

農業環境データにおけるSVMを用いた危険値予測

古瀬 純大¹ 伏見 卓恭² 岩崎 清斗³ 大久保 誠也⁴ 斉藤 和巳^{1,5,4}

概要: 本研究では、農作物の成長においていい影響を及ぼさないであろう環境データを事前に予測することによって農作物に効率的な成長を促す事を目的としている。特に本研究では環境データの中でも飽差とCO₂に焦点を当て、一定時間後に飽差・CO₂の値がある一定の値を超えるもしくは下回ることを予測分類する。そこで、SVMを用いた教師あり学習による環境データの予測分類を提案する。評価実験では、静岡県内のバラ農家に、データ収集用小型デバイスを設置し、各種環境データを収集した。その後、飽差およびCO₂データに対しラベル付けを行いSVMを用いて予測分類を行った。実験結果から今回用いたデータではSVMを用いたアプローチが有望であることが明らかになった。

キーワード: 危険値予測, SVM (Suport Vector Machine), 可視化, 農業環境データ

Prediction of Critical Value from Agricultural Environment Data using SVM (Suport Vector Machine)

JUNDAI FURUSE¹ TAKAYASU FUSHIMI² KIYOTO IWASAKI³ SEIYA OKUBO⁴ KAZUMI SAITO^{1,5,4}

Abstract: In this study, by predicting in advance environmental data that will not have a positive effect on crop growth, It aims to promote efficient growth of crops. In particular, this study focuses on HD and CO₂ in environmental data, and predicts and classifies that after a certain period of time, the values of HD and CO₂ exceed or fall below certain values. Therefore, we propose a predictive classification of environmental data by supervised learning using SVM. In the evaluation experiment, a small device for data collection was installed at a rose farmhouse in Shizuoka Prefecture to collect various environmental data. After that, the HD and CO₂ data were labeled and the predictive classification was performed using SVM. The experimental results show that the SVM approach is promising with the data used in this study.

Keywords: critical value prediction, SVM (Suport Vector Machine), visualization, agricultural environment data

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、例えば、大規模ビニールハウス内での温度上昇

¹ 神奈川大学 理学部
Faculty of Science, Kanagawa University
² 東京工科大学 コンピュータサイエンス学部
School of Computer Science, Tokyo University of Technology
³ 静岡県工業技術研究所 機械電子科
Industorial Research Institute of Shizuoka Prefecture
⁴ 静岡県立大学 経営情報学部
School of Management and Information, University of Shizuoka
⁵ 理化学研究所 革新知能統合研究センター
Center for Advanced Intelligence Project, RIKEN

を防ぐには、直射日光を避けるためにカーテンを閉めることが有効な手段とされているが、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。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、農作物に対してこのような危険状況を予測する問題は重要となる。すなわち本研究では、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、一般に、推定する将来時間 T を大きく設定すれば推定精度が下がること、および、危険状況度合いの設定により、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ことが課題となる。そこで本研究では、SVM (Support Vector Machine) を採用し学習パラメータを調整することにより、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。そして、上述するような危険状況度合いや、推定する将来時間 T を変えて構築した分類器の精度を、PR (Precision-Recall) 曲線として可視化することにより、精度の視点より、設定可能な危険度や将来時間の分析を可能とする方法論を提案する。実際のバラ農家のビニールハウスから得られた現実データを用いた評価実験では、提案法が有望であることを示す。

本稿は以下に示す構成である。第2章で提案手法に関連する既存研究について整理する。第3章で提案手法の定義と計算アルゴリズムについて解説する。第4章で実データを用いた評価実験およびその結果について議論する。最後に本稿のまとめと今後の課題について述べる。

2. 関連研究

本研究では、SVM を用いた危険値予測を提案するとともに、提案手法を農業環境データに適用する。たとえば、大塚 [2] らの研究は環境情報システムを実装し局所的な異常気象の予測を目的としていた。また気象庁提供データと比較することにより観測地点の急激な気圧変化を判断できることを示した。これに対し、本稿ではビニールハウス内の環境データを収集し分析することで急激な CO2 や HD の変化を予測できるようにしている。

3. 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 $\mathcal{D}_t = \{x_{(t,0)}, x_{(t,1)}, \dots, x_{(t,n-2)}, y_t\}$ 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” 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 PR 曲線として可視化することにより、評価を行う。

提案手法は、学習フェーズと利用フェーズからなる。学習フェーズでは、学習データを用いて SVM の学習を行い、分類器を構成する。利用フェーズでは、実際に分類器に値を入力し、時刻 T 後の危険値を求める。

学習フェーズは、以下のように行う。

- (1) S 個の組 $(\mathcal{D}_t, z_t), t \in \{0, \dots, S-1\}$ からなるデータ D を準備する。ここで、

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

である。

- (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、成功率を計算する。
- (4) 得られた成功率をプロットすることで、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) Recall と Precision をそれぞれ縦軸と横軸にしたグラフをプロットする。

(2) 各グラフを評価することで、適切な k と T を求める。

得られるグラフは図2のようになる。グラフ中、横軸は precision で縦軸が recall、また左下が0である。また、HD が条件 k の値となる。グラフには、さまざまな条件における値がプロットされている。Recall が小さいことは予測すべき危険を検知できず危険に直面することを、precision が小さいことは予知した危険が外れていることを意味している。つまり、右上にプロットされている点ほどよい結果となる。そして、 T の値が小さいときは近い未来の危険しか予知できないこととなる。

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、 $D_t, t \notin \{0, \dots, S-1\}$ に対する z_t を求める。

4. 評価実験

評価実験により、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）から得られた環境情報データを用いて、提案手法による学習器の構築を行った。

環境情報データは、ビニールハウスに環境情報収集デバイスを設置することにより収集した。デバイスの外観と搭載センサを、図1と表1に示す。デバイスは親機と子機にわかれている。子機には温湿度センサが搭載されており、親機にデータを送信する。親機には、温湿度以外のセンサと通信モジュールが搭載されており、データの収集・保存を行う。本実験で用いたデータは、[1]で用いたものと同様であり、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、各農家に対して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、農産物の育成に、ある一定範囲での飽差制御が重要であるとされている。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。

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

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 $3 \sim 6 \text{g/m}^3$ がよいとされている。

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つ

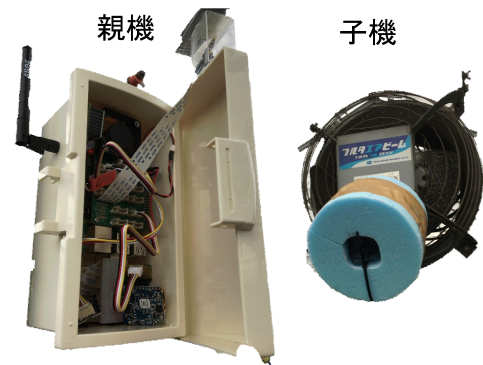


図1 データ収集用デバイス外観

表1 搭載センサ

種類	型番
照度センサ	LTR-308ALS
CO2 センサ	ELT-S-300L-3V
カメラ	Raspberry Pi PiNoir CameraModule V2 赤外線カメラモデル
温湿度センサ	SHTC3

表2 ハウスAの危険値の数

	HD			CO2	
	> 5	> 10	> 20	< 300	< 400
ハウスA	2865	1642	342	4520	6170

表3 ハウスBの危険値の数

	HD			CO2	
	> 5	> 7.5	> 10	< 200	< 225
ハウスB	2200	1497	983	678	4378

た。そのためハウスBの危険度合いの値 k を、飽差、CO2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5, 10, 20、二酸化炭素CO2に関する危険度合いを200, 300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5, 7.5, 10、二酸化炭素CO2に関する危険度合いを200, 225とした。推定時間設定 T はハウスA,B共通で10, 30, 60, 90, 120とした。各条件において、データ D に含まれる危険値の数を、表2及び、表3に示す。

5. 可視化結果

ハウスAにおける飽差HDでの評価結果を図2に、ハウスBにおける飽差HDでの評価結果を図3に示す。

図2(a)及び図3(a)は、危険度合いを変更した評価の結果である。どちらのどのグラフでも時間 T が大きくなればなるほど precision の値が小さくなっているのが分かる。また、recall に関してしてみると、HDの危険度合いが小さければ小さいほど recall は低下していることがわかる。一方、危険度度合いが大きい場合においては、precision が低下しているのにも関わらず recall が上昇するという状

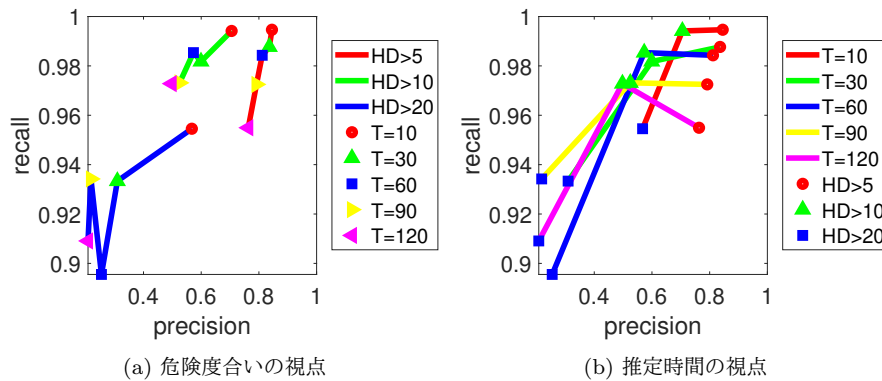


図 2 ハウス A の飽差 HD での評価結果

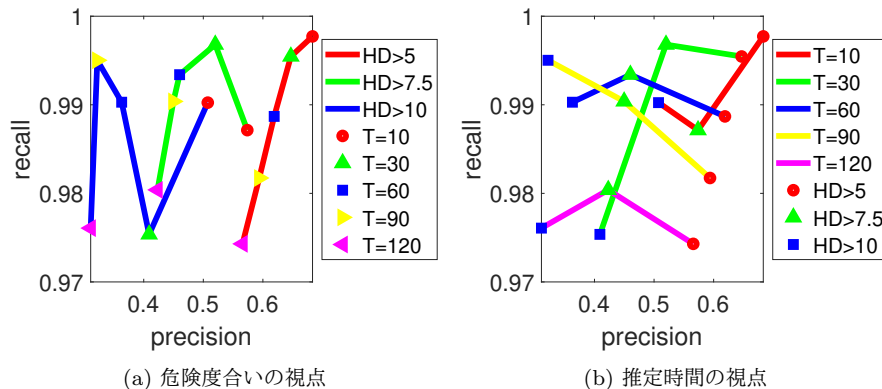


図 3 ハウス B の飽差 HD での評価結果

況が存在し、危険度合いが大きくなるとその変化率も大きくなることが分かった。このことから、危険度合いが大きくなるにつれ recall との相関関係が小さくなると推察される。また、ハウス A のグラフは危険度合いが $HD > 5$ のときと $HD > 10$ のときで recall にあまり変化はみられないが、危険度合いが $HD > 20$ のときになった途端に recall が減少している。

図 2(b) 及び図 3(b) は、推定時間を変更した評価の結果である。こちらのグラフでは先ほどのグラフと逆で、線が推定時間を、点が危険度合いを表している。こちらのグラフも危険度合いを評価したグラフと同様に、推定時間が短く危険度合いが低いほど precision が高くなり、危険度合いが高くなると recall が高くなる状況が多く見受けられた。特に図 3(b) の $T = 90$ のときのグラフはその傾向が顕著に表れている。

ハウス A における二酸化炭素 CO_2 での評価結果を図 4 に、ハウス B における二酸化炭素 CO_2 での評価結果を図 5 に示す。

図 4(a) 及び図 5(a) では、危険度合いを変更した評価を示している。図 4(a) をみると危険度合いの設定値により precision の値が大きく変化しているのが分かる。それに対し recall は、 CO_2 の危険値が 200 のときでも 300 のときでも、同じような変化をしていることが分かる。このことより、 CO_2 の危険値よりも、推定時間が recall に大きく関係していると推察される。これは図 4(b) から推察される。

図 4(b) をにおいては、どの推定時間であっても、危険度合いが 200 のときでも 300 のときでも、recall があまり変化してないのがわかる。一方、ハウス B においては同様の傾向は見られない。図 5 をみてもハウス B はハウス A に比べ危険度合いにより recall の変化が大きくなっていることが読み取れる。これは、危険度合いによるサンプル数に、ハウス A とハウス B で差があるためだと思われる。

6. 考察

図 4 のハウス A の飽差 HD での評価結果をみると、危険度合いが $HD > 20$ のときでは $HD > 5$ 、 $HD > 10$ のときに比べて precision と recall が小さくなっていることが危険度の視点及び推定時間の視点から読み取れる。これは、 $HD > 20$ のときは他の場合に比べサンプル数が少ないためであると考えられる。しかし、第 4 節の評価実験で述べたように、植物の光合成には飽差の値は $3 \sim 6 \text{ g/m}^3$ がよいとされている。さらに第 3 節で「誤った危険を予知するよりも、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。」と述べているように、recall はできる限り高いほうがいい。そう考えると $HD > 20$ のときを予測分類する必要性は感じない。高くとも $HD > 10$ となることを予測し手を加えることで飽差を下げることで植物に適した飽差量に保つことができると思われる。また、できるだけ先の予想をしたいと考えたと $T = 10$ では短すぎ、 $T = 30, 60$ のときはあまり recall が変わらなく、 $T = 90$ となると recall が

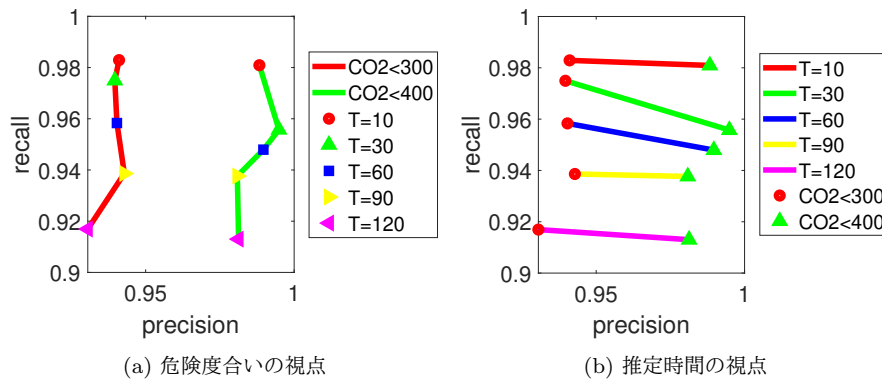


図 4 ハウス A の二酸化炭素 CO₂ での評価結果

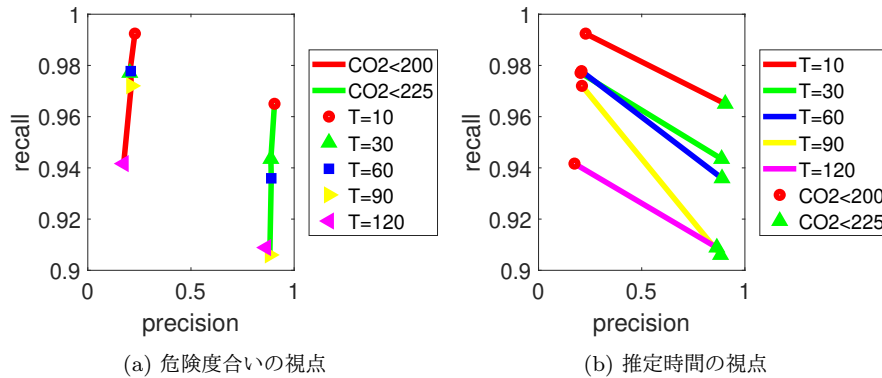


図 5 ハウス B の二酸化炭素 CO₂ での評価結果

小さくなってしまふので $T = 30 \sim 60$ の間が一番適していると思われる。これらをまとめるとハウス A の飽差の予想分類では、 HD が $5 < HD < 10$ で推定時間が $T = 30 \sim 60$ のとき良い結果が得られると考えられる。これは、ハウス B でも同じことがいえる。ハウス B はハウス A に比べ全体的に飽差の値が小さいため危険値の上限が $HD > 10$ となっている。グラフを見ても分かるように $HD > 5$, $HD > 7.5$, $HD > 10$ の場合で recall の変化の幅にあまり違いがみられない。しかし、ハウス B の場合は推定時間の変化による評価は不規則な感じだが $T = 30$ のときに限り、規則的な変化を見せている。このことと、recall ができる限り高い方がいいということを含めて考えるとハウス B の飽差の予測分類は HD が $5 < HD < 7.5$ で推定時間が $T = 30 \sim 60$ のとき良い結果が得られると考えられる。

図 4 と図 5 のハウス A とハウス B の CO₂ の危険度合いによる評価結果を見ると、両方とも危険度合いによる recall への影響は少なく precision への影響が高いことが分かる、逆に推定時間による recall への影響は高く、precision への影響は少ないことが分かる。また、CO₂ 濃度が高いほど光合成速度が高くという事実があるのでビニールハウス内の CO₂ 濃度も出る限り高い値を維持していきたい。これに、推定時間が recall に大きく関わってくるという実験結果を加味すると危険値設定はハウス A では $CO_2 < 400$ 、ハウス B では $CO_2 < 225$ 、推定時間は両ハウスともに $T = 30 \sim 60$ の間で有意なものになると考えられる。

以上の事より、提案手法は農家ごとにより危険値は異なってくるが推定時間が $T = 30 \sim 60$ の間での危険値予測であれば有用であると考えられる。

7. おわりに

本研究では、農業環境データを SVM を用いて分析し危険値を予測分類する手法を提案した。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 を用いたアプローチが有望であることが明らかとなった。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 (C)(No.18K11441) の助成を受けたものである。

参考文献

- [1] 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, 情報処理学会研究報告. MPS, Vol. 2018-MPS-121, No. 1, pp. 1-6 (2018).
- [2] 大塚孝信, 伊藤孝行: ワイヤレスセンサネットワークによる環境情報収集システム実装と異常検知手法, 情報処理学会第 76 回全国大会 (2014).