

センサーデータに基づく農業環境の分析 Analysis of Agricultural environment by Sensor data

工藤 秀明[†]
Hideaki Kudo

岩崎 清斗[‡]
Kiyoto Iwasaki

大久保 誠也[†]
Seiya Okubo

斉藤 和巳[†]
Kazumi Saito

1. はじめに

近年, IoT (Internet of Things) 技術は急速に発展している. それにともない, 農業・医療・教育・工芸などの専門知識と熟練が不可欠な分野においても, 専門家や熟練者が, どのような状況で, どのような活動をしているかの情報を, ビッグデータとして比較的容易に入手可能となってきた. 一方で, 特に農業では, 担い手の高齢化による労働力不足が深刻化しており, 作業の合理化による生産性の向上や, 熟練農家が持つ技術の継承が課題となっている.

農業において, 環境は非常に重要な要素である. そして, 熟練農家は, 様々な手段を用いて環境を制御することで, より多くの収穫や, 高い品質の作物を得ている. しかしながら, これらの熟練農家の持つ技術が, 生産性や環境に, どのような影響をどのような形で与えているのかは, 明らかになっていない部分が多い.

本研究の目的は, 熟練農家に特有な環境データの解析である. そこで, 環境の時系列データをクラスタリングすることで, 各農家が環境をどのように制御しているか明らかにする手法を提案する. また, 静岡県内のバラ農家 4 件から得られた環境データに対し, 提案手法を適用することにより, 妥当性の評価を行う.

2. 分析法

時刻 1 から T までの時系列データを T -次元ベクトル $\mathbf{x} = (x_1, \dots, x_T)$ と表す. 総数 N の時系列データ集合 \mathcal{X} が与えられたとき, 任意の時系列ペア \mathbf{x} と \mathbf{y} に対し, 本稿では距離 $d(\mathbf{x}, \mathbf{y})$ として, 標準ユークリッド距離

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{t \in \{1, \dots, T\}} (x_t - y_t)^2}$$

を採用する. また, 類似度 $s(\mathbf{x}, \mathbf{y})$ として, この距離から導かれる

$$s(\mathbf{x}, \mathbf{y}) = 1 - d(\mathbf{x}, \mathbf{y}) / \max_{\mathbf{x}', \mathbf{y}' \in \mathcal{X}} d(\mathbf{x}', \mathbf{y}')$$

を採用する. k -medoids 問題とは, 与えられた非負整数 k に対して, 下記の目的関数 f を最大にする k 個のベクトルの集合 $\mathcal{G}_k \subset \mathcal{X}$, $|\mathcal{G}_k| = k$ を求める問題である.

$$f(\mathcal{G}_k) = \sum_{\mathbf{x} \in \mathcal{X}} \max_{\mathbf{g} \in \mathcal{G}_k} \{s(\mathbf{x}, \mathbf{g})\}. \quad (1)$$

クラスタリングを行う際には, \mathcal{G} の各ベクトルが, 各クラスタの代表ベクトルとなる.

目的関数 f の最大化の標準解法となる貪欲法の詳細を以下に示す.

1. 反復制御変数を $h = 0$ とし, 結果を格納する集合を空 $\mathcal{G}_0 = \emptyset$ に初期化;
2. 集合 \mathcal{G}_h を固定し, \mathcal{X} から最良要素 $\mathbf{g}_{h+1} = \arg \max_{\mathbf{x} \in \mathcal{X}} \{f(\mathcal{G}_h \cup \{\mathbf{x}\}) - f(\mathcal{G}_h)\}$ を計算;
3. 最良要素 \mathbf{g}_{h+1} を追加 $\mathcal{G}_{h+1} \leftarrow \mathcal{G}_h \cup \{\mathbf{g}_{h+1}\}$ し, $h = h + 1$ に設定;
4. $h = k$ ならば終了, さもなければステップ 2. へ戻る.

貪欲法では厳密解は得られないが, 妥当な精度で最悪ケースの解品質を理論的に保証できる. 詳細には, 厳密解を \mathcal{G}^* とすると, 貪欲法で求まる近似解 \mathcal{G} の精度は, 関係式 $f(\mathcal{G}) \geq (1 - 1/e)f(\mathcal{G}^*)$ で抑えられる. ここで, e は自然対数の底である. 貪欲解により, 最悪でも厳密解の 63% 程度以上の性能が保証される [1].

本稿では, 複数の時系列データ集合 $\mathcal{X}^{(A)}, \mathcal{X}^{(B)}, \dots$ の特性を分析するための方法論を提案する. 具体的には, 複数のビニールハウス A, B, \dots で, ある一定期間に同一時刻で計測された, 各日毎の農業環境時系列データの集合より, 各ビニールハウスでの農業環境の特性を特徴付ける. 以下に提案分析法を示す.

1. medoids 数 k を適切に設定し, 各時系列データ集合 $\mathcal{X}^{(*)}$ に対し, 上述した貪欲法に基づき, k 個の medoids 集合 $\mathcal{G}_k^{(*)} \subset \mathcal{X}^{(*)}$ を計算;
2. 求めた $\mathbf{g}_1^{(*)}, \dots, \mathbf{g}_k^{(*)}$ を時系列でプロットし, 各時系列データ集合の特性を分析する.

3. 農業環境データの収集

提案手法の有効性を検証するための環境データを, 小型デバイスを設置することにより収集した.

収集したデータは, 静岡市内にある 4 つの農家における, バラのビニールハウスのものである. 各農家のビニールハウスは, 詳細は異なるが, 環境を制御する手段を備えている. たとえば, 温度等を調整する暖房や, 照度等を調整するカーテン等がある. 各ビニールハウスの環境は, これらの設備が自動もしくは手動で動作することにより, 制御されている. ただし, 実験期間中, 農家 A (ハウス A) のみは, 出荷期間でなかったため, これらの設備はほとんど使用されていない.

環境データの収集のため, データ収集用の小型デバイスを作成し, 各ビニールハウスに設置した. 小型デバイスは, RaspberryPi3 と各種センサーからなる. 小型デバイスの詳細を表 1 に示す. 本デバイスは, おおよそ 5 分に 1 回の頻度で各種情報の収集を行う. 4 つ

[†]静岡県立大学 経営情報学部

[‡]静岡県工業技術研究所 機械電子科

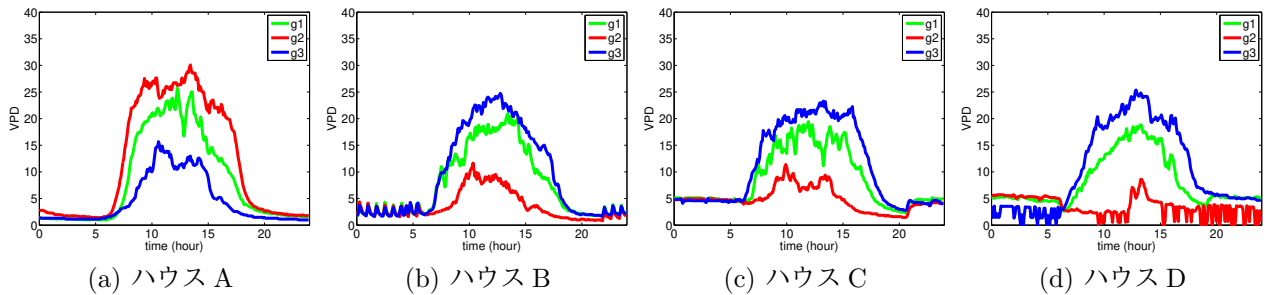


図 1: 飽差に基づく各ビニールハウスでの農業環境の比較

表 1: 環境データ収集用デバイス詳細

| センサ名 | 型番 |
|-------|---------------------|
| 本体 | RaspberryPi3 ModelB |
| 温度センサ | BME280 |
| 湿度センサ | BME280 |
| 照度センサ | TSL2561 |
| カメラ | KP-CAM32 |

表 2: 各クラスタのサイズ

| | ハウス A | ハウス B | ハウス C | ハウス D |
|----|-------|-------|-------|-------|
| g1 | 14 | 17 | 17 | 20 |
| g2 | 18 | 11 | 13 | 12 |
| g3 | 10 | 14 | 12 | 10 |

の農家に共通したデータ収集期間は 2018 年 3 月 26 日 16 時 10 分から 2018 年 5 月 8 日 9 時 55 分までであり、本解析では 2018 年 3 月 27 日 0 時 0 分から 2018 年 5 月 7 日 24 時 00 分までのデータを使用した。

収集されたデータは、おおよそ 5 分に 1 回と、厳密に 5 分間隔では無い。また、電源投入時からデータを収集するため、各農家の収集データ時刻に差が存在する。そこで、2018 年 3 月 27 日 0 時から 5 分間隔のデータを、補間することによって作成した。補間は R の Pandas を用いて行った。補間方法は index であり、時間に応じた按分が行われている。最終的なデータは、1 日あたり 288 次元ベクトルが、42 日分である。

また、小型デバイスは飽差 VPD を直接測定するセンサーは備えていないため、以下の式により求めた。

$$VPD = (100 - H) * \frac{217 * \frac{6.1078 * 10^7 * T^{7.5 * T}}{T + 237.3}}{T + 273.15} / 100$$

ここで、 H は湿度、 T は気温である。

4. 評価実験

得られた環境データに対して、提案手法を適用した。本実験では、42 個のベクトルを 3 つの集合に分割し、その代表ベクトルを評価した。実験結果を図 1 に示す。横軸は時刻、縦軸は飽差である。また、クラスタリングした際の、各クラスタのサイズを表 2 に示す

どのハウスでも、最初に中央付近の値として g1 が選

ばれ、その後、小さい値と大きい値の g2, g3 が選ばれている。また、ハウス B から D については、g1 から g3 のクラスタのサイズも大きく変わらない。

ハウス B からハウス D では、小さい値のグループの値は、日中でも $3 \sim 10 \text{g/m}^3$ 程度の値に収まっている。一方で、ハウス A においては、小さい値のグループの値は、日中だと $10 \sim 15 \text{g/m}^3$ に達している。これは、環境を制御する設備を使用したか使用していないかによって生じていると考えられる。一般には、飽差の値は $3 \sim 6 \text{g/m}^3$ がよいとされている [2]。環境を制御しているハウスにおいては、多くの日でこの値付近になっているであろうことが示唆される。一方で、高い値になってしまう日、もしくは制御を行っていない日があることがわかる。

また、ハウス B とハウス D においては、夜間時に値が上下している。実際には、ハウス B とハウス D は、温度がある範囲内に収まるように自動制御されており、収集データにおいても温度が上下していた。その結果が飽差にも反映されている。

以上のように、提案手法を用いることにより、いくつかの示唆が得られることが明らかとなった。

5. おわりに

本研究では、農業における環境データを分析する手法について提案した。提案手法は、農業環境データを 1 日ごとの時系列データとして扱い、各データをクラスタリングすることにより特徴付ける。提案手法の有効性を示すため、小型デバイスにより静岡市内のビニールハウスのデータを収集し、得られたデータに提案手法を適用した。結果、各農家が環境を制御した影響等、いくつかの事象が明らかとなった。小型デバイスの改良や、熟練農家の技術が環境に与える影響の評価などが、今後の課題である。

謝辞 本研究は、科学研究費補助金基盤研究 (C)(No.18K11441) の助成を受けた。

参考文献

- [1] 室田一雄：離散凸解析の考えかた 最適化における離散と連続の数理，共立出版 (2007)。
- [2] 農山漁村文化協会 (編)：野菜 vol.7: もっと知りたい環境制御技術-日中 CO2 濃度，飽差，葉面積を管理する，農山漁村文化協会 (2014)。