

SVMを用いた農業環境データの危険値設定法 Assignment of Critical Value to Agricultural Environment Data using SVM

古瀬純大[†]
Jundai Furuse

伏見卓恭[‡]
Takayasu Fushimi

大久保誠也[§]
Seiya Okubo

斉藤和巳^{†§¶}
Kazumi Saito

1. はじめに

近年, IoT (Internet of Things) 技術が急速に発展している. それにともない, 農業・医療・教育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に関する情報を, 容易に入手可能となってきた. 一方, 特に農業分野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, 作業の合理化や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. 熟練農家は, 様々な手段を用いて環境を制御することにより, より多くの収穫や高い品質の作物を得ている. 一方, かなりの高温状態が続くなど, 農作物に対して致命的に危険なダメージを与えることは確実に回避しなければならない. そのために, 例えば, 大規模ビニールハウス内での温度上昇を防ぐために, カーテンを閉めることにより直射日光を遮る. しかしながら, 高速動作する設備導入には多大なコストがかかるため, 通常の大規模ビニールハウスでは, カーテンが全部開いている状態から完全に閉め切るまでに然るべく時間が必要となる. そのため, 危険を避けるためには, 事前に危険を察知し, あらかじめ装置を動かさし始めることが必要となる. よって, 現時点の農業環境データから, 然るべく準備時間を確保した将来時間において, 危険状況を予測する問題は重要となる. また, 農家ごとにビニールハウス内の環境は大きく異なるが, これはビニールハウス内の設備や農作業にかかる時間, 作物に対する手入れの時間なども影響を与えている. そのため設備の機能などによってはビニールハウス内の環境制御に時間がかかってしまう場合などもあり, 農家によってどれぐらい先の時間でどの程度値が下がったか, もしくは大きくなったかを検知したいか異なってくる.

本研究では, 任意の危険値, 推定時間の時に, どれぐらいの精度, 正確さで予測分類できるかを可視化する手法を探求する. このような危機回避のための予測問題を解決することにより, 農作業の合理化や技術の継承などへの貢献を目指す. このような危険値予測問題では, 危険状況度合いの設定によっては, そのような危険状況を示すデータが極端に少なくなるため, 妥当な精度の推定器構築の学習は困難になる. 我々の研究 [1] では, SVM (Support Vector Machine)

と学習パラメータの調整により, 正例と負例のサンプル数が極めてアンバランスな問題での分類器を構築する手法を提案した. 具体的には, 上述するような危険状況度合い k や, 推定する将来時間 T を変えて構築した分類器について, 任意の危険度合い k , 推定する将来時間 T における精度を, Precision と Recall を求めることで測定し, その結果をグラフで表すことにより, ある危険値, 推定時間においてどれぐらいの精度で予測分類できるのかを可視化する. 研究 [1] においては, 有効性の評価を実験により行ったが, 二酸化炭素 CO₂ の危険値の予測分類は2パターン, 飽差 HD の危険値の予測分類は3パターンのみのため, その前後での危険値の予測分類でも有効かという事や関係性は不明であった. また, 危険度合いが増加することによる precision や recall の変化も, 前回作成したグラフからは読み取りづらく関係性も不明慮な点も多かった. 本稿では, カーテン開閉など環境制御のために必要となる農家の稼働件数の視点も導入した可視化法を提案する. そして, より広い範囲の危険値について評価実験を行うことにより, 前回よりも詳細な有効性の評価を実施する.

2. 提案手法

本研究では, 時刻 t の環境データから, 時刻 $t+T$ の危険値を予測する. つまり, 以下の問題を解く.

問題: 時刻 t における n 種類の環境データ $D_t = \{x_{(t,0)}, x_{(t,1)}, \dots, x_{(t,n-2)}, y_t\}$ と閾値 k が与えられる. このとき, 時刻 $t+T$ 時点の $y_{(t+T)}$ が k を上回る (もしくは下回る) か否かを判定しなさい.

典型的な場合では, 時刻 t は現時刻となる.

この問題を解くために, 本研究ではSVMを用いた手法を提案する. つまり, 事前に環境データを収集することにより学習データを作成し, そのデータを基にしたSVMの学習を行う. また, ハウスごとに条件は大きく異なるため, “どのぐらいの精度で”・“どのぐらい先の時間まで”・“どのぐらいの頻度で”危険を予想できるのかが, ハウスによって異なってくる. したがって, 事前にパラメータ k, T について評価を行う必要がある. そのため, 分類器の精度を precision と recall に分けグラフとともに, 農家のカーテンの開閉や炭酸ガスの放出, 放水といった環境制御の稼働件数の視点も導入した可視化法を提案する. これは, 前回の研究 [1]

[†]神奈川大学 理学部

[‡]東京工科大学 コンピューターサイエンス学部

[§]静岡県立大学 経営情報学部

[¶]理化学研究所 革新知能統合研究センター

で提案した手法の拡張法と言える。

提案手法は、SVMでデータを学習し、分類器を構成する学習フェーズと、実際に分類器に値を入力し、時刻 T 後の危険値を求める利用フェーズから構成される。利用フェーズでは、学習フェーズで得られた学習器を用いて、危険値を予測するので、本稿では、学習フェーズについてのみ説明する。

1. S 個の組 $(D_t, z_t), t \in \{0, \dots, S-1\}$ からなるデータ D を準備する。ただし、

$$z_t = \begin{cases} 1 & (y_{(t+T)} < k \text{ (or } y_{(t+T)} > k)) \\ -1 & (\text{otherwise}) \end{cases}$$

2. データ D を、学習用データ D_L と調整用データ D_M に分割する。
3. 以下を繰り返す。
 - (a) パラメータ T, k を決める
 - (b) 学習用データ D_L を用いて SVM の学習を行い、分類器を構成する。
 - (c) 得られた分類機に対して調整用 D_M データを入力し、結果の precision と recall を表す $pT(k)$ と $rT(k)$ に加えて、SVM で危険値を超える判定された判定件数 $XT(k)$ と実際に危険値を超えた異常件数 $YT(k)$ を計算する。

4. 得られた $pT(k)$ と $rT(k)$ のペア群と、 $XT(k)$ と $YT(k)$ のペア群をプロットすることで、各危険度合い別、推定時間別にどれぐらいの精度で予測分類が可能なのか、どの程度の頻度で起こるのかを可視化し、最終的に利用する T, k ならびに学習器を決める。

本研究では、SVMのカーネル関数として、ラジアル基底関数 (RBF, radial basis function, ガウシアンカーネル) を用いている。間違いデータの許容値は 0.1, 境界の複雑さを示す gamma は 15 としている。

ステップ (4) では、学習器の評価を行う。危険予知において、もっとも重要なことは危険を避けることである。つまり、誤った危険を予知することよりも、予知しない危険が発生してしまうことのほうが、大きな問題となりえる。そのため、本研究では全体の正解率ではなく、Precision (適合率。危険と予測したもののうち、本当に危険だったものの率) と Recall (再現率。本当に危険だったもののなかで、予測も危険であったものの率) を評価する。具体的には、以下のように行う。

1. 横軸に危険度合い k , 縦軸に $pT(k)$ と $rT(k)$ をプロットしたグラフ (a) を描く。

2. 横軸に危険度合い k , 縦軸に $XT(k)$ と $YT(k)$ をプロットしたグラフ (b) を描く。

3. 各グラフを評価することで、適切な k と T を求める。

3. 評価実験

実験により、提案手法の特徴と有効性の評価を行った。具体的には、静岡県内にある2件のバラ農家のビニールハウス (ハウス A と B) から得た環境情報データを用いて、提案手法で学習器を構築した。環境情報データは、これまでの研究 [1, 2] と同様に、ビニールハウスに環境情報収集デバイスを設置して収集した。本実験で用いたデータは、ハウス A が 2019 年 6 月 21 日 0 時 0 分から 2019 年 10 月 16 日 23 時 50 分までの 10 分刻みのデータである。1 日あたり 144 次元ベクトルで、それが 118 日分であるため、16992 個のデータとなり、ハウス B が 2019 年 6 月 21 日 0 時 0 分から 2019 年 8 月 14 日 24 時 00 分までの 10 分刻みのデータである。1 日あたり 144 次元ベクトルで、それが 54 日分であるため、7776 個のデータとなる。各時刻におけるデータ D_t は、温度・湿度・照度・CO2 濃度・飽差となる。ここで、飽差とは、ある温度と湿度の空気に、あとどれだけ水蒸気の入る余地があるかを示す指標であり、農産物の成育に、ある一定範囲での飽差制御が重要であるとされている [3]。小型デバイスは飽差 HD を直接測定するセンサーを備えていないため、以下の式により求めた。

$$HD = (100 - H) * \frac{217 * \frac{6.1078 * 10^{7.5 * C}}{C + 237.3}}{C + 273.15}}{100}$$

ここで、 H は湿度、 C は気温である。また一般には、植物が光合成をおこなう際、植物によって差はあるものの飽差の値はおおよそ 3 ~ 6g/m³ がよいとされている [3]。

本稿では、飽差 HD (Humidity Deficit) と二酸化炭素 CO2 を対象とした解析を行う。つまり、二酸化炭素がある値 k 以下になるか否かの予測と、飽差の値がある値 k 以上になるか否かの予測を行った。また、 D_L と D_M は、 D をランダムに 2 分割することにより生成した。今回の実験で扱ったハウス A と B の環境データを比較すると、ハウス B の方が飽差、CO2 の値が全体的に低かった。そのためハウス B の危険度合いの値 k を、飽差、CO2 とともに小さい値とした。ハウス A においては、飽差 HD に関する危険度合いを 3 ~ 30 で 1 刻み、二酸化炭素 CO2 に関する危険度合いを 400 ~ 210 で 10 刻みとした。ハウス B においては、飽差 HD に関する危険度合いを 3 ~ 20 で 1 刻み、二酸化炭素 CO2 に関する危険度合いを 290 ~ 195 で 5 刻みとした。推定時間設定 T はハウス A と B で共通で 30, 60, 120 とした。

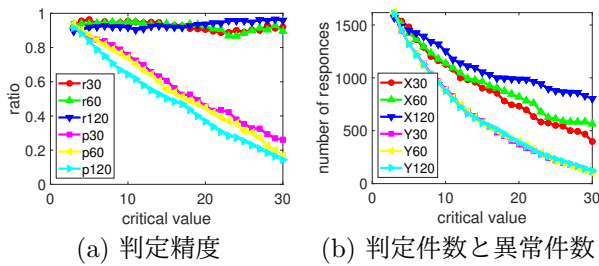


図 1: ハウス A の飽差 HD での評価結果

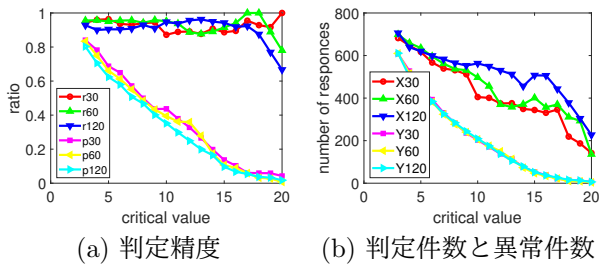
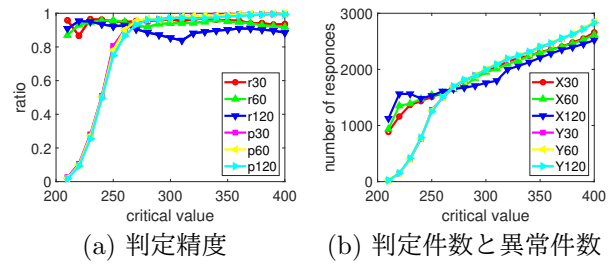
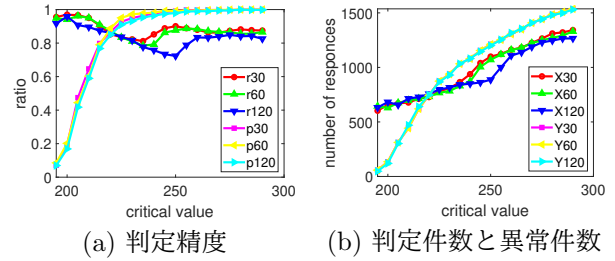


図 2: ハウス B の飽差 HD での評価結果

4. 可視化結果

ハウス A における飽差 HD と二酸化炭素 CO₂ での評価結果を図 1 及び図 2 に、ハウス B における飽差 HD と二酸化炭素 CO₂ での評価結果を図 3 及び図 4 に示す。ここで、各図の凡例の数値 30, 60, 120 は推定時間 T を表す。すなわち、第 2 章で説明したように、横軸の飽差 HD や二酸化炭素 CO₂ の危険値に対し、各図の (a) は、recall の $rT(k)$ と precision の $pT(k)$ をプロットしたものであり、各図の (b) は、SVM が危険と判定した件数 $XT(k)$ と実際の異常件数 $YT(k)$ である。ここで、 $YT(k)$ が実際の異常件数であり、SVM による判定件数 $XT(k)$ のグラフが $YT(k)$ に近ければ、行わなければならない環境制御回数、実際に必要な環境制御回数に近くなるので無駄な労力が小さくて済む。逆に、 $XT(k)$ が $YT(k)$ から離れれば、行わなければならない環境制御回数が多くなってしまふので労力が大きくなる。

まず、ハウス A の飽差 HD での判定精度については、図 1(a) より、危険度合いが増加した場合や推定時間が増加した場合でも recall に大きな変化はなかったが、precision の割合は危険度合いが増加するにつれて減少している事が分かる。推定時間に関しては、30 分後と 60 分後では precision は大きく変化しないが、推定時間が 120 分後になると他 2 つに比べて実質的に低下していることが分かる。このことから飽差 HD の precision は推定時間が長くなればなるほど precision の値が急激に悪くなっていくと想定できる。図 2(a) より、これらの事はハウス B の飽差 HD での判定精度でも同様の事がいえる。一方、ハウス A と B の飽差 HD での判定件数と異常件数については、図 1(b)

図 3: ハウス A の二酸化炭素 CO₂ での評価結果図 4: ハウス B の二酸化炭素 CO₂ での評価結果

と 2(b) より、危険度合いが増加するにつれて XT と YT のグラフが離れていき、さらに推定時間が大きくなるほど XT と YT のグラフの距離は離れていることが分かる。

次に、ハウス A の二酸化炭素 CO₂ での判定精度については、図 3(a) と 4(a) より、飽差 HD の結果と似ており危険度合いが増加しても recall は大きく変化することなく平均的に高い値を維持し続けていることが分かる。また、推定時間が増加しても、飽差 HD と同様に recall に大きな違いはみられない。一方、precision については、飽差 HD の場合は危険度合いの増加に伴いそれに recall の値も減少したが、二酸化炭素 CO₂ の場合では、危険度合いが増加しても、ある危険値に達する前では、非常に高い値を維持しているものの、ある危険値を境にして急激に precision が減少していくことが分かる。また、飽差 HD では推定時間が増加する程 precision が減少するという傾向にあったが、二酸化炭素 CO₂ の場合は推定時間が増加しても、precision の値に大きな変化は見られなく同じように変化していったことも分かる。

これに対し、ハウス B の二酸化炭素 CO₂ での判定精度については、図 4(a) より、ハウス A に比べて全体的に少し低い傾向となるものの、ある危険値を境に recall が増加していく傾向がみられる。また、ハウス B の precision については、ハウス A と同様に危険度合いが増加してもある危険値に達する前では、高い値を維持しているが、ある危険値を境にして大きく減少していくことも分かる。一方、ハウス A と B の二酸化炭素 CO₂ での判定件数と異常件数についても、図 3(b) と 4(b) より、飽差 HD の場合とは大きく異なり、ある危険値に達する以前では、推定時間が大きくなってあまり XT

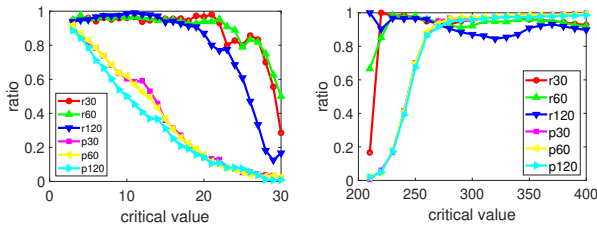


図 5: ハウス A における前回データでの飽差 HD と二酸化炭素 CO₂ の判定精度

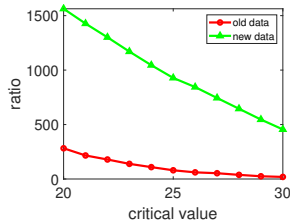


図 6: 学習データに占める危険値を超えるデータ数

と YT のグラフの距離が変化する事はなかったが、こちらも precision と同様に、ある危険値に達すると XT と YT の距離が離れていっていることが分かった。

5. 考察

まず初めに、図 1～図 4、それぞれのグラフの recall に注目する。前回の研究 [1] では危険度合いの増加による recall の変化があまり明確には読み取ることができなかったが今回の研究では、飽差 HD、二酸化炭素 CO₂ ともにどのような値の予測分類でも recall は比較的高い値になるという事が分かった。今回の研究ではハウス A のデータの期間を 2 倍弱に増やした。そこで比較対象として前回と同じデータで同様の実験をハウス A で行った。その結果が図 5 である。図 5 をみて分かるように特に飽差 HD における予測分類では前回のデータでは飽差 HD の値が 20 を超えたあたりから recall が減少し始めている。図 6 は飽差 HD の recall が減少を始めた 20 付近からの 60 分後の予測分類を行った際の学習データに占める危険値を超えるデータ数をプロットしたものである。グラフをみても分かるが今回用いたデータの方がデータの数が約 4～5 倍になっていることが分かった。データ数が 2 倍弱しか増えていないにも関わらず危険値を超えるデータ数が約 4～5 倍に増えた理由として考えられるのは今回追加で用いた 8 月 16 日以降のハウス A の飽差 HD の値が高かったためと考えられる。よって飽差 HD の recall の減少はデータの不足による学習不足であると考えられる。このことから、より広い範囲での詳細な予測分類を行うためには、より妥当なデータで行う必要がある。次は precision についてみていくと今回の研究では飽差 HD の場合はハウス A、ハウス B ともに危険度合いの増

加に伴いそれに比例するように precision の値も減少してき、飽差 HD に対し二酸化炭素 CO₂ の場合はハウス A、ハウス B ともあるところまでは precision も recall と同じように非常に高い値を維持しているが、あるときを境にして precision が減少するといった傾向にあることが分かった。

さらに、図 1～図 4 の (b) のグラフでは飽差 HD の場合は危険度合いが増加すると判定件数 (XT) も増加し推定時間が増加するとその変化がより顕著なものになり労力も大きくことが分かった。しかし、逆に二酸化炭素 CO₂ の場合は、危険度合いが増加しても判定件数 (XT) には変化がほとんどなく、推定時間が増加しても飽差 HD の場合とは違い判定件数 (XT) の変化もわずかしかなくことが分かった。また、図 3、図 4 のハウス A とハウス B の (b) のグラフで異常件数 (YT) が判定件数 (XT) を下回り始めた場所が (a) のグラフにおける precision が減少し始めた場所とほとんど一致している事が分かる。このことから、precision の減少は異常件数が大きな要因の 1 つになっていると考えられる。

6. おわりに

本研究では、SVM を用いた農業環境データの危険値設定法を提案した。提案手法の有効性を示すため、小型デバイスにより静岡県内のバラ農家のビニールハウスのデータを収集し、提案手法を適用した。結果、今回用いたデータに対して SVM を用いた危険値の設定法が有効であることを示した。今後の課題として、実際に IoT 機器により送られてくる時系列データに対する評価またその結果から環境状況を人為的に変化させた際の環境データにどれほど対応できるのかなどがあげられる。

謝辞 本研究は、JSPS 科研費 (No.19K20417) の助成を受けた。

参考文献

- [1] 古瀬純大, 伏見卓恭, 岩崎清斗, 大久保誠也, 齊藤和巳. "農業環境データにおける SVM を用いた危険値予測," 情報処理学会第 199 回知能システム研究会, 2020.
- [2] 岩崎清斗, 伏見卓恭, 大久保誠也, 齊藤和巳. "差分累積値に基づく農業環境データの可視化分析法," 情報処理学会第 121 回数理解モデル化と問題解決研究会, 2018.
- [3] 農文協 (編). "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," 野菜, vol.7, 農山漁村文化協会, 2014.