野菜

3.1 対象施設の概要と対象野菜

水耕栽培を行っている植物工場Aを研究の対象施設とする。植物工場Aの主な特徴は、(1)気温・湿度・CO₂濃度などの環境データの計測と制御、(2)無農薬で多品種の野菜を栽培、である。植物工場Aでは、地元の小売店と取引しており、「無農薬栽培された野菜」として販売している。対象野菜は、ミニトマトMである。植物工場Aでは、ミニトマトMを主に育てており、生育期間が長く、環境要素からの影響を受けやすいため、ミニトマトMを対象とする。

3.2 対象施設の業務内容

植物工場Aでのミニトマトは、「収穫、選果、保管、包装」の4つの工程を経て出荷される。収穫は、毎朝行われる。ミニトマトの苗は、約1,300本であり、苗には播種日が記録されたラベルが貼られている。ミニトマトが赤く熟しているかどうかを作業員が目視で判断する。赤く熟したミニトマトを収穫し、収穫日のラベルが貼付されたコンテナに入れる。収穫作業後、コンテナに入れられた収穫済みのミニトマトは、選果場へと運ばれる。

選果は、収穫作業後に行われる。選果を担当する作業員は、約5名である。選果作業は、野菜の規格に基づき、商品として提供できる野菜と提供できない野菜を分類する作業である。規格は野菜によって様々であるが、主な規格は、傷の有無、色の良し悪しである。選果作業が終了し、規格を満たしたミニトマトは、収穫日のラベルが貼付されたコンテナに格納されて、選果場内の冷蔵庫で保管される。

次に、保管についてである。選果作業後、ミニトマトはすぐに包装されるわけではなく、出荷可能なミニトマトを一定量収集してから、包装作業に移る。そのため、冷蔵庫内には、収穫日が異なるミニトマトが保管される。コンテナは収穫日別に分けられている。選果を終えたミニトマト

の保管期間は、2～3日間程度である。

包装では、出荷可能なミニトマトを一定量収集した後、ミニトマトを出荷用のパッケージに包み、出荷する。選果作業員は、包装作業も担当している。

4. 提案手法

本研究の課題は、「糖度を捉える上で有効な野菜の外観的要素を新たに調査・検証」、「外観的要素に加え、糖度に影響を与える環境要素のデータも学習させて予測モデルを作成」である。提案手法は、(1)糖度と外観的要素の関係性を調査、(2)外観的要素と環境要素の特徴量を生成し、機械学習を使用した高精度な予測モデルを作成、(3)機械学習予測モデルを使用して、糖度予測システムを開発、である。

4.1 糖度と外観的要素の関係性の調査

ミニトマトの外観的要素は、「色情報、大きさ、着色日数」など多様であると考えられる。これらから有力な要素を選択していく。糖度や外観的要素のデータを収集して分析し、有効な要素を糖度の予測モデルの作成時に使用する。

4.2 糖度予測モデルの作成

糖度を捉える上で有効な外観的要素と、環境要素を使用して新たに特徴量を生成する。生成した特徴量を使用して機械学習の予測モデルを複数作成し、各モデルの予測の精度を求める。各モデルの予測精度を比較して、最も精度の良い予測モデルを糖度の予測に最適なモデルと判断する。最も高精度な予測モデルを糖度予測システムで使用する。

4.3 糖度予測システムの開発

本研究では、選果作業員の各自のスマートフォンや PC からアクセスして必要な情報を入力すると、糖度を予測して結果を返す糖度予測システムを開発する。開発した糖度予測システムは、研究の対象施設で実際に使用し、システムの予測精度を評価する。また、本研究が提案する糖度予測システムの使用場面は、図1のように、選果場での包装作業の前段階を想定している。選果を終え、収穫日ごとに分けられたコンテナに保管されているミニトマトを糖度別に分類し、高糖度のミニトマトを選び出す。高糖度のミニトマトとして包装し、出荷するというフローを想定している。これにより、ミニトマトに「無農薬、かつ高糖度」という付加価値をつけた販売が可能となる。無農薬栽培された野菜を求める消費者や、無農薬栽培された野菜の中でも、高糖度のような付加価値の高い野菜を求める消費者に向けた販売を見込むことができる。商品規格を満たしたミニトマトの中から総合的に高糖度なミニトマトを選ぶこととなるため、誤って規格外のミニトマトを選んで包装することは防げる。

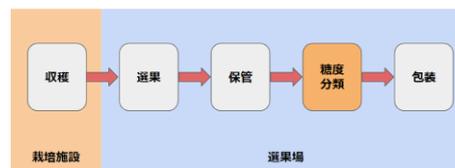


図1 想定する糖度分類の場面

5. 予備調査

予備調査の目的は、糖度を捉える上で、有効な外観的要素を調査・検証することである。糖度との関連性が高い外観的要素は、糖度の予測モデル作成時に使用する。

5.1 対象とする外観的要素

ミニトマトの外観的要素は「色情報、大きさ、着色日数」などが考えられるが、本研究では、「大きさ、着色日数」に着目する。理由を以下に示す。本研究では、包装の前段階のミニトマトを糖度別に選果することを想定している。一般的に選果作業の時点で、赤色でないミニトマトは、規格外野菜として廃棄される。そのため、図2のように、赤色のミニトマトのみが冷蔵庫内で保管され、包装される。



図2 選果後のミニトマトの例

選果後のミニトマトの色は、ほぼ均一の赤色であり、肉眼で微量な色の差異を判別することは困難であるため、色情報を外観的要素として用いないことにした。一方で、植物工場Aでのミニトマトの規格は、色の良し悪しと傷の有無であり、大きさに関しての規格はない。そのため、選果後の一つひとつのミニトマトの大きさにはばらつきがある。ミニトマトの大きさは、定規を使って数値として容易に表すことができる。これらのことより、ミニトマトの外観的要素として、大きさを対象とする。次に、着色日数を選んだ理由を述べる。着果を終えた緑色のミニトマトは、徐々に赤色に着色していき、十分に熟したところで収穫される。一般的には、着色を開始して収穫を迎えるまでの間に、糖度が変化する。本研究では、着色開始から収穫までの期間の環境要素が糖度に影響を与えるという仮説を立てる。そのため、着色開始から収穫までに要する日数を調査する必要があり、着色日数を外観的要素の対象とする。

5.2 糖度と大きさの関係性の調査

ミニトマトの糖度を捉える上で、大きさが有効であるかどうかを検証する方法として、相関分析をする。相関分析を実施するために、糖度データと大きさデータの収集が必要である。糖度データは、非破壊式糖度計[8]を使用して計測する。図3のように、定規を使用してミニトマトの縦

の長さや横の長さを計測して、縦の長さや横の長さを掛け合わせて算出した面積を大きさデータとする。1 ミリ単位で縦の長さや横の長さを計測する。



図 3 ミニトマトの大きさの計測

計測したデータを記録するため、Android アプリケーション（以下、アプリ）を開発する。このアプリには、収穫日、糖度データ、縦の長さ、横の長さを入力する。入力したデータは、Firebase の DB に JSON 形式で保存する。2018 年 10 月～12 月に栽培したミニトマト 290 件分の糖度データと大きさデータを予備調査で使用する。

糖度と大きさの 2 変数間の相関関係を調査するにあたり、シャピロ・ウィルク検定を実施して、糖度と大きさの 2 変数の正規性を検証する。正規性を検証する理由は、相関係数の検定方法は、母集団が正規分布に従う場合と従わない場合で異なるためである。帰無仮説を「標本は正規分布に従う」とし、有意水準は 5% とする。相関係数の検定について述べる。ピアソンの積率相関係数の検定は、母集団が正規分布に従うデータ同士の相関を、スピアマンの順位相関係数の検定は、母集団が正規分布に従わないデータ同士の相関を調査する方法である。帰無仮説を、「2 変数間の真の相関は 0 に等しい」とし、有意水準は 5% とする。どちらの検定方法を使用するかは、シャピロ・ウィルク検定の結果をもとに判断する。

5.3 着色日数の調査

ミニトマトの着色日数を調査するために、2019 年 1 月～2 月に栽培施設内で生育中のミニトマトを使用する。栽培施設内で生育中のミニトマトの複数の苗を定点撮影し、タイムラプス動画を作成する。作成した動画からミニトマトの着色していく様子を捉え、一つひとつのミニトマト着色日数を数える。

5.4 結果

糖度と大きさの相関を調査した結果を述べる。シャピロ・ウィルク検定を実施し、糖度と大きさの 2 変数間の分布の結果を表 1 に示す。

表 1 正規性の結果

変数	p 値
糖度	9.62e-07
大きさ	8.41e-14

シャピロ・ウィルク検定の結果、糖度と大きさの両者とも、p 値が有意水準 5% を下回り、対立仮説が採択された。

すなわち、両者は正規分布に従わないと判断した。次に、ノンパラメトリックな手法であるスピアマンの順位相関係数の検定を実施した。表 2 に結果を示す。

表 2 相関分析の結果

スピアマンの順位相関係数	p 値
-0.70	9.22e-44

スピアマンの順位相関係数の検定の結果、p 値は有意水準 5% を下回り、対立仮説が採択された。すなわち、2 変数間には相関があると言える。また、スピアマンの順位相関係数は -0.70 となり、糖度と大きさの 2 変数間には、負の強い相関があることが分かった。

次に、着色日数の調査結果について述べる。ミニトマトが着色していく様子をタイムラプス動画から確認できたのは、45 個であった。ミニトマトの着色日数の結果は、図 4 に示す。ミニトマトは 5～8 日間かけて着色していき、着色日数の最頻値は、7 日間であった。

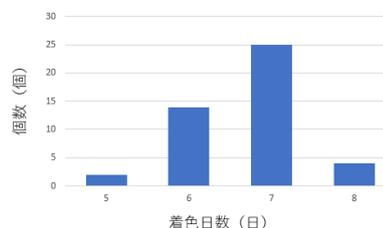


図 4 着色日数の結果

5.5 考察

明らかとなった点は 2 つある。まず 1 つは、ミニトマトの外観的要素である大きさは、糖度と強い負の相関がみられることである。ミニトマトの大きさが小さいほど糖度が高くなる傾向にあることがわかった。機械学習を使用して糖度の予測モデルを作成する際、モデルの属性に大きさを入れることで、正確に糖度を予測できる見込みがある。2 つ目は、ミニトマトの着色に要する日数は、7 日間が最も多いことである。着色するにつれて、糖度が変わると一般的に言われており、本研究ではその期間の環境要素が糖度に影響していると考えられる。図 5 に例を示す。例えば、収穫日が 9 月 9 日である場合、9 月 2～8 日までの 7 日間の環境データを使用して糖度を予測する。収穫日が 9 月 20 日である場合、9 月 13～19 日までの 7 日間の環境データを使用して、糖度を予測する。糖度予測システムは、図 1 に示すように、包装前の段階での使用を想定しており、ミニトマトは収穫日が記載されたコンテナに格納されている。コンテナに記載されている収穫日から 7 日間前までの環境データを使用し、機械学習することで、糖度を正確に捉えることができると考えた。

September 2019						
SUN	MON	TUE	WED	THU	FRI	SAT
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	...					

図 5 収穫日より 7 日間前までの環境データを使用

6. 予備実験

予備実験の目的は、高精度な糖度の予測モデルを作成することである。予備調査の結果を踏まえ、予測モデルの作成を進める。

6.1 対象データと目的変数

予備実験では、環境データ、大きさデータ、糖度データを使用する。植物工場 A では、栽培施設内の気温・湿度・CO₂濃度を自動的に計測し、DB に記録している。気温・湿度・CO₂濃度の日別の平均データを植物工場 A からいただき、これらを環境データとして使用することにする。2018 年 11 月～2019 年 9 月の間に生育したミニトマト 3,192 件分の大きさデータ、糖度データ、日別の環境データ（平均気温、平均湿度、平均 CO₂濃度）を対象とする。3,192 件の糖度は 3.0～6.0%の間に分布している。これらの糖度のデータを 4 つのクラスに分け、表 3 に示すように糖度クラスを目的変数として予測する。

表 3 目的変数

糖度(%)	糖度クラス
～4.0	クラス 1
4.1～4.5	クラス 2
4.6～5.0	クラス 3
5.1～	クラス 4

6.2 機械学習モデル

糖度クラスは機械学習を使用して予測する。本研究では、3 つの機械学習アルゴリズムを使用して予測モデルを作成する。対象の機械学習アルゴリズムは、K 近傍法（以下、KNN）、RF、XGBoost である。KNN は、対象のデータと類似度が最も高い K 個の標本を抽出し、多数決でクラスラベルを求めるアルゴリズムである[12]。RF は、アンサンブル学習の 1 つであり、決定木を複数作成し、それぞれの決定木モデルで予測した値から多数決で最終的な予測値を求める[13]。XGBoost は、RF と同様、アンサンブル学習のアルゴリズムの 1 つであり、勾配ブースティングと呼ばれるアルゴリズムを採用している[14]。また、機械学習の競技会の Kaggle で優れた結果を残したとの報告もある。グリッドサーチを使用して最適なハイパーパラメータを探索し、予測モデルを作成する。各予測モデルの精度を正解率を基に評価する。最も高精度な予測モデルを糖度の予測に最適とする。

6.3 特徴量の生成

予備調査の結果・考察で述べたように、ミニトマトの「大きさ、着色日数」は糖度を捉える上で重要な外観的要素としている。ミニトマトの大きさは、ミニトマトの糖度と高い相関関係にあり、大きさに着目することで糖度を正確に捉えることができると考える。本研究では、機械学習の特徴量に必ず大きさデータを取り入れることにする。また、着色期間の環境要素が糖度の向上に影響を与えると仮定しており、その着色に要する日数は 7 日間であった。そこで、表 4 のように、収穫日前の着色期間の 7 日間の環境データの特徴量を新たに生成する。図 5 を使用して例を説明すると、9 月 9 日に収穫したミニトマトは、着色期間が 9 月 2～8 日までの 7 日間であるとして、この期間の環境データの平均値を生成する。9 月 20 日に収穫したミニトマトは、着色期間が 9 月 13～19 日までの 7 日間であるとして、この期間の環境データの平均値を生成する。気温と湿度のデータから飽差という新たな要素を生成する。飽差とは、空気 1 立法メートル当たりの水蒸気の空き容量を示し、気温と湿度から求めることができる。

表 4 特徴量の生成

収集したデータ	新しく生成した特徴量
日平均気温	7 日間の平均気温
日平均湿度	7 日間の平均湿度
日平均 CO ₂ 濃度	7 日間の平均 CO ₂ 濃度
日平均気温・湿度	7 日間の平均飽差

6.4 結果

機械学習 3 手法において、大きさデータ、収穫日から 7 日間前までの環境データ（気温、湿度、CO₂濃度、飽差）の特徴量の組み合わせ 16 通りを試して、訓練セットとテストセットに対する正解率を求めた。表 5 に、各機械学習手法で最も正解率が高かったモデル、特徴量の組み合わせ、最適なハイパーパラメータ、訓練セットとテストセットに対する予測精度をまとめた。

表 5 予測精度の比較結果

機械学習	特徴量	最適なハイパーパラメータ	予測精度 訓練(上) テスト(下)
KNN	大きさ・気温・湿度・CO ₂ 濃度	n_neighbors:6	83.5% 83.2%
RF	大きさ・気温・CO ₂ 濃度	max_depth:5 min_sample_leaf:1 min_sample_split:3 n_estimator:50	83.0% 82.6%
XGBoost	大きさ・気温・湿度・CO ₂ 濃度	max_depth:5 n_estimator:200 subsample:0.5	85.5% 84.8%

表 6 環境データを使用・未使用の場合の精度の比較結果

機械学習	特徴量	最適なハイパーパラメータ	予測精度 訓練(上) テスト(下)
XGBoost	大きさ	max_depth:2 n_estimator:100 subsample:0.7	81.7% 81.2%
XGBoost	大きさ・気温・湿度・CO ₂ 濃度	max_depth:5 n_estimator:200 subsample:0.5	85.5% 84.8%

表 6 からわかるように、大きさデータのみを特徴量に使用した予測モデルは、大きさデータ、気温、湿度、CO₂濃度を特徴量として使用した予測モデルよりも予測精度が明らかに低かった。

6.5 考察

KNN, RF, XGBoost のそれぞれの予測精度は、正解率が 8 割程度であったが、中でも XGBoost が最も精度よく予測できた。XGBoost の予測精度が最も高い理由として、本研究で対象としたデータの特徴が関係していると考えられる。本研究で対象とした大きさデータや環境データは、それぞれ単位やスケールがそれぞれ異なっており、複雑なデータの集合体である。しかし、植物工場という比較的安定した環境であるため、それぞれのデータが驚異的に変化することや、異常値を取ることは少ない。そのため、ある決定木の予測誤差を、新たな決定木が学習するというサイクルを逐次的に行う勾配ブースティングのアルゴリズムを採用することで、自然と誤差を収束することができ、結果的に XGBoost の手法が最も精度が高くなったと考えられる。また、特徴量に飽差を使用した場合、精度向上が見込めなかった。飽差は、気温と湿度から算出され、ある気温と湿度の空気中に、水蒸気をどれだけ含むことができるかの指標であり、気温と湿度を集約した要素である。飽差を使用することで、気温や湿度の独自の特性が薄れてしまい、糖度を予測する上での重要性が低くなったと考えられる。一方で、特徴量に環境要素を 1 つも入れずに大きさのみで糖度を予測したとき、最も精度が低かった。提案手法で述べた「外観的要素のみだけでなく、環境要素を考慮することで、糖度を正確に捉えることができる」という仮説が正しいと判断した。また、収穫日前の 7 日間の環境データを使用して、学習を行った結果、どの予測モデルも精度が相応に高かった。この結果より、着色期間の環境要素が糖度に影響を及ぼしているという仮説も、概ね正しいと判断した。今回は、XGBoost を用いて「ミニトマトの大きさ、気温、湿度、CO₂濃度」を用いたときに、最も精度が高かったが、さらにデータ数を増やし、生育に影響を与える環境要素を

より多様的に考慮することで、精度を向上させることができると考えられる。

7. 糖度予測システムの開発

予備実験で作成した予測モデルを使用して、糖度予測システムを開発した。システムの構成図は、図 6 に示す。システムのユーザは、植物工場 A の選果作業員を想定している。実際のシステムの流れについて以下に述べる。各選果作業員のスマートフォンや PC からインターネットを経由して、糖度予測システムにアクセスする。「ミニトマトの収穫日、大きさデータ（縦の長さ、横の長さ）」を入力する。入力された収穫日のデータをもとに、収穫日から 7 日間前までの環境データ（平均気温、平均湿度、平均 CO₂濃度）を DB から取得する。DB の環境テーブルから取得した環境データと、入力した大きさデータを、XGBoost の予測モデルに当てはめる。予測モデルの予測結果は、「sweet 1 : 糖度 4.0%以下です, sweet 2 : 糖度 4.1~4.5%です, sweet 3 : 糖度 4.6~5.0%です, sweet 4 : 糖度 5.1%以上です」のいずれかを返し、システム画面に表示する。各選果作業員が入力した「収穫日、縦の長さ、横の長さ」のデータと予測モデルによる予測結果は、DB の予測結果テーブルに格納される。

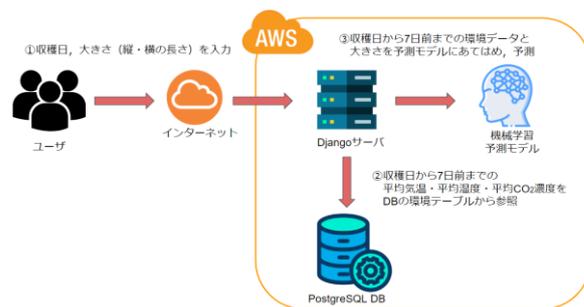


図 6 システム構成図

8. 本実験

本実験の目的は、開発した糖度予測システムの予測精度を定量的に評価することである。糖度予測システムの予測精度を正確に検証するために、複数回にわたって実験を実施する。複数回の実験の結果から、糖度予測システムの予測精度を総合的に評価する。

8.1 対象ユーザと対象データ

実験は、植物工場 A 内で行い、選果作業員 5 名に協力していただく。実験の実施日は、2019 年 12 月 15 日、18 日、21 日、2020 年 1 月 8 日であり、合計で 4 回実施する。1 回の実験で、選果作業員 2, 3 名を対象とする。それぞれの実験では、収穫日の異なるミニトマトを使用する。先にも述べた通り、ミニトマトは冷蔵庫で収穫日ごとに分けられたコンテナに保管されている。12 月 15 日実施の 1 回目の実験では、12 月 13 日、14 日、15 日に収穫したミニトマト（合計 80 件）を、12 月 18 日実施の 2 回目の実験では、12 月

17日、18日に収穫したミニトマト（合計52件）を、12月21日実施の3回目の実験では、12月19日、20日、21日に収穫したミニトマト（76件）を使用する。1月8日実施の4回目の実験では、1月6日、7日に収穫したミニトマト（60件）を使用する。

8.2 実験の進め方

糖度予測システムの予測精度を検証するために、非破壊式糖度計[8]を使用する。実験の流れは、次の(1)~(6)であり、この流れを繰り返し実施する。

- (1)選果作業員が各自のスマートフォンやPCから糖度予測システムにアクセスする。
- (2)ミニトマトが入ったコンテナから、ランダムにミニトマトを取り出す。
- (3)取り出したミニトマトの縦の長さ、横の長さを定規を使用して計測する。
- (4)糖度予測システムの入力フォームに、ミニトマトの収穫日、ミニトマトの縦の長さ、横の長さのデータを入れて、糖度を予測する。
- (5)非破壊式糖度計を使用してそのミニトマトの糖度を計測する。
- (6)非破壊式糖度計で計測した糖度データを実測値として、実測値と予測値の2つを比較する。

開発した糖度予測システムは、「sweet 3：糖度 4.6~5.0%です」のような予測の結果を返す。一方で、非破壊式糖度計の計測結果は、「糖度 4.7%」のように数字として返される。糖度予測システムが予測した糖度の範囲に、非破壊式糖度計で計測した値が含まれる場合、正しく予測できたことにする。一方で、糖度予測システムが予測した糖度の範囲から、非破壊式糖度計で計測した値が外れていた場合、正しく予測できなかったことにする。

8.3 結果

1回目から4回目の実験までの結果を、表7に混合行列としてまとめて正解率を求めたところ、4回の実験の平均の正解率は、82.0%であった。

表7 混合行列

	sweet_1	sweet_2	sweet_3	sweet_4
sweet_1	51	8	0	0
sweet_2	10	56	7	0
sweet_3	0	5	59	11
sweet_4	0	0	7	54

8.4 考察

糖度予測システムの予測精度は、正解率 82.0%であり、これは大まかに糖度別に分類することができていると言える。より実用的にしていけるためには、予測精度の向上が求められる。さらに予測精度を高めるための方法として、ミニトマトの生育位置に着目することが有効であると考えられる。図7に、上から見た栽培施設のイメージを示す。

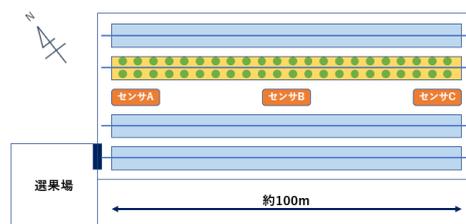


図7 上から見た栽培施設のイメージ

栽培施設は横に長く、全長で約100mある。黄色の矩形がミニトマトのポッドであり、緑色の丸がミニトマトの苗である。中央と両端には環境データを計測するためのセンサーが備えられており、10分ごとに計測している。栽培施設内の中央位置の環境データを着目することで、施設内のだいたいの環境データを把握できると考え、中央のセンサBの環境データを対象として実験を進めてきた。図8、9は、図7のセンサA、B、Cで2019年12月1日~12月25日に計測した気温、湿度のデータの日別の平均値を示している。青色の線はセンサAで計測したデータ、褐色の線はセンサBで計測したデータ、灰色の線はセンサCで計測したデータである。



図8 センサA、B、Cで計測した気温の変化



図9 センサA、B、Cで計測した湿度の変化

気温に関しては、センサA、Bにおいて、似たような値を取っているが、センサCは、比較的低い値が計測されている。植物工場Aの屋外に設置されている風向計によると、12月は東側から風が吹く傾向にあった。つまり、センサC側の外側の壁が冬季の風によって冷やされ、植物工場内のセンサC付近の空気の気温が低くなったと考えられる。

湿度に関しては、3つのセンサーで計測したデータは、似たような傾向が見られることがわかるが、センサAでは湿度が低くなる箇所がいくつか見られる。センサAの近くには、選果場と結ぶ通路があり、ドアで仕切られている。栽培施設に入るためには、この通路を通過し、ドアを開け閉

めする必要がある。作業員の出入りの頻度が非常に高いため、ドア付近では湿度は低くなり、センサ A で計測した湿度データが低くなったと考えられる。

このように、高度な環境制御を行っている植物工場においても、位置によって環境データの違いが見られることがわかる。位置ごとに環境データが異なるため、生育する場所によって環境要素がミニトマトの品質に与える影響の違いが出てくると考えられる。植物工場 A の現状は、収穫したミニトマトは、苗の生育位置を考慮せずに、1 つのコンテナに格納されて、選果場へ運ばれる。そのため、糖度予測システムを使用して糖度別に分類するときには、どの生育位置の苗から収穫したかを知ることができない。生育位置を明確に区別することは難しいが、「センサ A 付近、B 付近、C 付近」のようにコンテナを複数用意し、収穫した位置に該当するコンテナにミニトマトを入れることで、大まかに苗の位置情報を考慮することができる。このように苗の位置情報を加味し、中央位置のセンサ B で計測したデータだけでなく、センサ A、C などの異なる位置のセンサデータも機械学習させることで、より着実にミニトマトの糖度を捉えることができ、予測精度の向上を見込むことができると考える。

9. まとめ

本研究の目的は、植物工場で栽培された野菜の糖度を高精度に予測し、糖度別に選果できるようにすることである。研究目的を達成するために、高精度な予測モデルの作成と、糖度の予測システムの開発を目指した。ミニトマトを研究の対象野菜とし、(1)ミニトマトの糖度と外観的要素の関係を調査、(2)外観的要素と環境要素を使用して、機械学習の糖度予測モデルを複数作成し、予測精度を比較、(3)糖度予測モデルを使用して糖度の予測システムを開発、をアプローチとした。調査より、ミニトマトの大きさや糖度には高い相関があること、着色日数は7日間であることがわかり、これらの結果をもとに機械学習に使用する特徴量を生成した。複数の予測モデルを作成し、予測精度を比較したところ、大きさ、収穫日から7日間前までの気温、湿度、CO₂濃度を学習した XGBoost が最も精度が高かった。作成した予測モデルを使用した糖度予測システムを開発し、現場で予測精度を検証したところ、82.0%であった。中央付近の環境データを機械学習に使用して予測をしたが、同じ空間においても場所によって環境データが異なることがわかった。場所によって環境データがミニトマトの品質に与える影響度も異なると思われる。精度の向上を目標として、ミニトマトの生育位置のデータと生育位置付近の環境データを新たに機械学習に取り入れて予測し、精度を検証していくことが今後の課題である。

謝辞 本研究で使用したミニトマトや植物工場の環境データは、株式会社アプレの協力によるものである。ここに感謝の意を表す。

参考文献

- [1] 伊東幸恵, 大浦裕二, 清野誠喜. 農産物の残留農薬に対する消費者意識. 農林業問題研究, 2005, vol. 41, no. 1, p. 150-152.
- [2] 芦田恵樹, 清水智美, 建山和由. 太陽光併用型植物工場における光の効率的な利用システムの検討. 植物環境工学, 2007, vol. 19, no. 2, p. 59-65.
- [3] 岡田英博, 多田誠人, 坂井義明. 太陽光利用型植物工場における自動化の必要性について. 植物環境工学, 2011, vol. 23, no. 2, p. 44-51.
- [4] “一般社団法人日本施設園芸協会 平成 27 年度 次世代施設園芸導入加速化支援事業 (全国推進事業) 事業報告書 別冊 2 大規模施設園芸・植物工場実態調査・事例調査”. <https://www.maff.go.jp/j/seisan/ryutu/engei/sisetsu/pdf/daikibo1.pdf>, (参照 2020-01-05)
- [5] “財団法人愛知県農業振興基金尾張地域における野菜の新しい地域流通「消費者ニーズに対応する都市近郊農業の生産振興策」報告書”. <http://www.aichinoshinki.or.jp/chousa/H13kinkou.pdf>, (参照 2019-11-11).
- [6] 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.
- [7] “株式会社アタゴ手持屈折計”. <https://www.atago.net/product/?l=ja&f=new/productsmaster-top.php#AHG52038>, (参照 2019-11-11).
- [8] “千代田電子工業製品の特徴”. <https://www.chiyoda-denshi.co.jp/oishika/feature/index.html>, (参照 2019-11-11).
- [9] Elhariri E., El-Bendary N., Hassanien A.E., Badr A., Hussein A.M.M., and Snasel V. Random Forests Based Classification for Crops Ripeness Stages. Proceedings of the Fifth International Conference on Innovations in Bio-Inspired Computing and Applications IBICA 2014. Advances in Intelligent Systems and Computing, 2014, vol. 303, p. 205-215.
- [10] 鈴木真美, 松尾誠治, 梅田大樹, 岩崎泰永. CO₂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.
- [11] 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.
- [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パラメータデータの解析: インドネシア水田地域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.