

科学研究費助成事業 研究成果報告書

平成 28 年 6 月 14 日現在

機関番号：13801

研究種目：挑戦的萌芽研究

研究期間：2014～2015

課題番号：26660198

研究課題名(和文) ウィンドウサイズ自動調整型サポートベクターマシンによる高精度な微気象データ予測

研究課題名(英文) Support vector machine with window-size self-adjusting for high-accuracy predicting of micrometeorological data

研究代表者

峰野 博史 (MINENO, HIROSHI)

静岡大学・情報学部・准教授

研究者番号：40359740

交付決定額(研究期間全体)：(直接経費) 3,000,000円

研究成果の概要(和文)：微気象データは、非常に多数の複雑な要素からなる現象の時系列データであり予測が難しい。本研究では、時系列データ予測に関するSVRをベースとし、予測精度が最大となるようにトレーニングデータのウィンドウサイズを自動調整するSW-SVRを研究開発した。SW-SVRは、現在の環境に特化した予測モデルの構築と再構築を自律的に繰り返すことで、計算量を削減しつつ高精度な予測を実現する。提案システムのプロトタイプ実装として窒素吸収量予測制御システムを実装し現場実証実験の結果、SW-SVRは従来の予測手法やオンライン学習器と比べ、窒素吸収量の予測誤差を大幅に削減できることを確認した。

研究成果の概要(英文)：Micrometeorological data, such as temperature, humidity, and wind speed, has a complicated correlation among different features, and its characteristics change variously with time. In this paper, we propose a new methodology for predicting micrometeorological data, sliding window-based support vector regression (SW-SVR) that involves a novel combination of SVR and ensemble learning. To represent complicated micrometeorological data easily, SW-SVR builds several SVRs specialized for each representative data group in various natural environments, such as different seasons and climates, and changes weights to aggregate the SVRs dynamically depending on the characteristics of test data. We implemented nitrogen absorption amount prediction control system as the prototype system and evaluated the prediction performance. The results demonstrated that SW-SVR reduced remarkably the prediction error of nitrogen absorbed amount compared with conventional machine learning and online learning.

研究分野：総合領域

キーワード：農業気象・微気象 時系列データ予測 機械学習

・ 研究開始当初の背景

様々な自然現象や社会現象に対して、その現象の未来における状況や様子を予測したいという需要は非常に多い。これまで、若手研究(A)、挑戦的萌芽、SCOPE 地域 ICT 振興型研究開発などを通じ、無線センサネットワーク研究の観点から高度農業 ICT を実現するための研究開発を進めてきたが、モデル予測制御システム化する際の予測アルゴリズムに関して、回帰モデルのような統計解析よりも、高度な予測を実現可能な機械学習の適用を試みてきた。一般に、現象を数学的・物理的にモデル化できる場合は、非常に正確な予測が実現可能であるが、降雨予測、気象予測のように気圧や水蒸気量など非常に多数の複雑な要素からなる現象の時系列データ予測は容易ではない。そのため、ANN (Artificial Neural Network) や SVR (Support Vector Regression)を用いた気象予測研究が注目されつつある。

特にカーネルトリックと呼ばれる方法を用いて非線形のパターン関数を構成できる SVM (Support Vector Machine) を回帰に対応させた SVR を採用し、ある地域・期間の実気象データへ適用し、予測精度を ANN と比較するという研究が主流である。ただし、できるだけ少ない計測点数(トレーニングデータ)で精度良い応答曲面を得るには、具体的に、どのような時系列データ予測に対し、どの程度のトレーニングデータが必要で、どの程度の予測時間、予測精度が得られるのかといった関係は試行錯誤という状況である。

2. 研究の目的

時系列データ予測に関して上記 SVR をベースとし、モデル予測制御システムにおける予測アルゴリズムとして組み込めるよう意識した上で、予測精度が最大となるようにトレーニングデータのウィンドウサイズを自動調整する SW-SVR (Sliding Window-based Support Vector Regression) を研究開発する。初年度は、本 SW-SVR の構想を実装し、オープンな気象観測データ等を用いてトレーニングデータのウィンドウサイズ自動調整効果を検証する。どのような時系列データ予測に対し、どの程度のトレーニングデータが必要で、どの程度の予測時間、予測精度が得られるのかの分析を進め、次年度、SCOPE 地域 ICT 振興型等で共同研究を進めてきた静岡県農林技術研究所の施設園芸環境データ(温度、相対湿度、土壌水分量など)に対して、データ収集周期内で高精度なりアルタイム時系列データ予測可能な手法になるか検討を深める。

3. 研究の方法

研究期間を 2 年に設定し、以下の手順で進めることで本研究分野の手法として確立させることを目指した。

i) SW-SVR の設計と実装

図 1 に研究開発する SW-SVR の概要(構想時)を示す。まず、予測したいデータとどれだけ先のデータを予測したいのかを決定し、トレーニングデータを作成する。この作業は既の実測値の得られているデータを用いて作成することができる。次にトレーニングデータの量(サイズ)となる学習期間を決定する。学習期間が長ければ長いほど高精度な時系列データ予測が可能と考えるが、トレーニングに要する時間が増加するため、予測値と将来の実測値が得られるまでの時間内で処理可能なウィンドウサイズを上限に基準を設定する。また、予測結果は、トレーニングデータの質にも依存するため、前述のウィンドウサイズを基準に一定の粒度でサイズを変更させ、異なる学習期間でパラレルに予測値を算出し、予測値と実測値の 2 乗平均平方根誤差 (RMSE) が最少となるウィンドウサイズを次の予測の基準として採用する。この際、異なる学習期間での予測値算出は分散処理可能なため、CPU 負荷状況によっては複数 PC で並列処理させることも検討する。

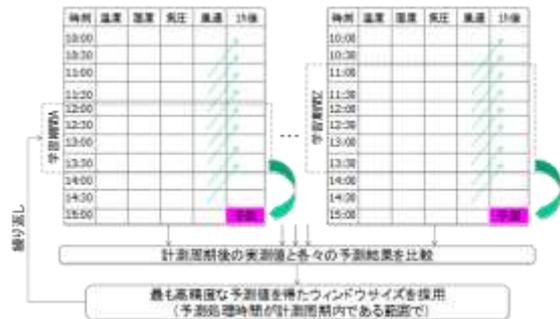


図 1. SW-SVR の概要 (構想時)

以上の処理を繰り返すことで、どのような時系列データ予測に対し、どの程度のトレーニングデータが必要で、どの程度の予測時間、予測精度が得られるのか、といった関係を分析可能な SW-SVR の設計と実装を進める。

ii) SW-SVR の詳細評価

SVM では、ANN と異なり局所解問題は発生しないが、トレーニングデータに応じて最適のカーネル関数を選択する必要がある。一般的には、カーネル関数として Linear, Polynomial, RBF (Radial basis function), Sigmoid の 4 種類、SVR のタイプとして eSVR と vSVR の 2 種類があり、これら組み合わせの中から最適のカーネル関数と SVR のタイプを適用する必要がある。また、SVR のタイプに応じて、調節すべきパラメータがいくつかあり、不適切な値を設定すると過学習に陥りやすい。gamma と cost というパラメータの設定に関して、まずは大雑把な範囲で k-交差検定をグリッドサーチで実施し、最適なパラメータのありそうな範囲を絞ることができたらその範囲で再びグリッドサーチを行うことで、適切なパラメータの組み合わせを探索できると考える。

4. 研究成果

4. 1 概要

微気象データのように経年変化のある時系列データに対して、適切な学習データを自動的に抽出し予測精度が向上するよう自律的に予測モデルを再構築し続けることのできる機械学習アルゴリズム SW-SVR (Sliding Window-based Support Vector Regression) の詳細設計と実装評価を進めた。SW-SVR は、現在の環境に特化した予測モデルの構築と再構築を自律的に繰り返すことで、計算量を削減しつつ高精度な予測を実現する。SW-SVR を用いた施設園芸環境向け知的制御システムのプロトタイプとして窒素吸収量予測制御システムを研究開発した。

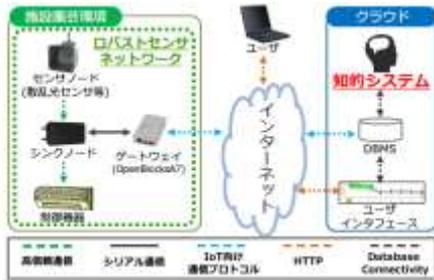


図 2. システムアーキテクチャ

4. 2 システムアーキテクチャ

SW-SVR を用いた施設園芸環境向け知的制御システムのアーキテクチャを図 2 に示す。提案システムは大別して、施設園芸環境下で環境データの収集と制御信号の送信を行うロバストセンサネットワーク (図 2 左) と、クラウド上で SW-SVR を用いて将来の環境を予測し、制御信号を生成する知的システム (図 2 右) から構成される。ロバストセンサネットワークと知的システム間の通信には、Message Queue Telemetry Transport (MQTT) 等の Internet of Things (IoT) 向け通信プロトコルが適用可能である。知的システムや Data Base Management System (DBMS) は施設園芸環境下に配置することもできるが、クラウド上に配置することで高温多湿な現場環境へ設置する電子機器類の負荷を軽減し、またクラウド型サービスとして統合的な予測分析を可能とする。

4. 3 SW-SVR

SW-SVR は、時系列データ予測に特化した SVR の改良アルゴリズムである。SVR は SVM を回帰分析に応用したものであり、多様な分野で優れた認識性能を示すことが知られている学習器の一つである。SVM は、カーネル関数で学習データを高次元空間に写像することで線形分離可能なデータに変換し、その後マージン最大化の基準で識別平面を定めることで、汎化性能の高い学習器を構築する。ただし、自然環境は時間経過で特性が変化するため、SVR で自然環境を高精度に予測するためには、定期的なモデルの再構築が望まれる。しかし、SVR の計算量は学習

データ量 n に対し $O(n^2)$ から $O(n^3)$ であり、計算量削減の工夫を施さずに再構築を行うことは計算量の増大に繋がるため、即時性が求められる機器制御システムへの適用は困難だった。SW-SVR では、時間経過に伴う環境の特性変化に対応させるために自律的な再構築を行いつつ、モデル再構築時には学習データの抽出を行う。学習データの抽出によって、計算量を削減できるだけでなく、ノイズとなり得るデータを除去することで予測誤差も削減できる。そのため、学習データを抽出しない従来の予測手法と比べ高精度な予測が期待できる。



図 3. SW-SVR 概要

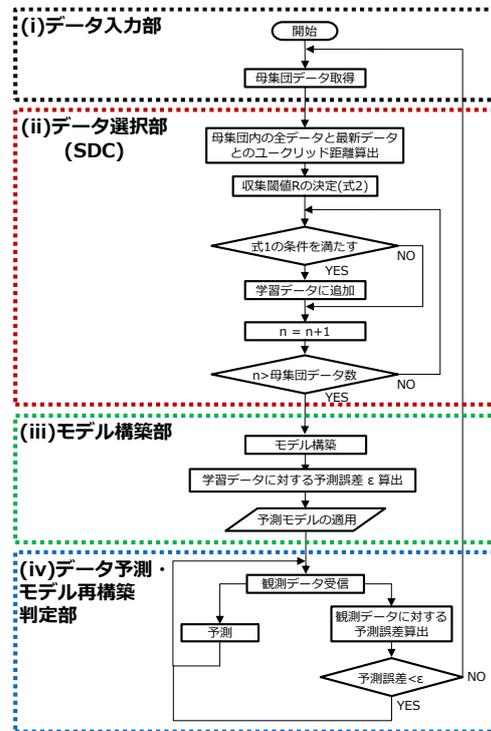


図 4. SW-SVR フローチャート

SW-SVR の概要を図 3 に示す。SW-SVR は、学習データ抽出 (図 3 (1)) と自動再構築 (図 3 (2)) の二つの特徴を持つ。学習データ抽出では、時系列データが持つデータの規則性に注目することで、最新の観測データに関連した最低限の学習データだけを抽出し、予測時点での環境に特化した予測モデルを構築する。自動再構築では、時間経過に伴う環境の特性変化に対応するため、予測モデルの有効性を観測データが観測される度に周期的に判定する。現在使用している予測モデルが、時間経過に伴う環境の特性変化に対応不能と判定した場合、SW-SVR は予測時点

の環境に特化した予測モデルを再構築する。

SW-SVR のフローチャートを図 4 に示す。

(i) データ入力部では、母集団として一定量の学習データを入力する。(ii) データ選択部では、入力された観測データから現時点の予測に必要な最低限の学習データだけを抽出する。ここで抽出した学習データを用いて、(iii) モデル構築部で現在の環境に特化した予測モデルを構築する。(iv) データ予測・モデル再構築判定部では、モデル構築部で構築した予測モデルを用いて予測を行い、同時に現在の予測モデルの予測結果と観測データを比較することで予測誤差を算出する。予測誤差が閾値を上回った時点で予測モデルの再構築を行う。

4. 4 学習データ抽出

SW-SVR の学習データ抽出 (図 3 (1)) には、最新の観測データ G との類似度が収集閾値 R 以内の学習データを、モデル構築に使用する学習データとして抽出する Short-distance Data Collection (SDC) を研究開発した。SDC は、時系列データの遷移特性には規則性が存在し、ある環境とそれに類似した過去の環境は遷移特性も類似することを想定し、SDC で最新の観測データと類似した学習データだけを抽出することで、最新の環境に特化した予測モデルを構築する。

SDC の抽出条件を式 (1) に、収集閾値 R の導出式を式 (2) に示す。ここで、最新の観測データ G の各成分を (g_1, g_2, \dots, g_D) 、 n 番目の学習データ P_n の各成分を $(p_{n1}, p_{n2}, \dots, p_{nD})$ 、説明変数の次元数を D 、母集団のデータ数を N とする。式 (1) と式 (2) では、類似度を表す距離の定義として広く用いられているユークリッド距離を採用した。ただし各次元のスケールの違いを除去するためデータの標準化を実施する。SDC は、式 (1) の条件を満たす P_n を探索し、条件を満たした P_n のみを学習データとして採用する。収集閾値 R の導出式である式 (2) は、母集団の全学習データと最新の観測データとの距離を基とした標準偏差を求める式である。この式が表す標準偏差は、最新の観測データに対する学習データのばらつき大きさに応じて変動するため、式 (2) から学習データ量を決定することで、最新の観測データに対する学習データのばらつき大きさに関わらず、予測に不要な学習データを常に削減する。

$$R > \sqrt{\sum_{i=1}^D (g_i - p_{ni})^2} \quad (1)$$

$$R = \sqrt{\frac{\sum_{n=1}^N \sum_{i=1}^D (g_i - p_{ni})^2}{N}} \quad (2)$$

4. 5 自動再構築

時系列データ予測の場合、予測時間経過後の観測データから予測誤差を算出できるため、予測モデルの有効性の指標として予測誤

差を用いることができる。そのため、SW-SVR の自動再構築 (図 3 (2)) では、予測誤差が一定の閾値を上回ったとき、環境の特性変化に予測モデルが対応不能と判断し、予測モデルの再構築を行う。

再構築の判定を行う閾値の決定方法について述べる。SW-SVR はモデル構築終了後に、モデル構築に用いた学習データに対して予測を行い、予測誤差 ε を算出する。学習データに対する予測誤差 ε は、円内に分布するデータに対する予測誤差と捉えることができるため、もし最新の観測データが円外に位置した場合、最新の観測データに対する予測誤差は ε を上回ると考える。そのため、 ε をモデル再構築の閾値に設定することで、最新の観測データと学習データの特性の差を識別可能となり、環境の特性変化に伴う自律的なモデルの再構築を実現できる。

4. 4 プロトタイプ実装と評価

プロトタイプとして、空气中に散乱している光 (散乱光) の強さを計測する散乱光センサを用いて、制御当日にトマトが吸収する窒素量を予測し、その予測結果に応じた窒素供給制御を行うシステムを開発した。毎日決められた時刻に行われる窒素供給制御に対し、どの程度の量の窒素を供給すべきかをシステムで判断し、適切な量の窒素を供給し続けることで高付加価値な果実栽培を目指した。

本実験の栽培方式であるポット式水耕栽培方式では、給液タンクに接続されているポンプが給液タンク内の溶液を汲み出し、植物が定植されている栽培ベッドへ溶液を流す。栽培ベッドを流下した溶液は再び給液タンク内に格納される。ポンプは実験を実施する農林技術研究所で既に使用されていた機器制御システムが制御し、本提案システムが窒素タンクと給液タンク間に接続されている電磁弁を制御する。本提案システムは予測結果に基づき供給すべき窒素量を決定し、その量を供給するために必要な時間だけ電磁弁を開放する。実験では、植物が根から養液を吸収する時間帯が日中であることに着目し、早朝 6 時に 6 時間後までの窒素吸収量を予測し、予測結果を基に窒素供給制御を実施した。

4. 5 知的システムの実装

知的システムは、学習データ生成部と予測制御部から構成され、SW-SVR は予測制御部で動作する。学習データ生成部では、予測に用いる学習データを定期的に生成する。ただし農業データには、様々な要因によるノイズが含まれており、不正確な学習データの生成を避けるためには、取得時刻の検査、異常値の検査、センサ誤差の除去等の前処理が必要である。また、ポット式水耕栽培では、給液タンク内窒素量の減少量と、植物の窒素吸収量が一致するため、6 時間後の窒素吸収量の予測として、6 時間後までの給液タンク内窒素量の減少量を目的変数とした。給液タンク内窒素量は、給液タンク内水量と給液タンク内窒素濃度の積をとることで算出した。

表 1. 説明変数に利用するデータ

データソース	生成されるデータ
散乱光センサ	気温, 相対湿度, 葉面積指数
EC メータ	給液タンク内窒素濃度
気象庁	磐田市気温, 風速, 降水量, 日射量
日本気象協会	磐田市の予想気温, 予想風速

説明変数として利用するデータの一覧を表 1 に示す. また, 時系列データ予測に対して学習データの抽出を行う場合, 学習データが生成された時間も重要な指標となるため, UNIX 時間も説明変数に加えた. 散乱光センサが生成する気温, 相対湿度, 葉面積指数の三種類のデータから, 窒素吸収と関連する蒸発散量を誤差 3%以内で推定できることを確認済みのため, 上記三種類のデータを説明変数に利用した. また, 本実験の栽培方式では, 植物は給液タンク内の養液から窒素を吸収するため, 給液タンク内の窒素濃度も窒素吸収量と密接に関係することから, 給液タンク内窒素濃度も説明変数に利用した. さらに, 植物の活動は気候に大きく影響を受けるため, 施設園芸環境内の環境だけではなく, 気象庁提供の気象データと日本気象協会提供の天気予報データも説明変数に加えた.

予測制御部では SW-SVR が動作する. ただし本実験での予測制御のため, 図 3 のデータ予測・モデル再構築判定部の仕様を一部変更した. 本来 SW-SVR は観測データが入力されるたびに予測と再構築判定を行うものであるが, プロトタイプ実装では制御時刻である早朝 6 時にのみ予測結果が必要とされるため, 予測は早朝 6 時に入力される観測データに対してだけ行い, 再構築判定はその 6 時間後である 12 時に行う. また, 本実験では予測時刻である早朝 6 時の予測を高精度に実施する必要があるため, 12 時に行われた再構築判定の結果再構築すべきと判断された場合, 翌日の早朝 6 時の観測データに対して SDC を適用させ再構築を行う.

予測後には, 予測結果から制御信号を生成する. まず, 予測結果である 6 時間後までの窒素吸収量の予測値を基に, 制御時刻から 6 時間後までに供給すべき窒素量 (窒素供給量) を決定する. その後, 制御機器である電磁弁に対する制御信号として, 窒素供給量を供給するために必要な電磁弁の開放時間を算出する. 生成された制御信号は現場の電磁弁に即時に送信され, 制御が実施される.

4. 6 評価結果

学習期間と評価期間を表 2 に示す. ただし SW-SVR の各再構築の際には, 再構築時点で既に生成されている評価期間内の学習データも予測モデル構築に利用する. また, 実際の利用を想定した評価のため, 評価期間内のデータを全て予測するのではなく, 予測制御を行う早朝 6 時のデータのみ予測するため, 予測対象の評価データ数は 54 件となる.

表 2. 学習期間と評価期間

	学習期間	評価期間
期間	2014/10/31~2014/11/25	2014/11/26~2015/2/1
日時	6 時~18 時	6 時
データ量	4148 件	54 件

予測誤差の指標には Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percent Error (MAPE) の三種類を用いる. N は評価期間のデータ数, y_i は i 番目の真値, \hat{y}_i は i 番目の予測値を示す.

評価対象は, SVR の他に機械学習のアルゴリズムとして広く利用される Artificial Neural Network (ANN) と Random Forest (RF) を採用した. さらに, SDC と類似した考え方で予測を行う k 近傍法 (k -Nearest Neighbor algorithm : KNN) も比較する. k 近傍法は, 観測データと全ての学習データの距離を計算することで, 観測データと近傍しているデータを k 個抽出し, 抽出データの目的変数の平均値を予測値とするアルゴリズムである. SDC との相違点は, SDC が SVR で利用する学習データを抽出するための手法であることに対し, k 近傍法は抽出した近傍データから直接的に予測値を算出するための手法である. また, オンライン学習器 PA をサポートしており, 農業支援向けアプリケーションとしても利用されている Jubatus も評価対象とした. PA は, 追加学習として予測誤差を許容範囲 ϵ 以内に抑えるように, 予測モデルの重みパラメータを一定の条件下で更新する. PA は, SVR と同様にマージン最大化を採用しており, 重みパラメータを極力小さくするように更新を行う. PA を含むオンライン学習は, 予測モデルの構築の際に全ての学習データを利用せず, 現在の学習データだけを用いて予測モデルの重みパラメータを更新するため, SVR 等のバッチ学習と比べ計算量が非常に少ない.

各学習アルゴリズムにおけるパラメータは, グリッドサーチで事前に準最適解を算出した. チューニング時の予測モデルの評価方法には, ホールドアウト検定を用いた. ホールドアウト検定は, ある一部を評価データとして選び, 残りを学習データとする方法である. 一般的に用いられている k -交差検定と比べ, 学習データと評価データの選び方による偏りの影響を受けやすいが, 学習データと評価データを特定の期間に限定して評価できる. 時系列データ予測を実アプリケーションで利用する際には, 過去の学習データを用いて未来の評価データを評価する. そのため, 学習期間と評価期間を適切に設定することで, データの時系列性を考慮した評価ができる. ホールドアウト検定は, 時系列データ予測の予測モデルを評価する手法として妥当であると考えられる.

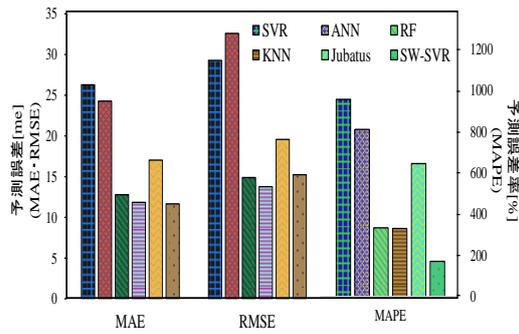


図 5. 予測誤差

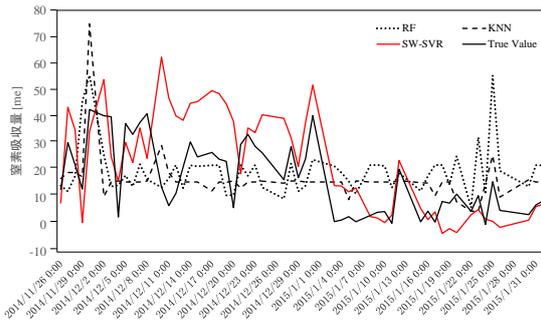


図 6. 予測値と真値

評価期間全体における予測手法毎の予測誤差を図 5 に示す。各予測指標で最小値を示した学習アルゴリズムは、MAE と MAPE では SW-SVR であり、RMSE では KNN であった。一方で、RMSE においても、SW-SVR は平均値以下の予測誤差を示した。RMSE では、KNN, RF, SW-SVR の順に低い予測誤差を示した。一方で、MAPE では、最小値を示した SW-SVR は、KNN や RF の予測誤差を約 50%削減した。誤差指標の違いで最適な予測手法が異なった理由は、各誤差指標は評価期間内の全ての予測誤差を異なる尺度で数値化しているためである。特に、MAE と RMSE は誤差の絶対的な値を扱う一方で、MAPE は相対誤差を用いて予測値が実測値に対して何%誤っているかを表す。そのため、実測値に対する相対的な誤差を示す MAPE は、値が大きく変動する実測値に対して予測値がどれほど追従しているかを示す。農業データは、植物の生長や環境の変化等の時間経過に伴い値が大きく変動するため、長期的な運用を考えた場合には、真値に追従する予測が実現できることが望ましい。そのため、農業データに対する予測手法を評価する際には、真値への追従性を表す MAPE が最も重要視すべき誤差指標であると考えられる。

上記のことを、各予測手法の評価期間内の全ての予測誤差から明らかにするため、評価期間内における各予測手法の予測値と真値を図 6 に示す。ただし、全ての誤差指標において SW-SVR より大きい誤差を示した SVR, ANN, Jubatus は、グラフの煩雑化を避けるため省略する。SW-SVR は、12 月の中旬では予測誤差が増大しているが、その他の期間では値が大きく変動する真値に追従した予

測が実現できている。一方、RF と KNN は、評価期間内の予測値が大きく変動することが少ないため、大幅な予測誤差を示すことは少ないものの、真値に追従した予測が実現できていない。以上から、真値の大きな変動にも追従可能な SW-SVR は、農業データに対する予測に最適な予測手法であるといえる。また、実験終了後に収穫された果実は糖度 8%以上の高糖度で、実験期間中の尻腐れ果等の異常果の発生もほとんどなく、付加価値の高い高糖度トマトを継続的に生産する可能性が明らかとなった。

5. 主な発表論文等

[雑誌論文] (計 2 件)

1. Yukimasa Kaneda, Hiroshi Mineno, “Sliding window-based support vector regression for predicting micrometeorological data,” Expert Systems with Applications, Vol. 59, pp. 217-225 (15 Oct. 2016, IF: 2. 24, Q1).
2. 兼田千雅, 井林宏文, 鈴木雄也, 黒田正博, 大石直記, 峰野博史, “SW-SVRを用いた施設園芸環境向け知的制御システムの評価,” 情処論トランザクション, CDS, Vol. 6, No. 1, pp. 22-32 (2016-05-24).

[学会発表] (計 2 件)

1. Yukimasa Kaneda, Hirofumi Ibayashi, Naoki Oishi, Hiroshi Mineno, “Greenhouse environmental control system based on SW-SVR,” 19th Int’ 1 Conf. on Knowledge-Based and Intelligent Information & Engineering Systems (KES2015), pp. 860-869, 8 Sep. 2015 (Singapore).
2. 兼田千雅, 井林宏文, 大石直記, 峰野博史, “無線散乱光センサと SW-SVR を用いた養液栽培制御システム,” 情処 DICOMO シンポジウム, pp. 433-440, 8. Jul. 2015. (岩手県八幡平市) (優秀プレゼンテーション賞).

[産業財産権]

○出願状況 (計 1 件)

名称: 予測システム, 予測方法, および予測プログラム

発明者: 峰野博史, 鈴木雄也, 兼田千雅

権利者: 静岡大学

種類: 特開

番号: 2016-099738

出願年月日: 平成 26 年 11 月 19 日

国内外の別: 国内

6. 研究組織

(1) 研究代表者

峰野 博史 (MINENO HIROSHI)

静岡大学・情報学部・准教授

研究者番号: 40359740