

食用ほおずきを対象とした積算日射量と温度分布を用いた収量予測モデルの開発

今田寛也^{†1} 南野謙一^{†1}

岩手県立大学 ソフトウェア情報学部^{†1}

1. はじめに

近年、スマート農業が推進され、AI 技術によりセンサーで収集された環境データをもとに収量予測モデルを開発し、出荷量確保・雇用労働力確保に役立つ試みが行われている。岩手県では 2005 年頃から岩泉市を中心に食用ほおずきの栽培に取り組んでおりブランド化に成功している。しかし、全国的に栽培地域は少なく、収量予測モデルの開発はあまり行われていない。食用ほおずきの栽培方法は同じナス科であるトマトと似ており、種から植えて苗になったら畑に移し成長させる。

そこで本研究では、先行研究で開発したトマトの収量予測モデル[1]をもとに改良を加え、食用ほおずきの収量予測モデルを開発する。機械学習を用いて訓練データにより収量予測モデルを開発し、テストデータを用いた検証を行うことで、モデルの有効性を明らかにする。

2. 収量予測モデルの開発手法

2.1 使用するデータ

本研究では、(株)ネクスグループの食用ほおずき(ハウス栽培)の収量・環境データを利用し、重回帰モデルによって収量予測を行う。重回帰モデルとは、1つの目的変数を複数の説明変数を使用して予測を行う手法である。本研究では、説明変数に環境データの日射量と気温、目的変数には、収量データを使用する。また、ネクスグループでは A 棟と I 棟の2つのハウスで栽培をしており、それぞれのデータを利用して2つのモデルを開発する。データ収集期間は、環境データが2020年12月1日から2021年8月31日、収量データが2021年2月1日から2021年8月31日である。環境データをランダムに訓練データ、テストデータの2つに分け、訓練データを使って学習させてモデルを開発し、テストデータを利用してモデル評価を行う。

2.2 説明変数

先行研究[1]を参考にし、説明変数をハウス内の日射量、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、最も相関が

表1 説明変数の候補と評価

A 棟				
H 関数	49 日間	49 日間	42 日間	42 日間
日射量積算	4 週間	3 週間	4 週間	3 週間
決定係数	0.80	0.79	0.74	0.73
誤差平均値	1.54	1.55	1.78	1.81
I 棟				
H 関数	49 日間	49 日間	42 日間	42 日間
日射量積算	4 週間	3 週間	4 週間	3 週間
決定係数	0.92	0.92	0.94	0.94
誤差平均値	1.37	1.39	1.17	1.2

高くなる期間を採用する。日平均気温は一定期間の温度分布を現した Histogram 関数(以下、H 関数)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。今回は 10°C から 31°C までの気温の分布を説明変数とする。提供されたデータには、定植日のデータがなかったため、先行研究[1]で採用されていた 8 週間から、食用ほおずきの栽培で多い 6 週間までを調査し、こちらも同様に最も相関が高くなる期間を採用する。

2.3 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、計測された収量データは、収穫しなかった日や、生産調整を行った日も含めたものであり、個々の日のデータが正確とはいえないため、7日間移動平均を求めた値を用いた。しかし、7日間移動平均においても、収量が 0kg になってしまう期間が存在しており、正確に予測することが難しかった。そのため、7日間から範囲を広げた結果を調査し、最も相関が高くなった 14 日間移動平均を採用した。また、移動平均は収穫日を起点とし、前後 7 日間の計 14 日間とした。

3. モデル評価

3.1 A 棟のモデル評価

A 棟のモデルにおける相関が最も高くなる説明変数について調べた結果、H 関数では 49 日間(7 週)、日射量積算では 4 週間の時が最も決定係数が高くなり、0.796 となった(表 1)。また、誤差の平均値も 1.54 となり、グラフの精度も最も高くなった。

Development of a Yield Prediction Model Using Cumulative Solar Radiation and Temperature Distribution for Edible Hozuki Plants

HIROYA IMADA[†], KEN-ICHI MINAMINO[†]

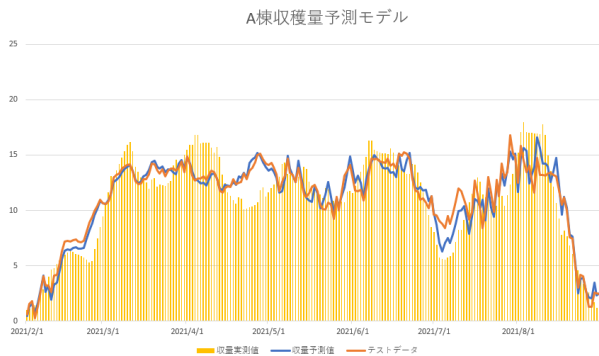


図1 A棟の収穫量予測モデルの評価



図2 I棟の収穫量予測モデルの評価

2021年2月1日から2021年8月31日のネクスグループの訓練データで学習させた重回帰モデルと収量モデルとの適合を示す(図1)。1年間を通して学習したデータの収量の再現ができていたが、細かい部分をうまく再現することができなかつた。しかし、テストデータにおいて自由度調整済み決定係数が0.72となり、実測値を大きく外さず予測ができた。

3.2 I棟のモデル評価

I棟のモデルにおける相関が最も高くなる説明変数について調べた結果、H関数では42日間、日射量積算では4週間の時が最も決定係数が高くなり、0.94となった(表1)。また、誤差の平均値も1.17となり、グラフの精度も最も高くなった。2021年2月1日から2021年8月31日のネクスグループの訓練データで学習させた重回帰モデルと収量との適合を示す(図2)。I棟では1年間を通して学習したデータを大きなずれもなく高精度で再現できていた。特にA棟では精度が低かった3月から5月の期間の予測が非常に高い精度で予測できている。テストデータにおいて自由度調整済み決定係数が0.93となり、こちらも高精度な予測ができていた。

3.3 考察

I棟の重回帰モデルの決定係数が高いことにより、食用ほおずきもトマトと同じように積算日射量とH関数を用いることによって訓練データの収量を

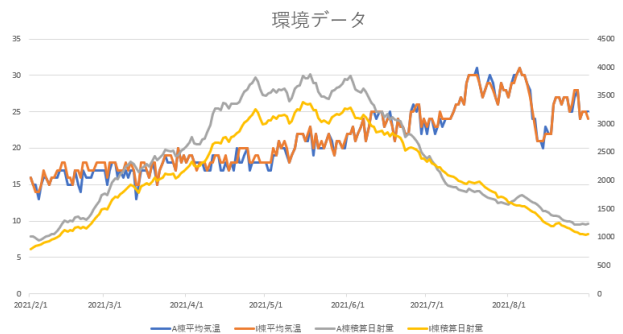


図3 A棟とI棟の環境データの差異

高い精度で再現することができた。しかし、A棟はI棟と比べて予測がうまくできていない。原因として、説明変数では、2つの棟の環境データの平均気温には大きな差はないが、積算日射量はA棟の方が高い傾向がある(図3)。また、積算気温の推移と収量の推移が似ていることから、平均気温よりも積算気温の方が収量に影響していると推測される。目的変数では、A棟は収量が多い日もあれば、少ない日があり、落差がI棟と比べて多い。これら2つの要因が原因でI棟よりも収量予測モデルの精度が落ちたと推測される。また、2つの収量予測モデルを開発する上で、データが今年1年間分しかないことや、収量が0kgの日が多いなど、安定しておらず、14日間の移動平均のデータを使用しているなど、データ不足である可能性がある。そのため、来年度以降のデータを利用してこのモデルを利用した際に、予測値と実数値に差異が起りうる可能性がある。

4. まとめ

本研究では、ネクスグループの食用ほおずきのハウス栽培の収量を対象に、収量予測モデルを開発し、有効性を評価し、問題点について考察した。I棟に関してはデータ不足という懸念点があるが、収量の実測値の傾向は捉えており、実用可能である。A棟では、大まかには実測値の傾向は捉えられているが、収量が上がっている期間で予測値が下がってしまっているなど、不安定な部分がある。

今後の課題として、環境データ、収量データともに学習データを増やす必要がある。また、本研究では先行研究[1]を参考に説明変数を採用したが、今回使用しなかった環境データを説明変数に採用するかを検討する必要もある。また、露地栽培も行っている地域もあるため、気象実況・予測データを用いて収穫量予測モデルを開発する必要もある。

参考文献

- [1] 横田悠, 南野謙一, "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", 情報処理学会第83回全国大会, 2021.