

少ない学習データでサポートベクター回帰を用いた排水機場水位予測モデルの可用性 Availability of the water level prediction model at a drainage pumping station for a small amount of training data using support vector regression.

○木村延明*・皆川裕樹*・福重雄大*・馬場大地**

○Nobuaki KIMURA・Hiroki MINAGAWA・Yudai FUKUSHIGE・Daichi BABA

1. はじめに

近年、簡便な準備や迅速な予測の観点から、リアルタイム水位予測への適用が試みられているデータ駆動型の深層学習（例：長・短期記憶，Long short-term memory [LSTM]¹⁾）は、大量のデータを用いることで、高精度の予測が可能である。しかし、データが少ない場合には、予測精度が低下することが知られている。そこで、少ないデータでも、高精度な予測ができる他の機械学習の導入が検討されている。例えば、サポートベクター回帰（Support Vector Regression [SVR]²⁾は、最近の COVID-19 感染者数の時系列予測に³⁾、また、石油掘削に関して小サンプル数によるコスト推定のために利用された⁴⁾ 報告等がある。このように、SVR は少ないデータにも使われている。

我々の研究グループでは、浸水回避や効率的なポンプ操作の支援のために排水機場調整池の水位をリアルタイムで簡易的に予測できる LSTM モデルを開発してきた。しかし、対象とする低平農地の排水機場調整池や排水水路で長期間に渡る大量の水位データを蓄積している地区は非常に少ないために、リアルタイムで利用できる水位予測モデルの普及が困難である。近年の農業分野の ICT 推進のお陰で、観測データの蓄積を行う地区も見られるものの、データ量が少ない場合が多い。このような地区を対象に、少ないデータでも水位予測が可能な機械学習の検討が必要である。なお、ここでいう「少ないデータ」とは、水位データの季節変化の傾向を学習するのが困難な 2~3 年間までと定義する。

本研究では、多分野では少ないデータで予測が可能と考えられる SVR を、データ量の蓄積が少ない地区における排水機場調整池の水位予測手法として採用し、少ないデータでも良好な予測が可能かどうかを検討する（図-1）。



図-1 既往研究と新規研究の手法の概略図

Concept & data flow of our proposed methodology.

2. 方法

本研究の対象地区は水田が主な低平農地である。この地区の排水機場の吐出水槽で計測された水位データは、約 3 年間に渡る 1 時間毎のデータ（約 28,000 点）で、灌漑期（4 月～8 月）と非灌漑期で約 0.5 m の水位差が見られるものである（図-2）。雨量データは、近郊のアメダスデータを利用する。本研究で導入する SVR は、回帰モデルの一種で、線形・非線形事象のデータを扱うことができ、2 つ以上の変数の関係性を調べる手法である。本 SVR では、データに合わせて最適なハイパーパラメータ（ C ：正則化係数=10， ϵ ：不感度係数=0.01，カーネル関数=Radial basis function）を調整した。入力データは、雨量・水位データの過去 6 h～現在（1 h 間隔）までの 7 個を使い、出力は 1 h 間隔で将来の 1 h～6 h の水位である。比較用として、既往研究で開発され、既に最適化されたハイパーパラメータを適用した LSTM モデルを用いる。K 分割交差検証法（ $K=3$ ）を用いて予測精度の評価を行い、二乗平均平方根誤差（RMSE）と Nash-Sutcliffe 係数（NSE）で定量評価を行う。

* 農研機構－農村工学研究部門 The Institute for Rural Engineering, NARO

** (株)アーク情報システム ARK Information Systems, INC.

キーワード 排水機場水位予測，少ないデータ，Support vector regression，深層学習

3. 予測結果

2つの水位予測モデルSVRとLSTMについて、全区間のリードタイム1~6時間の水位予測値と観測値との相関図を図-3に示す。各パネルの線形回帰からのバラツキ具合を確認すれば、リードタイム6hの相関図を除いて、SVRによる予測結果の方がLSTMの結果に比べて良好であることが理解できる。SVRの定量的な精度評価では、NSEで若干の改善、RMSEで6~28%の改善が見られた。特に、リードタイム1hと3hのLSTMの相関図は、S字カーブを示した。この意味は、水位が高い値で予測が過小評価され、低い値で過大評価されたことを示している。つまり、極大値・極小値となるやや極端な値を予測する場合に、LSTMの弱点と考えられる。一方、SVRでは線形回帰に沿って散点図が描かれ、より精度の高い予測が行われたことを示している。以上から、SVRは少ないデータでも相対的に良好な水位予測が可能なが示された。

謝辞：本研究はJSPS科研費JP 21K05838と(財)鹿島学術振興財団研究助成により実施された。また、農研機構の中矢哲郎氏より戦略的イノベーション創造プログラム委託研究の成果の提供を受けた。

参考文献

- Hochreiter, S. and Schmidhuber, J. (1997) Long short-term memory. *Neural Comput.*, 9(8), pp. 1735-1780.
- Drucker, H., Burges, C.C., Kaufman, L., Smola, A.J., and Vapnik, V.N. (1997) Support vector regression machines, in advances in neural information processing systems 9, NIPS 1996, pp.155-161, MIT Press.
- Parbata, D. and Chakraborty, M. (2020) A python based support vector regression model for prediction of COVID19 cases in India. *Chaos, Solitons & Fractals*, 138.
- Kor, K. and Altun, G. (2020) Is support vector regression method suitable for predicting rate of penetration? *J. Petrol. Sci. Engr.*, 194.

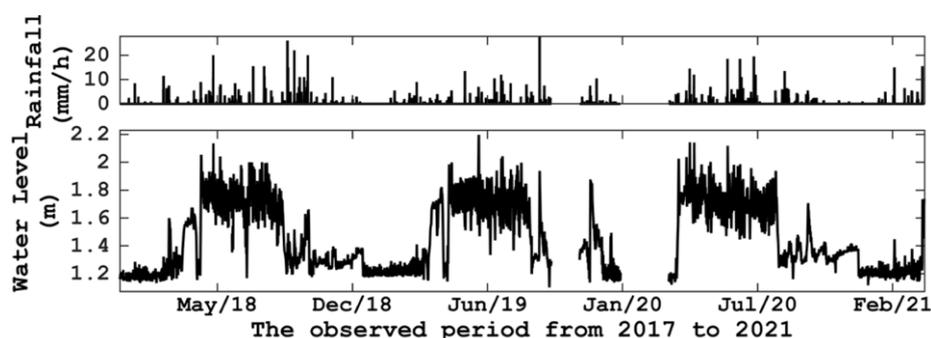


図-2 観測データ (上段：雨量，下段：水位)
Observed data (upper panel: rainfall, under panel: water level).

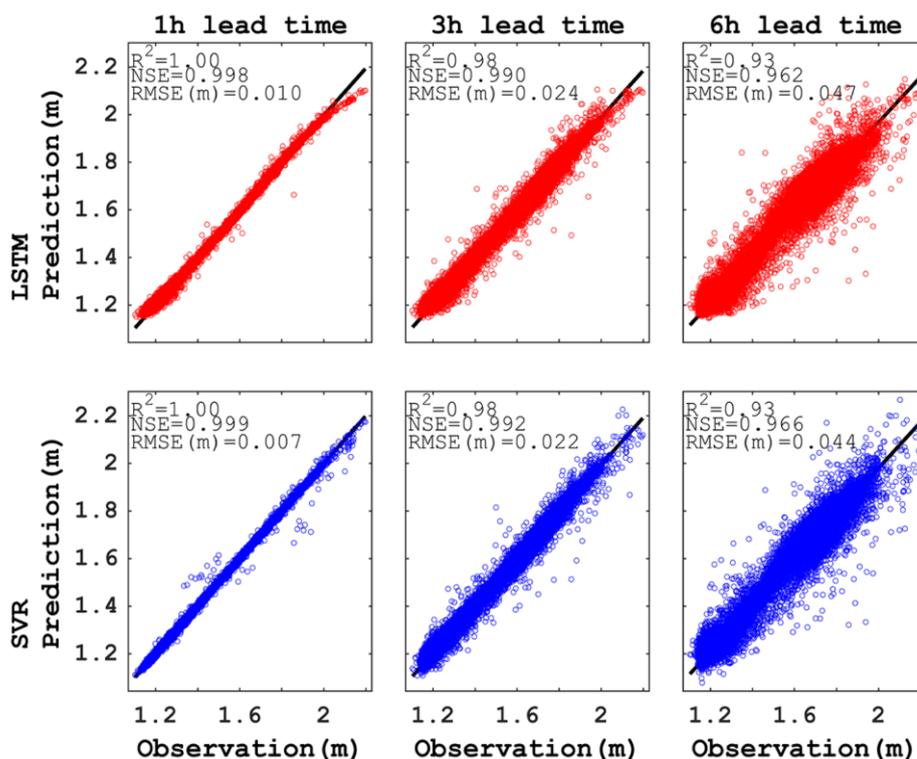


図-3 全区間のリードタイム1h, 3h, 6hの水位予測結果と観測値の相関図 (上段：LSTM, 下段：SVR), 黒線は線形回帰を示す。
Predicted water levels in 1 to 6 h lead times for a whole period vs. observations (upper panels: LSTM, lower panels: SVR) with linear regression (black line).