

コンシューマ・システム論文

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

兼田 千雅^{1,a)} 井林 宏文² 鈴木 雄也²
 黒田 正博³ 大石 直記⁴ 峰野 博史^{1,5}

受付日 2015年10月1日, 採録日 2016年2月23日

概要: 農業分野では、センサネットワークを用いて農作物の栽培環境や生育状況を収集し、そのときの状況に応じた制御を自動的に行うシステムの開発が進められている。しかし、既存システムでは、農業従事者の経験や勘による制御を適切に再現しきれていないという課題があった。そこで本稿では、農業従事者の栽培で実施される予測制御を人の介入なしに高精度かつリアルタイムに再現するため、SW-SVRを用いた施設園芸環境向け知的制御システムを提案する。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、プロトタイプシステムを用いた現場実証実験を実施した結果、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し、栽培後には糖度8%以上の高糖度なトマトを収穫することに成功した。

キーワード: 農業, 環境制御システム, 時系列予測, サポートベクター回帰, 無線センサネットワーク

Evaluation of Agricultural Environmental Control System Using SW-SVR

YUKIMASA KANEDA^{1,a)} HIROFUMI IBAYASHI² YUYA SUZUKI²
 MASAHIRO KURODA³ NAOKI OISHI⁴ HIROSHI MINENO^{1,5}

Received: October 1, 2015, Accepted: February 23, 2016

Abstract: In the agriculture, environmental control systems are becoming more widespread and sophisticated. The systems collect the environmental information using sensor networks and control depending on the situation. However, it is difficult to reproduce the control based on the experience and the intuition of farmers using existing systems. In this paper, we propose agricultural environmental control system using SW-SVR. The proposed system can do farm work like farmers without relying on famers. This system uses SW-SVR that can predict future environment with high accuracy and low calculation cost. We implemented nitrogen absorption amount prediction control system as the prototype system using SW-SVR and evaluated the prediction performance of SW-SVR. The results demonstrated that SW-SVR reduced remarkably the prediction error of nitrogen absorbed amount compared with conventional machine learning and online learning. In addition, the proposed system cultivated high quality tomatoes that have over 8% of sugar content.

Keywords: agriculture, environment control system, time series prediction, support vector regression, wireless sensor network

¹ 静岡大学大学院総合科学技術研究科
 Graduate School of Integrated Science and Technology,
 Shizuoka University, Hamamatsu, Shizuoka 432-8011, Japan

² 静岡大学大学院情報学研究科
 Graduate School of Informatics, Shizuoka University,
 Hamamatsu, Shizuoka 432-8011, Japan

³ 情報通信研究機構
 National Institute of Information and Communications
 Technology, Koganei, Tokyo 184-8795, Japan

⁴ 静岡県農林技術研究所
 Shizuoka Prefectural Research Institute of Agriculture and
 Forestry, Iwata, Shizuoka 438-0803, Japan

⁵ 科学技術振興機構, さきがけ
 JST, PRESTO, Kawaguchi, Saitama 332-0012, Japan

a) kaneda@minelab.jp

1. はじめに

センサ技術の進歩によって様々な環境データを収集可能となり、また収集した環境データを利用して各種機器を制御するシステムもさかんに実用化されつつある。農業分野では、センサネットワークを用いて農作物の栽培環境や生育状況を把握し、そのときの状況に応じた制御を自動的に行うシステムの開発が進められている。一方で、従来システムの多くはシステムユーザが入力した設定値とセンサデータの比較結果に基づき制御が行われる。そのため、システムユーザが季節や天候等の状況に応じて適切に設定値を設定する必要があるが、農業の経験が浅いシステムユーザは、機器制御に関する設定値をつねに適切に設定することは難しい。

農業従事者が実際に行っている制御の1つに、自身の経験や勘をもとに行う予測制御がある。たとえばメロン栽培では、メロンの成長度や当日の環境データから当日メロンが吸収する水分量を予測し、その予測結果をもとに必要な量の灌水を毎朝行うことが望まれる。予測制御を用いることで作物に対して過不足のない制御を実現し、高品質な作物の栽培が期待できる。しかし、時間経過や季節変化とともに特性が激しく変化する環境では、農業従事者が暗黙的に行っている予測をコンピュータで再現することは困難だった。農業データに適用可能な予測手法も多く研究されているが、既存の予測手法は高精度予測と計算量削減の両立が実現されておらず、必要な制御を必要なタイミングで実施するために高い精度と即時性が求められる環境に、既存手法を適用することは難しかった。

本研究では、農業従事者の栽培で実施される予測制御を人の介在なしに高精度かつリアルタイムに再現するため、Sliding Window-based Support Vector Regression (SW-SVR) [1] を用いた施設園芸環境向け知的制御システムを提案する。提案システムは、施設園芸環境下に構築したロバストな双方向多点センサネットワークにSW-SVRを適用したものであり、SW-SVRが予測に基づき制御信号を生成する。SW-SVRは、現在の環境に特化した予測モデルの構築と、特性変化に応じたモデル再構築を自律的に繰り返すSupport Vector Regression (SVR) の改良アルゴリズムであり、詳細は3.3節で述べる。

以降、2章で関連研究について述べる。3章でSW-SVRを用いた施設園芸環境向け知的制御システムを提案し、4章でプロトタイプを実装する。5章で提案システムの評価実験を、6章で追評価を行う。7章で本稿をまとめる。

2. 関連研究

施設園芸環境の様々な環境データをセンシングし、制御機器を自動的に制御する機器制御システムの研究が進められている [2]。従来の機器制御システムを用いて高品質な

作物を栽培するためには、農業従事者の経験や勘をもとに行われる予測制御を高精度に再現し、植物に対して過不足のない制御をリアルタイムに実施することが重要となる。そのため、機器制御システムは、農業従事者が暗黙的に実施している将来予測を機械学習や統計モデルで予測し、現在必要とされる制御を決定すべきである。しかし、農業従事者が予測に用いているデータは、時間経過や季節変化とともに特性が変化する農業データであるため、ある期間の学習データで構築した予測モデルを長期間使用し続けることは予測誤差の増大につながる。そのため、農業データに対して高精度な予測を維持するためには、オンライン学習を採用したり予測モデルを繰り返し再構築したりすることで、つねに最新の環境を予測モデルに学習させ続ける必要がある。

オンライン学習を採用する方法として、SVRをオンライン学習に応用したIncremental and decremental support vector machine learning [3] の理論に基づく手法が提案されている [4]。また、機械学習の分散処理基盤として近年注目されているJubatus [5] では、SVRをオンライン学習に応用したPassive Aggressive (PA) [6] を採用している。つねにデータが観測され続ける時系列データの将来予測を行う場合、ある時刻に観測されるデータ X と、 X から予測時間後に観測されるデータ Y を組み合わせることで、 X を説明変数、 Y を目的変数とした学習データを時間経過とともに生成し続けることができる。そのため、学習データが増大するたびに予測モデルのパラメータを少ない計算量で更新するオンライン学習は、計算量を削減しつつ最新の環境をつねに学習し続けることができ、即時性を維持しつつ高精度な予測が期待できる。

一方で、オンライン学習を含む既存の学習方法では、各学習データに対して学習の必要性の有無を考慮せず、蓄積されている学習データすべてを学習する。自然環境予測の場合、予測対象のデータの特性に応じて学習データを適切に抽出することで、予測誤差が削減できることが分かっている [7]。そのため、さらなる予測精度の向上のためには蓄積されている学習データすべてを学習するのではなく、予測に必要な最低限の学習データのみを用いることが重要であると考えられる。

以上のことから、特性が変化する自然環境を高精度に予測するためには、最新の環境をつねに学習し続けるとともに、予測対象のデータに応じて学習データを適切に抽出する仕組みが重要である。本研究では、予測対象の観測データと類似した学習データを適切な量だけ抽出して予測モデルを構築し、環境の特性変化に応じて予測モデルの再構築を自律的に繰り返す手法を提案する。適切な学習データの抽出を行うことで、予測精度の向上だけでなく、再構築における計算量の削減も実現できるため、即時性が必要とされる環境での予測制御にも適用可能と考える。

3. SW-SVR を用いた施設園芸環境向け知的制御システム

3.1 概要

本研究では、農業従事者の栽培で実施される予測制御を人の介在なしに高精度かつリアルタイムに再現するため、時系列データ予測に特化した SW-SVR を用いた施設園芸環境向け知的制御システムを提案する。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。

提案システムは、予測制御のために施設園芸環境下で確実にデータを収集し、また制御信号を制御機器に確実に送信する必要があるため、施設園芸環境下にロバストな双方向多点センサネットワークを構築する。提案システムは、SW-SVR で計算量を削減しながら高精度な予測を実現しつつ、ロバストな双方向多点センサネットワークでデータの収集と機器の制御を確実に行うことで、農業従事者の栽培で実施される予測制御を人の介在なしに高精度かつリアルタイムに再現する。

提案システムの稼働率について述べる。提案システムはシステムの利用者が限定され、また使用するセンサノードや制御機器の増加等の拡張性を必要とし、さらに栽培期間中の確実なシステム稼働を保証する必要がある。以上から「社会的影響が限定されるシステム」[8] に該当し、99.99%以上の稼働率が必要とされる。ただし、提案システムの稼働率に関する分析は他論文 [9] で記述することとし、本稿は知的制御の内容を中心にまとめる。

3.2 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 1 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 1 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 1 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、

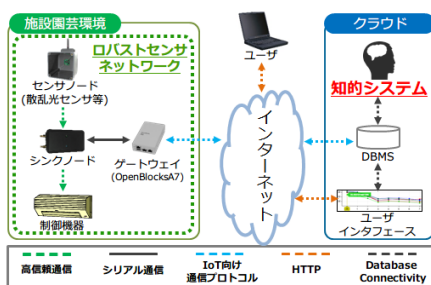


図 1 システムアーキテクチャ
Fig. 1 System architecture.

Message Queue Telemetry Transport (MQTT) [10] 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。

提案システムの動作手順を図 2 に示す。はじめに、ロバストセンサネットワークで環境データを収集し、収集したデータをクラウド上 DBMS に蓄積する。次に、蓄積されたデータをもとに知的システムが SW-SVR で予測を行い、予測結果に基づく制御信号を生成する。最後に、知的システムで生成された制御信号が、ロバストセンサネットワークを経由して現場の制御機器を制御する。

3.3 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 Support Vector Machine (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 1 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 [11]、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つながるため、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、時間経過にともな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。

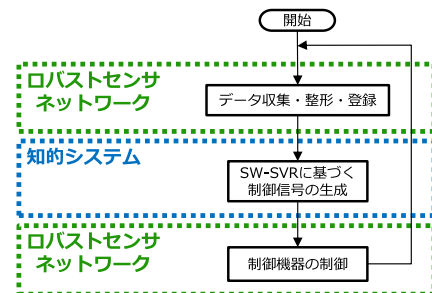


図 2 システム動作フロー
Fig. 2 System operation flow.



図 3 SW-SVR 概要
Fig. 3 Overview of SW-SVR.

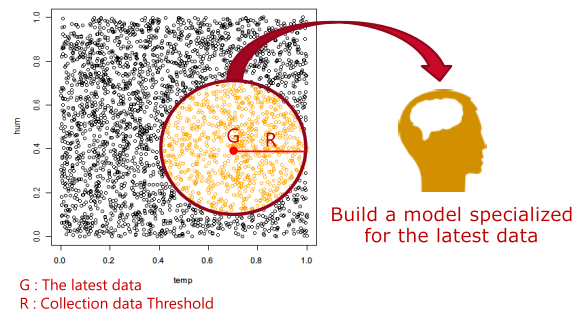


図 5 SDC 概要
Fig. 5 Overview of SDC.

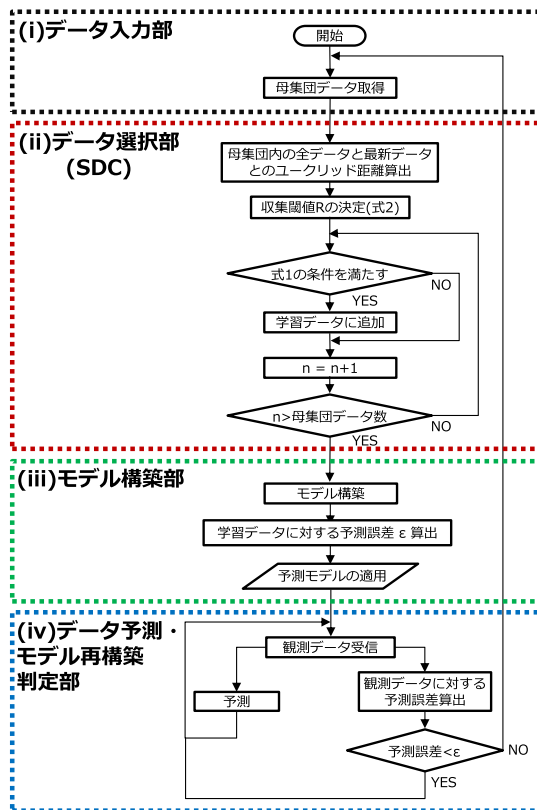


図 4 SW-SVR フローチャート
Fig. 4 Flowchart of SW-SVR.

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 2 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着目することで、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、時間経過とともに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たびに周期的に判定する。現在使用している予測モデルが、時間経過とともに環境の特性変化に対応不能と判定した場合には、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。

SW-SVR の計算量について述べる。モデル構築時の計算量は、データ選択部の学習データ抽出手法の計算量と、モデル構築部の SVR の計算量から決定される。詳細は次節で述べるが、本稿で採用する学習データ抽出手法の計算量は、SW-SVR に入力する学習データ量 n に対して $O(n)$ である。一方、SVR の計算量は、SVR に入力する学習データ量 n' に対して $O(n'^2)$ である。このため、 n 件の母集団データから n' 件の学習データを抽出したときの SW-SVR の計算量は、抽出率が低い場合 ($n > n'$) は学習データ抽出アルゴリズムの計算量である $O(n)$ となり、抽出率が限りなく 100% に近い場合 ($n \approx n'$) は SVR の計算量である $O(n'^2)$ となる。以上より、抽出率が極端に高い場合を除き、SW-SVR は SVR と比べ計算量を削減できる。

3.4 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) を用いる (図 5) [12]。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もと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の大きさに応じて変動する。そのため、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の大きにかかわらず、予測に不要な学習データをつねに削減する。

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

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

3.5 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、図 5 の円内に分布するデータに対する予測誤差ととらえることができるため、もし最新の観測データが図 5 の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、環境の特性変化にともなう自律的なモデルの再構築を実現できる。

4. プロトタイプ実装

4.1 散乱光センサを用いた窒素吸収量予測制御システム

提案システムのプロトタイプとして、空気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、付加価値の高い果実の栽培を目指す。

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。



図 6 使用センサー一覧

Fig. 6 Sensors prototype used.

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すでに使用されていた機器制御システムが制御し、提案システムは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。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。今回の実験では、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、その予測結果をもとに窒素供給制御を実施した。

4.2 ロバストセンサネットワークの実装

プロトタイプでは、無線センサを用いてロバストセンサネットワークを構築した。無線センサは、センサノードや制御機器を任意の位置に設置することを容易にし、システム導入や運用の負担軽減に役立つ。一方で、無線センサを用いる場合、無線通信の品質を低下させる植物の葉や金属パイプ等の障害物の存在を考慮する必要がある。そこで今回の実装では、無線周波数帯を低くし波長を長くすることで障害物への回折性が高まりパスロスを抑えられることに着目し、無線通信に 429 MHz 帯 IEEE802.15.6 方式 [13] を採用した。各センサノードとシンクノード間は、無線モジュールを介して 429 MHz 帯 IEEE802.15.6 方式による無線通信で確実に送受信される。ロバストセンサネットワークの評価として、429 MHz 帯 IEEE802.15.6 方式と、多用途に用いられている 2.4 GHz 帯無線通信を比較したところ、429 MHz 帯 IEEE802.15.6 方式は 2.4 GHz 帯無線通信と異なり施設園芸環境下でも優れたパケット到達率を示した [14]。

プロトタイプ実装で使用したセンサノードを図 6 に、各センサノードの配置図を図 7 に示す。図 6(a) は施設園芸環境内の環境データの計測と植物の成長度 (葉面積指数) の推定を行う散乱光センサであり、気温、相対湿度、光量を収集する。光量センサは筐体内に取り付けられており、直射日光が筐体内に差し込まないように開口部を北側に向けて設置することで、散乱光の強さを計測できる。群落の上下の散乱光を計測することで、推定誤差 3% 以内で葉面積指数を推定可能であるため [15]、散乱光センサは栽培ベッド

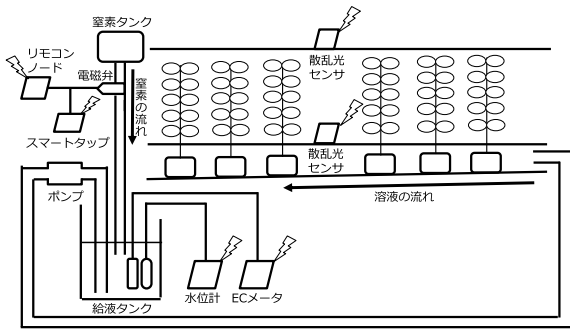


図 7 各種センサ配置図
Fig. 7 Sensors layout.

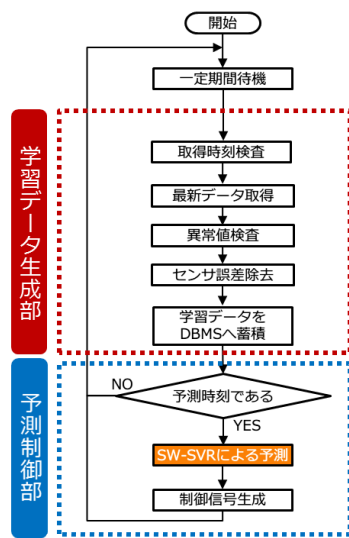


図 8 知的システムフローチャート
Fig. 8 Flowchart of intellectual system.

の上部と下部に設置する。また、給液タンクに関するデータを取得するため、窒素濃度を計測する Electric Conductance (EC) メータ (東亜ディーケーケー製, CM-31P) (図 6(b)) と、水位を収集する水位センサ (クリマテック株式会社製, CWG-H10-C20) (図 6(c)) を用いた。給液タンク内の濃度と水位を計測することで、予測対象である窒素吸収量を算出できる。さらに、制御機器の制御のため、制御機器の消費電力を計測するスマートタップ (図 6(d)) と、制御機器への制御信号を送信する赤外線リモコンノード (図 6(e)) を用いた。赤外線リモコンノードは、知的システムが生成した制御信号をシンクノード (図 1) から受信し、制御信号をもとに適切な赤外線信号を制御機器に対して送信する。

4.3 知的システムの実装

プロトタイプにおける提案システムの知的システム (図 2 中央) の動作フローを図 8 に示す。知的システムは、学習データ生成部 (図 8 上部) と予測制御部 (図 8 下部) から構成され、SW-SVR (図 4) 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

表 1 説明変数に利用するデータ

Table 1 Data that was used as independent variables.

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	磐田市の気温, 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

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。

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 [7], UNIX 時間も説明変数に加えた。

表 1 の各データを説明変数として採用した理由を以下に示す。散乱光センサが生成する気温、相対湿度、葉面積指数の 3 種類のデータからは、窒素吸収と関連する蒸発散量を誤差 3% 以内で推定できるため [15], 上記 3 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 [16] と日本気象協会提供の天気予報データ [17] も説明変数に加えた。

予測制御部では図 4 に示す SW-SVR が動作する。ただし本実験での予測制御のため、図 4 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 時にも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もとに、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。

5. 評価実験

5.1 実験内容

知的システムの評価実験では、4章で述べたプロトタイプを用いて、提案手法 SW-SVR と様々な予測手法で窒素吸収量予測モデルを構築し、SW-SVR の予測精度を評価した。SW-SVR の計算時間に関しては、今回使用可能な学習データが約 4,200 件と少なく、モデル構築の際の計算時間が制御に影響を与えないため、計算時間の評価は省略した。また、実験期間終了後に収穫された作物を評価することで、SW-SVR による予測制御の有用性を評価した。

5.2 実験方法

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、再構築時点ですで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、本実験は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すべ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。

予測誤差の指標には Mean Absolute Error (MAE) (式 (3)), Root Mean Square Error (RMSE) (式 (4)), Mean Absolute Percent Error (MAPE) (式 (5)) の 3 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。

$$MAE = \frac{\sum_{i=1}^N |y_i - \hat{y}_i|}{N} \quad (3)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}} \quad (4)$$

$$MAPE = \frac{100}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (5)$$

評価対象は、SVR のほか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

すべ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 [18]。Jubatus がサポートしている 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すべ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。チューニング対象としたパラメータの一覧を表 3 に示す。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る。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際の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る。

5.3 実験結果・考察

評価期間全体における予測手法ごとの予測誤差を図 9 と表 4 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。

表 3 チューニング対象パラメータ

Table 3 Tuning parameters.

	Parameter1	Parameter2
SVR	Cost	γ
ANN	隠れ層のユニット数	Weight decay
RF	使用する説明変数の数	-
KNN	k	-
Jubatus	sensitivity	regularization_weight
SW-SVR	Cost	γ

表 2 学習期間と評価期間

Table 2 Learning periods and evaluation periods.

	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

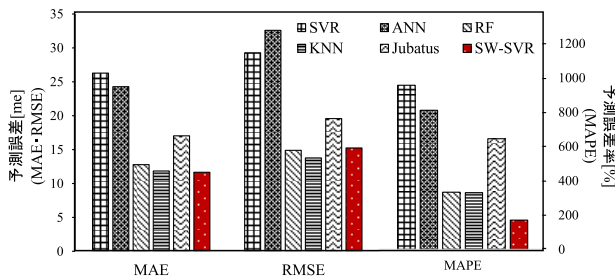


図 9 予測誤差

Fig. 9 Prediction error.

表 4 予測誤差 (詳細)

Table 4 Prediction error (detail).

誤差指標	予測手法						Avg.	Std.
	SVR	ANN	RF	KNN	Jubatus	SW-SVR		
MAE	26.28	24.29	12.80	11.85	17.03	11.67	17.32	6.50
RMSE	29.29	32.59	14.90	13.78	19.58	15.26	20.90	8.10
MAPE	956.95	810.98	333.73	329.71	644.79	170.90	541.18	310.24

農業データに対する予測に最適な予測手法を検討する。表 4 から、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すべ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表す。農業データは、植物の生長や環境の変化等の時間経過とともに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すべ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 10 に示す。ただし、すべ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。図 10 から、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。一方で、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。

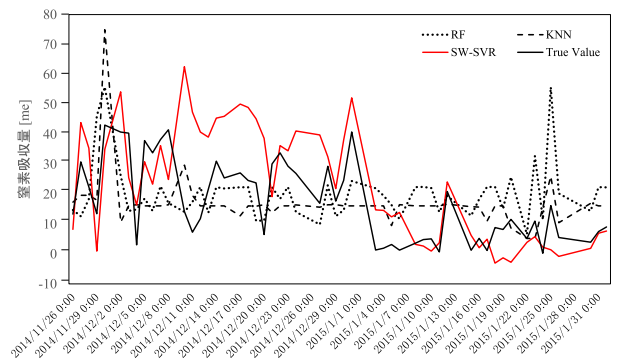


図 10 予測値と真値

Fig. 10 Prediction values and true values.

最後に、SW-SVR による予測制御の有用性を評価する。実験終了後に収穫された果実を評価した結果、提案システムは、糖度が 8%以上の高糖度なトマトを栽培することに成功した。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、提案システムを運用することで、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。

6. 追評価

6.1 実験内容

5 章において、農業データに対する予測精度について検討し、SW-SVR が優れた予測性能を示すことが明らかとなった。一方で、SW-SVR は再構築のためにシステム運用中に予測モデルの再構築を行うため、計算時間に関する検討も行う必要がある。特に、実際の様々なアプリケーションを想定した場合、本稿のプロトタイプより扱うデータ量が多い場合や、予測範囲が 6 時間以外であることも想定される。しかし、プロトタイプを用いて収集した農業データのデータ量は、表 2 に示されるように非常に少ないため、計算時間に関する詳細な分析ができていない。そのため、本章ではオープンデータとして提供されている AMeDAS [16] の気象データを利用し、追評価を行った。ただし、SW-SVR は AMeDAS の気象データに対しても優れた予測精度を示すことが明らかとなっているため [12]、本稿では計算時間に関する評価に限定する。

6.2 実験方法

追評価では、AMeDAS の東京の気象データを用いて、提案手法 SW-SVR と既存手法 SVR の双方で気温の予測モデルを構築し、モデル構築時間を評価した。SW-SVR と SVR を比較することで、再構築の際に学習データ抽出を行わずにすべての学習データを利用する場合と比べ、どれほどモデル構築時間を削減することができるかを分析した。

学習期間は 2011 年から 2013 年の期間内で 1,000 件から 100,000 件の範囲で学習データ量を変化させ、学習期間から以後 1 カ月間を評価期間とした。学習期間と評価期間のデータ収集頻度は 10 分であり、毎日 144 件のデータが取

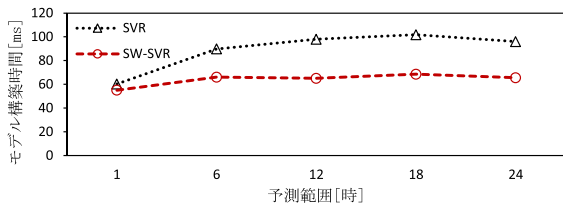


図 11 モデル構築時間 (学習データ 1,000 件)

Fig. 11 Model building time (1,000 learning data).

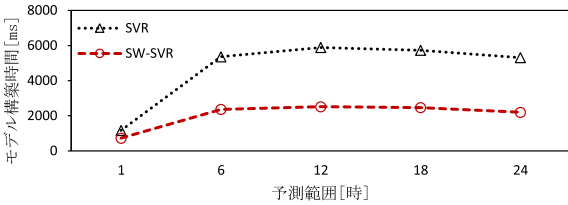


図 12 モデル構築時間 (学習データ 10,000 件)

Fig. 12 Model building time (10,000 learning data).

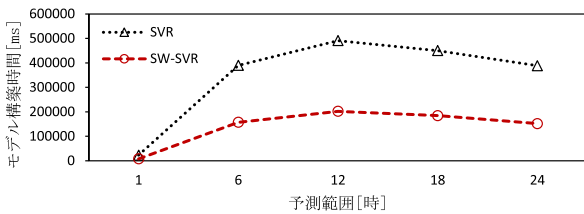


図 13 モデル構築時間 (学習データ 100,000 件)

Fig. 13 Model building time (100,000 learning data).

表 5 モデル構築時間削減率

Table 5 Reduction rate of model building time.

学習 データ量	予測範囲[時]					平均 削減率
	1	6	12	18	24	
1,000 件	8.71%	26.46%	33.67%	32.67%	31.77%	26.65± 10.41%
10,000 件	37.62%	55.88%	57.21%	56.95%	58.60%	53.25± 8.79%
100,000 件	65.77%	59.65%	58.87%	59.05%	60.90%	60.85± 2.86%

集される。予測範囲は、1 時間から 24 時間の範囲で変化させた。

6.3 実験結果・考察

予測精度の評価結果として、学習データ量が 1,000 件のときの結果を図 11 に、学習データ量が 10,000 件のときの結果を図 12 に、学習データ量が 100,000 件のときの結果を図 13 に示す。また、SVR と比較したときの SW-SVR のモデル構築時間削減率を表 5 に、学習データ抽出率を表 6 に示す。SW-SVR は、すべての実験条件下において、つねに SVR より短い予測モデル構築時間を実現した。特に、学習データ量の違いに関しては、学習データ量が多くなるほど、SW-SVR がモデル構築時間を削減できることが

表 6 学習データ抽出率

Table 6 Extraction rate of learning data.

学習 データ量	予測範囲[時]				
	1	6	12	18	24
1,000 件	61.65%	61.86%	61.80%	62.03%	61.73%
10,000 件	61.66%	61.87%	62.26%	61.61%	61.98%
100,000 件	61.31%	61.65%	61.61%	61.80%	60.86%

明らかとなった。この理由は、SW-SVR は学習データの標準偏差だけをもとに学習データを抽出したため、学習データ量が多いほど削減したデータ量も多くなったからである。また、予測範囲の違いに関しては、予測範囲が 1 時間より長くなるほど、SVR、SW-SVR とともにモデル構築時間が長くなる傾向が確認できた。この原因は、予測範囲の長期化にともなって予測が困難になるため、学習データ内に含まれるサポートベクトル (SV) が増大したためと考える。しかし、SW-SVR は予測範囲の長期化にともなうモデル構築時間の増大率も、SVR と比べて緩やかである。そのため、SW-SVR は予測範囲の変化にも柔軟に対応可能と考える。

以上の結果より、学習データ量や予測範囲が異なる様々なアプリケーションを想定した場合でも、SW-SVR は高い予測精度を維持しつつ、再構築の際の計算量を大幅に削減することができるため、様々な応用が可能である。また、SDC の計算量は $O(n)$ と小さいため、SVR 以外のバッチ学習の再構築にも適用可能である。一方で、SW-SVR を用いた場合でも、SVR と同様の増大率ではないが、学習データ量に比例してモデル構築時間が増大する。そのため、厳密な即時性が求められるアプリケーションの場合には、予測モデルを再構築せずに運用するか、より長期間の場合にはオンライン学習を採用するべきと考える。今後、SW-SVR のモデル構築時間についてさらに検討し、適用可能なアプリケーションの整理を進める。

7. おわりに

本研究では、農業従事者の栽培で実施される予測制御を人の介在なしに高精度かつリアルタイムに再現するため、SW-SVR を用いた施設園芸環境向け知的制御システムを提案し、プロトタイプシステムを実装した。提案システムの評価の結果、従来の予測手法やオンライン学習器と比較して、窒素吸収量の予測誤差を大幅に削減し、栽培後には糖度 8% 以上の高糖度なトマトを収穫することに成功した。また、気象データを用いて計算時間の評価を行った結果、今後様々なアプリケーションに応用可能であることが明らかとなった。

今後の課題として、SW-SVR の予測誤差が増大した理由を分析し、さらなる予測精度の向上を目指す。特に、12 月中旬に SW-SVR の予測誤差が他の予測手法と比べ大幅に増大した理由は、SDC が予測に必要な学習データも削減し

たためと考える。SDC は、予測対象データとの類似度をすべての説明変数の値の差分から重み付けせずに算出するため、目的変数と関係度の低い説明変数が与えられた場合、算出された距離は真の類似度を適切に表現できない場合がある。そのため、収集閾値決定方法や説明変数の再検討を行うだけでなく、類似度計算方法の改良として説明変数の要約や重み付けの前処理を適用することで、予測に必要な学習データを、さらに適切に抽出する手法の確立を目指す。

謝辞 本研究は、SCOPE 地域 ICT 振興型研究開発 (H25~26) と挑戦的萌芽研究 (26660198), JST さきがけ (H27~30) により実施したものである。

参考文献

[1] Suzuki, Y. et al.: Proposal to sliding window-based support vector regression, *Knowledge-Based and Intelligent Information & Engineering Systems*, pp.1615-1624 (2014).

[2] Khan, R. et al.: Wireless sensor network based irrigation management system for container grown crops in Pakistan, *World Applied Sciences Journal*, Vol.24, No.8, pp.1111-1118 (2013).

[3] Cauwenberghs, G. et al.: Incremental and decremental support vector machine learning, *Advances in Neural Information Processing Systems 13: Proc. 2000 Conference*, Vol.13, MIT Press (2001).

[4] Ma, J. et al.: Accurate on-line support vector regression, *Neural Computation*, Vol.15, No.11, pp.2683-2703 (2003).

[5] Jubatus: Distributed Online Machine Learning Framework, Jubatus, available from (<http://jubat.us/en/>) (accessed 2015-07-01).

[6] Crammer, K. et al.: Online passive-aggressive algorithms, *The Journal of Machine Learning Research* 7, pp.551-585 (2006).

[7] Suzuki, Y. et al.: Analysis of support vector regression model for micrometeorological data prediction, *Computer Science and Information Technology*, Vol.3, No.2, pp.37-48 (2015).

[8] システム基盤における上流工程での非機能要求合意を目指して、情報処理推進機構, 入手先 (<http://www.ipa.go.jp/files/000004741.pdf>) (参照 2015-07-01).

[9] 井林宏文, 他: 施設園芸環境向け高信頼無線環境制御システムの検討と分析, 情報処理学会論文誌コンシューマ・デバイス&システム, Vol.5, No.4, pp.10-20 (2015).

[10] MQTT, MQTT.org, available from (<http://mqtt.org/>) (accessed 2015-07-01).

[11] 向山 学, 他: 幾何学的学習アルゴリズムによる最大マージン識別法, 信学論 J88-D-II(4), pp.687-694 (2005).

[12] 鈴木雄也, 他: 近傍データ収集法を用いた SW-SVR の改良, 情処研報第 73 回 MBL, pp.1-8 (2014).

[13] Kuroda, M. et al.: Interoperable and Diligent Body Area Networks over IEEE802.15.6 for Real-time Monitoring, *IEEE Engineering in Medicine and Biology Society*, pp.2224-2227 (2011).

[14] Ibayashi, H. et al.: Highly reliable wireless environmental control system for home gardening, *2014 IEEE 3rd Global Conference on Consumer Electronics (GCCE2014)*, pp.725-726 (2014).

[15] 大石直記: 温室環境および植物生態情報取得のための散乱光マルチセンサユニットの開発, 日本生物環境工学会

(2014).

[16] Japan Meteorological Agency, 気象庁, 入手先 (<http://www.jma.go.jp/jma/indexe.html>) (参照 2015-07-01).

[17] Japan Weather Association, 日本気象協会, 入手先 (<https://www.jwa.or.jp/english/>) (参照 2015-07-01).

[18] ビッグデータのリアルタイム分析基盤「Jubatus」を活用し、センサデータ機械学習検証システムを構築, 住友精密工業株式会社, 入手先 (http://www.spp.co.jp/pdf/news/1398318789/1398318789_0.pdf) (参照 2015-07-01).



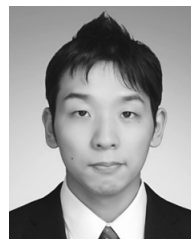
兼田 千雅 (学生会員)

2015 年静岡大学情報学部卒業。同年静岡大学大学院総合科学技術研究科進学。現在センサネットワーク、センサデータマイニング、モバイルコンピューティング、農業支援システムに関する研究に従事。



井林 宏文 (学生会員)

2014 年静岡大学情報学部卒業。同年静岡大学大学院情報学研究科進学。高信頼無線センサネットワークシステムに関する研究に従事。



鈴木 雄也

2015 年静岡大学大学院情報学研究科修士課程修了。同年ソフトバンク入社。センサネットワーク、センサデータマイニング、モバイルコンピューティング、農業支援システムに関する研究に従事。



黒田 正博

1980 年三菱電機入社。2002 年旧独立行政法人通信総合研究所勤務, 以降, 同研究所は情報通信研究機構に統合。2013 年から静岡大学創造科学技術大学院客員教授, 2000 年静岡大学大学院理工学研究科博士課程修了。医療健康分野を中心にボディエリアネットワークの省電力ネットワーク技術と省電力セキュリティ技術の研究開発と実用化に従事。



大石 直記

1987年東京農工大学農学研究科修士課程修了。同年静岡県農業試験場入社。静岡県庁を経て、2002年より静岡県農業試験場勤務。2007年より静岡県農林技術研究所に名称変更。トマト栽培における夏季高温対策，高度環境制御に関する研究に従事。



峰野 博史 (正会員)

1999年静岡大学大学院理工学研究科修士課程修了。同年日本電信電話(株)入社。NTT サービスインテグレーション基盤研究所を経て、2002年10月より静岡大学情報学部助手。2006年九州大学大学院システム情報科学府博士(工学)。2011年4月より静岡大学情報学部准教授。モバイルコンピューティング，センサネットワーク応用システムに関する研究に従事。電子情報通信学会，IEEE，ACM各会員。