

ビッグデータ分析技術に関する研究

— 農業関連データの活用法検討 —

電子科 岩崎清斗* 望月建治*
静岡県立大学 斉藤和巳**

Study on Big-Data Analysis Technology

— How to utilize data for agriculture —

Kiyoto IWASAKI, Kenji MOCHIZUKI and Kazumi SAITO

Big-data analysis technology is expected to increase the utility value of data and contribute to economic growth. In recent years, the agricultural IoT for collecting data, such as environmental information and the working records, have been spreading to agricultural corporations. However, big-data analysis technology is indispensable to connect the collected data with the environmental control methods and the optimization of agricultural works.

In this study, we analyzed the big-data collected in agricultural field such as environmental information and entry-exit records of workers, and examined how to use them. Firstly, we calculated the humidity deficit, which serves as an index of the wilting condition of the plant, from the temperature and the humidity in the greenhouse. Secondly, we defined the ideal state of the humidity deficit: it does not change more than 10g/m^3 within one hour, the median for a day is within $3\sim 6\text{g/m}^3$, and the lower quartile and the upper quartile are less than 3g/m^3 and more than 6g/m^3 respectively. As a result, we found that the ideal state can be determined by visualizing the humidity deficit in the greenhouse using a box plot. In addition, from the analysis of the environmental information around the greenhouse and the entry-exit records of workers using the classification tree of machine learning, it was found that the entry-exit time of the workers was not affected by the environmental information.

Keywords : big-data, IoT (Internet of Things) , agriculture

キーワード : ビッグデータ、IoT (Internet of Things)、農業

1 はじめに

IoT (Internet of Things) やインターネットの普及により、日々膨大なデータが収集、蓄積されている。それらビッグデータの分析技術は、データの利用価値を高め、経済成長に貢献するものと期待されている。

このような状況下、農業分野では、作業合理化による生産性向上、熟練技術やノウハウの継承が大きな課題となっている。近年、環境情報や作業記録等のデータ収集を目的に、農業法人等で農業IoTの普及が進んでいるが、課題の分析、解決策の発見等の成果に結びつけるためには、生育環境の情報を含めたビッグデー

タ分析技術が必要不可欠である。

本研究では、農業分野で収集、蓄積した生育環境情報等について統計処理言語であるR^{*1}を用いたデータ可視化等の処理を行い、得られた分析結果の活用法について検討した。

Rは、オープンソースの統計解析向けプログラミング言語及びフリーソフトウェアである。充実したライブラリ、マルチプラットフォーム対応等の理由から、ビッグデータ処理ツールとして広く利用されている。

*) 現 機械電子科 **) 現 神奈川大学

* 1) The R Project for Statistical Computing : Getting Started. <https://www.r-project.org/> (2017.4.5アクセス)

【報告】

2 方法

2.1 生育環境データの収集及びクレンジング

御前崎市で葉物野菜の温室栽培を営む農業法人（写真）の協力により、温室内生育環境情報を収集した。協力先は、10分間隔で温室内の温度と湿度を収集する農業IoTサービスを利用しており、本研究では平成29年8月3日～9月13日までのデータを用いた。



写真 葉物野菜の温室栽培

また、作業者が携帯するビーコンの信号を温室内に設置した記録装置が受信することにより、作業者の温室への入退室記録の収集を行った（図1）。

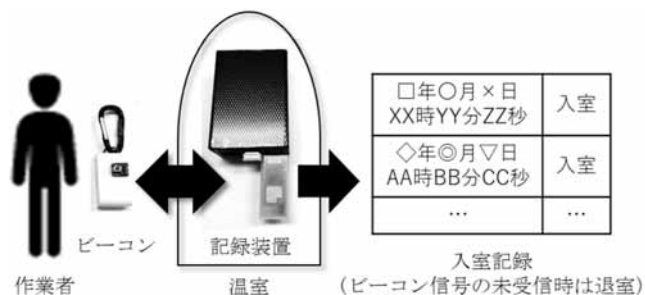


図1 温室入退室記録方法

表1のソフトウェアの構成により、温度、湿度、飽差（空気中にあとどれくらいの水分を含むことができるかを表す指標）、表2に示す生育環境情報と入退室記録より時系列データの作成を行った。温度を $T^{\circ}\text{C}$ 、湿度を $H\%$ 、飽差を HDg/m^3 とすると、Tetensの実験式¹⁾より、 $\text{HD} = (1 - H/100) (1,325.3926 \times 10^{(7.5T/(T+237.3))}) / (T + 273.15)$ から求めることができる。

表1 ソフトウェアの主な構成

OS	Microsoft Windows 10 Pro (64bit)
使用言語	R・3.3.3
使用したRライブラリ	rgl、rpart、partykit
開発環境	RStudio-1.0.136

2.2 理想状態の定義

植物の萎れを抑制するためには、

- ・急激な飽差の変化を避ける
 - ・飽差の最適値（ $3 \sim 6 \text{ g/m}^3$ ）を維持、かつ定常（常に $3 \sim 6 \text{ g/m}^3$ を維持した）状態を避ける
- といった飽差管理が効果的であるといわれている²⁾。ここでは、作成した時系列データから、次のa～cの条件を満たす日を理想状態と定義し、判別を行った。

- 一時間以内の変化量が 10 g/m^3 未満：第 m 日目の時刻 $t = \{1, \dots, 144\}$ の飽差を $x_{m,t}$ 、 $n = \{1, \dots, 6\}$ ステップ前との差分を $d_{m,t}(n) = ((x_{m,t} - x_{m,t-n})) / 2$ （ただし、 $m \geq 2$ かつ $t - n \leq 0$ ならば、 $x_{m-1, 144+t-n}$ となる。また、 $d_{1, 1 \leq t \leq 6} (1 \leq n \leq t)$ は、欠損値とする）とすると、 $\max\{d_{m,t}(n)\} < 10$ を満たす。
- 中央値が $3 \sim 6 \text{ g/m}^3$ 以内：aの条件を満たす第 m' 日目について、時刻 t の飽差 $x_{m',t}$ の代表値が $3 \sim 6 \text{ g/m}^3$ を満たす。この代表値として、外れ値の影響を受けにくい中央値（母集団 $x_{m',t}$ を小さい順に並び替えたときの中心となる値。ただし、母集団が偶数のとき、中心の二値の平均を取る）を用

表2 生育環境情報の詳細

データ名	入手先	変数名	項目	詳細
温室内環境情報	協力先より提供	Temp Humi HD	温度($^{\circ}\text{C}$) 湿度($\%$) 飽差(g/m^3)	協力先が利用する農業IoTサービス（インターネット上で閲覧可）
温室周辺環境情報	気象庁HP	O_Temp O_Humi O_HD		気象庁が公表する御前崎市の環境情報
入退室記録	ビーコン履歴	Stay	入室：1 退室：0	記録装置がビーコンの信号を受信することにより入退室を管理

いる。

- c. 第1四部位～第3四分位が $3 \sim 6 \text{ g/m}^3$ の範囲外：
bの条件を満たす第 m 日目について、飽差 $x_{m,t}$ の50%以上が $3 \sim 6 \text{ g/m}^3$ のとき、飽差が定常状態であったと定義し、理想状態の対象から除外する。

なお、b、cについては、箱ひげ図 (box plot) を用いた判別を行う。箱ひげ図は、データの中央値や分散、異常値等を、図2のように可視化することができる。ここでは、Rのboxplot関数を用いて箱ひげ図を作成した。

2.3 分類木による入退室記録の分析

作業者の温室滞在時は、どのような傾向があるかを把握すべく、入退室記録について分類木³⁾ (classification tree) を用いた分析を検討した。分類木とは、機械学習による非線形判別手法の一種であり、変数を何らかの基準で分岐させ、判別や予測モデルの構築を行う。モデルの理解が容易であることから、データ構造を捉える目的で広く利用されている。ここでは、Rのrpart関数を用いたモデルの作成と分類木の可視化を行った。

3 結果および考察

3.1 生育環境データの収集及びクレンジング

2.1より作成した時系列データの一部を表3に示す。また、温室内と温室周辺の温度、湿度、飽差の関係を表した三次元プロットを図3に示す。飽差は、温度上昇と共に増加、湿度上昇と共に減少し、全体的に

温室内の方が、飽差が高いことが分かった。

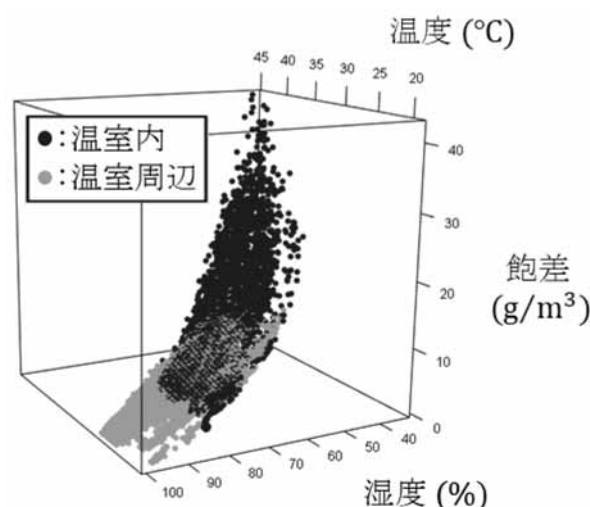


図3 環境情報の三次元プロット

3.2 理想状態の定義

差分 $d_{m,t}(n)$ を計算した結果を図4に示す。同日内の差分は、全て重ねて表示され、2.2aを満たす日が十六日間確認された。2.2aを満たした日の飽差について、箱ひげ図を用いた結果を図5に示す。2.2b、cより、8/16、8/29、9/3、9/4、9/7の五日間が理想状態であったと考えられる。8/7、9/12の二日間は、飽差の変化が小さいことから2.2cを満たしており、理想状態から除外した。

3.3 分類木による入退室記録の分析

結果を図6に示す。分類木は、複雑さのパラメータ cp の値を調整することで、データ構造をシンプルに表

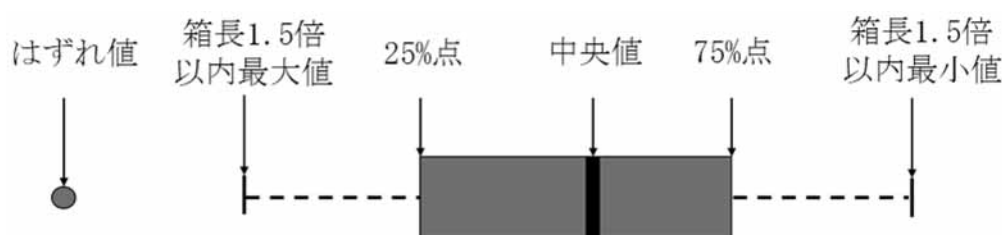


図2 箱ひげ図の概要

表3 時系列データの一部

Date	Time	Temp	Humi	HD	O_Temp	O_Humi	O_HD	Stay
8/3	0:10	23.6	80.6	4.132153	24.3	82	3.989123	0
8/3	0:20	23.6	79.7	4.323851	24.5	83	3.810301	0
8/3	0:30	23.6	79.1	4.451649	24.2	83	3.746262	0
8/3	0:40	23.6	78.8	4.515549	24.4	81	4.234598	0
8/3	0:50	23.6	78.3	4.622047	24.3	81	4.210741	0

現することができる。これを分類木の剪定といい、print cp関数を用いた交差確認法により、最適なcpは0.103743であった。この結果より、作業者は周囲の環境情報よりも、時間帯や日付（曜日）を軸に規則的な作業を行っていることが考えられる。

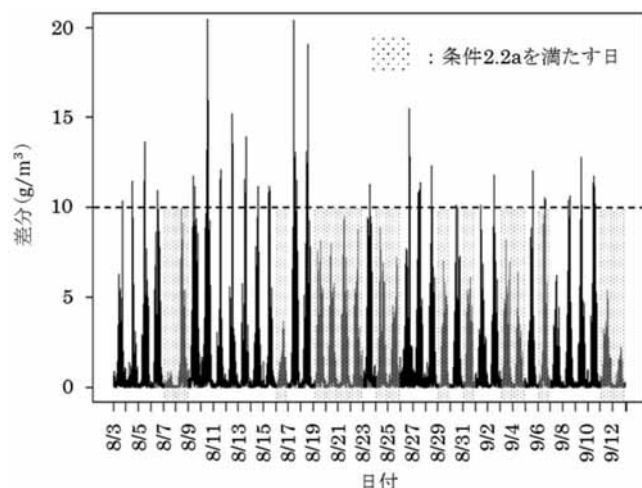


図4 飽差の差分

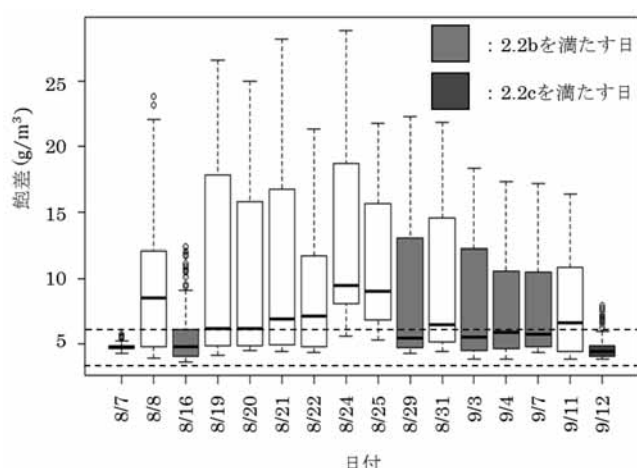


図5 飽差の箱ひげ図

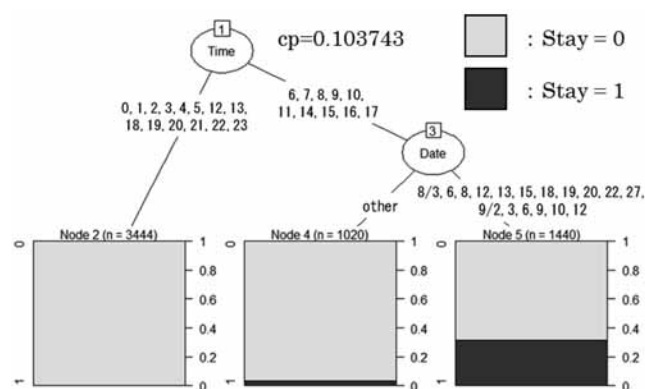


図6 入退室記録の分類木

4 まとめ

温室内環境情報について分析した結果、定義した温室の理想状態の判別が可能となった。また、作業者の温室への入退室記録を分析した結果、決められた時間帯や日付で作業を行うといった規則性の抽出が可能となった。

Rによる指標変換や差分、箱ひげ図、分類木等のビッグデータ分析技術を農業関連データに適用し、データを可視化することで、データの隠れた構造を捉えることが可能となった。その結果、栽培管理の最適化や生産性の向上が期待できる。また、これらの手法は、データ量や変数が増加した場合でも同様の手順で分析可能であるため、今後、IoTの普及が進み、農業分野に限らず分析対象となるデータが膨大となった場合でも、データ可視化手法としての活用が可能である。

謝辞

本研究を行うにあたり、データを御提供いただいた株式会社菜彩ファーム、iway.jpの皆様に感謝いたします。

参考文献

- 1) 岩崎泰永：CO₂施用と湿度管理，「最新農業技術野菜vol.7」，第1版（一般社団法人農山漁村文化協会，東京），農山漁村文化協会 編集，pp.59-61（2014）
- 2) 斉藤章：オランダに学んだ環境制御の取り入れ方，「最新農業技術野菜vol.7」，第1版（一般社団法人農山漁村文化協会，東京），農山漁村文化協会 編集，pp.76-79（2014）
- 3) 金明哲：樹木モデル，「Rによるデータサイエンス」，第1版第4刷発行（森北出版(株)，東京），森北博巳，pp.234-240（2009）