

## トマトハウス栽培を対象とした機械学習による収量予測モデルの開発

横田悠<sup>†</sup> 南野謙一<sup>†</sup>岩手県立大学ソフトウェア情報学部<sup>†</sup>

## 1. はじめに

近年、センサーで計測して集めたビッグデータを解析し、収量を予測することで、効率的に栽培管理するスマート農業が進められている。ハウス野菜栽培では収穫時に必要とされる人員の調整や、見込まれる収量に基づいた販売計画作成などの観点より、収量予測への需要が高まっており、センサーを設置し収集したデータから機械学習[1]により収量予測モデルの開発が行われている。しかしながら、気温や日射量、土壌の性質など多くの外的要因により定まる農作物の収量を予測するには、センサーデータの選別、データの事前処理などを行い、対象圃場に適合させなければ精度の高いモデル開発はできない。そこで、本研究では、岩手県内でトマトハウス栽培をしている企業との共同研究による、実用的なトマトの収量予測モデルの開発手法について提案する。

## 2. 機械学習のプロセス

機械学習には、岩手県に拠点を置く A 社、B 社のデータを用い、教師あり学習（重回帰）により収量予測を行う。本研究におけるモデル開発からその評価までのプロセスを図 1 に示す。

はじめに、各社所有ハウス内に設置されているセンサーより得られた室温、日射量の環境データ、収量データ（本研究では、A 社はミニトマトの、B 社は大玉トマトの収量予測モデルを開発する）を記録する。記録されたデータはそのまま機械学習用データとしては用いず、図 1 のプロセスを通して適したデータに加工、変換（データの事前処理）する。データセットが準備できたら、説明変数に環境データ、目的変数に収量データをセットし、重回帰モデルを開発する。

収量予測モデルは、訓練データにより学習を行い、訓練データとは異なるテストデータにより評価を行う。評価指標は決定係数、予測値と実測値の誤差とし、これらの結果を基にモデル評価を行い、モデル開発に用いるデータ選別等の改善をすることで、より精度の高いモデルの開発を目指す。

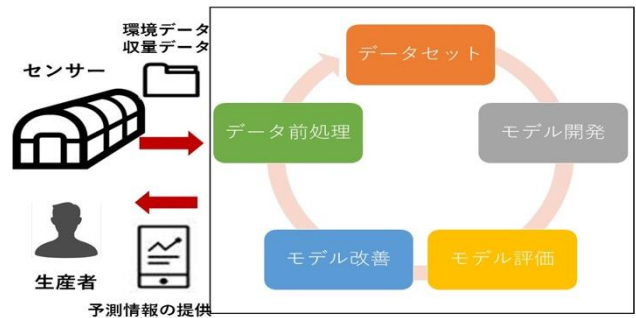


図 1 機械学習によるモデル開発プロセス

## 3. 収量予測モデルの開発手法

## 3.1 説明変数

説明変数はハウス内の日射量、日平均気温を加工したものをを用いる。日射量は一定期間積算した値とする。積算期間は、1 週刻みの積算日射量と収量の相関を調査した結果、最も相関が高くなった収穫日より 4 週間積算した値とする。日平均気温は一定期間の温度分布を表した H 関数、または U 関数のどちらかを用いる。H 関数は、定植日から収穫開始日までの日平均気温を対象とし、1 度刻みの度数分布表を作成して気温ごとの相対度数を求め、その値を用いる。しかしながら今回提供いただいたデータには定植日や収穫開始日などに該当する項目がなかった。そこで、H 関数を導入した先行研究[2]を参考にしたところ、定植日から収穫開始日までの期間は 7 週から 9 週となることが多かったため、本研究では間をとって 8 週を採用した。U 関数は、相対度数（H 関数）ではなく累計相対度数を用いたものである。

## 3.2 目的変数

目的変数には収量データを用いる。しかしながら、計測された収量は出荷量調整を行なった場合も含めたものであり、トマトが生育した環境本来の収量とは異なる。出荷量調整は収穫の前後 1 週間にのみ影響を与えるため[3]、B 社の予測では既存の収量データより 14 日移動平均を求めた値を用いる。A 社のミニトマトの収量は、日別では少量で変動が大きかったため、移動平均に加えて、収穫日を起点としその日までの 1 週間の収量を積算した値を目的変数とする。これにより、1 週間先の収穫時に必要とされる人



図2 A社の訓練データ(2019年)と重回帰モデル



図3 B社の訓練データ(2019年)と重回帰モデル  
員の調整や、見込まれる収量に基づいた販売計画作成などに役立つようにする。

## 4. モデル評価

### 4.1 決定係数による評価

2019年1月13日から2019年7月31日のA社の訓練データで学習させた重回帰モデルと収量データ(訓練データ)との適合を示す(図2)。このとき、説明変数にはH関数を用いた。その結果、決定係数は0.84となり、一年を通して大きなずれを生じさせず再現ができています。また、2019年1月13日から2019年7月31日のB社の訓練データでも同じように重回帰モデルと収量データ(訓練データ)との適合を示す(図3)。こちらも訓練データを高精度で再現できており、決定係数は0.91となった。さらに、A社の2019年データから予測日の前3か月分のデータを用いて訓練データとテストデータを分けた場合の検証も行った(図4)。このときの決定係数は予測日の翌日が0.75、2日後で0.60となり予測日の3日先までは実測値を大きく外すことのない予測ができた。また、予測日から7日先までの積算収量を予測し、実測値と比較したところ決定係数0.86を得た。

### 4.2 考察

重回帰モデルの決定係数が高いことより、積算日射量、H関数を用いることで訓練データの収量を高い

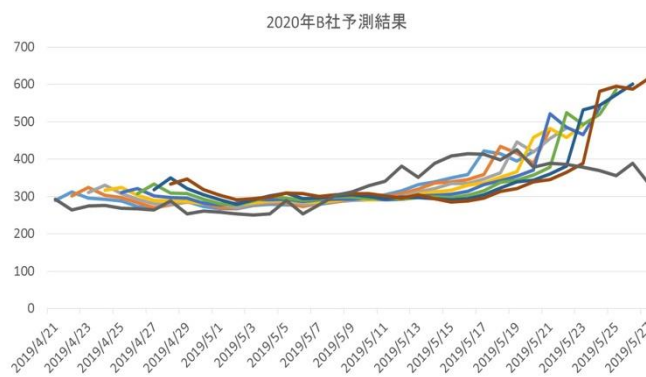


図4 B社のテストデータ(2020年)と重回帰モデル

精度で再現することができた。したがって、ハウスにおいて適切に温度管理などを行い、年が変わっても環境データに大きな変化がないようにすれば、毎年の収量が安定するため、過去の訓練データで開発したモデルを用いても精度の高い収量予測ができると考えられる。しかしながら、訓練データに用いた2018年から2019年にかけてのデータで開発した重回帰モデルでは、2020年の収量よりも多くの収量を予測してしまった。この原因としてA社、B社の収量データが年毎に安定していないこと、データ数の少なさにより、開発したモデルが訓練データに依存してしまうことが考えられる。これでは、1年分の収量と環境データを学習させても、翌年の収量は大幅に異なるため、過学習が起き、予測は困難になってしまう。

また、当日の前3ヶ月分のデータを学習させた重回帰モデル(図4)では、予測日から7日先までの積算収量であれば高い精度で再現することができた。

## 5. まとめ

本研究では岩手県内企業のハウストマト栽培の収量を対象に、機械学習よりモデルを開発し、その有用性について考察した。今後の課題として、学習データの前処理改善や、機械学習時の説明変数にどのデータを用いるかの見直しをし、未来の収量を高精度で予測できるモデルを開発することがあげられる。また、機械学習でトマトの収量を予測するため、さらに多くの学習データを用意する必要もある。

## 参考文献

- [1] 毛利拓也・北川廣野・谷一徳- scikit-learn データ分析実装ハンドブック”2019
- [2] 遠藤隆也・割澤伸一・山田一郎  
“トマト低段密植栽培における収穫日・収量の予測手法(計算モデル)の検討, 日本生物環境工学会:14-15.2017
- [3] 久枝和昇・仁科弘重  
“大規模トマト生産温室における生産性向上に関する研究—積算日射量に基づいたトマトの出荷量予測” 植物環境工学(J.SHITA):11-18.2007