

ENGINEERING

Prediction of the DO concentration using the machine learning algorithm: case study in Oncheoncheon, Republic of Korea

Heesung Lim¹, Hyunuk An^{1*}, Eunhyuk Choi², Yeonsu Kim^{3,*}

¹Agricultural and Rural Engineering, Chungnam National University, Daejeon 34134, Korea

²Rural research institute, Korea Rural Community Corporation, Ansan 15634, Korea

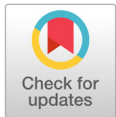
³Korea Water Resources Corporation, Daejeon 34350, Korea

*Corresponding authors: hyunuk@cnu.ac.kr, yeonsu0517@kwater.or.kr

Abstract

The machine learning algorithm has been widely used in water-related fields such as water resources, water management, hydrology, atmospheric science, water quality, water level prediction, weather forecasting, water discharge prediction, water quality forecasting, etc. However, water quality prediction studies based on the machine learning algorithm are limited compared to other water-related applications because of the limited water quality data. Most of the previous water quality prediction studies have predicted monthly water quality, which is useful information but not enough from a practical aspect. In this study, we predicted the dissolved oxygen (DO) using recurrent neural network with long short-term memory model recurrent neural network long-short term memory (RNN-LSTM) algorithms with hourly- and daily-datasets. Bugok Bridge in Oncheoncheon, located in Busan, where the data was collected in real time, was selected as the target for the DO prediction. The 10-month (temperature, wind speed, and relative humidity) data were used as time prediction inputs, and the 5-year (temperature, wind speed, relative humidity, and rainfall) data were used as the daily forecast inputs. Missing data were filled by linear interpolation. The prediction model was coded based on TensorFlow, an open-source library developed by Google. The performance of the RNN-LSTM algorithm for the hourly- or daily-based water quality prediction was tested and analyzed. Research results showed that the hourly data for the water quality is useful for machine learning, and the RNN-LSTM algorithm has potential to be used for hourly- or daily-based water quality forecasting.

Keywords: deep learning, DO (dissolved oxygen), LSTM (long short-term memory), machine learning, water quality prediction



OPEN ACCESS

Citation: Lim H, An H, Choi E, Kim Y. 2020. Prediction of the DO concentration using the machine learning algorithm: case study in Oncheoncheon, Republic of Korea. Korean Journal of Agricultural Science 47:1029-1037. <https://doi.org/10.7744/kjoas.20200086>

Received: October 29, 2020

Revised: November 13, 2020

Accepted: November 20, 2020

Copyright: © 2020 Korean Journal of Agricultural Science



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Introduction

격한 도시화와 경제 발전으로 인해 증가된 오폐수 방류량은 하천의 수질오염 문제를 유발하였으며 악화된 수질을 개선하기 위한 다양한 노력이 시도되고 있다. 하천의 수질 오염 정도를 판단하는 수질인자 중 DO (dissolved oxygen)의 변화는 역사적으로 하천 연구의 주요 대

상이 되어 왔다(Streeter and Phelps, 1925). 일반적으로 하천 수질 오염도를 BOD (biochemical oxygen demand)로 표시하고 있지만 DO는 유기생명체의 광합성, 질소의 질산화 과정, 탄소 순환 및 BOD 등 모든 수질인자들과 밀접한 관계를 갖고 있기 때문에 DO는 수질을 표시하는 주요한 기본인자이다(Wetzel, 1984).

수질예측에 대한 수학적 모형에는 확정론적 방법과 통계학적 특성을 이용한 추계학적 방법 그리고 최근에는 데이터 기반의 인공지능망을 이용한 형이 있으며, 각각 다양한 연구가 수행되었다. 확정론적 방법에서 기존 수리모형은 QUAL2E 모형, WASP 모형, CORMIX 모형, WQRRS 모형 등이 있으며 K-water (1996)은 DO의 경우 QUAL2E, WASP 모형 모두 상류부에서는 실측치보다 다소 작게, 하류부에서는 실측치보다 다소 크게 예측되는 현상을 밝혀냈고, QUAL2E 모형은 계산속도가 빠른 반면, WASP 모형은 계산속도가 느리다는 것을 확인하였다. Box and Jenkins (1970)에 의해 창안된 추계학적인 입출력 시계열 모형은 1974년 Huck and Farquhar (1974)에 의해 수질자료를 활용한 연구가 시작되었고, ARIMA (autoregressive integrated moving average)모형 등이 여러 연구에서 활용되어 왔다(Ahn et al., 2001). Cho and Um (2005)은 고령수질오염측정지점을 대상으로 ARIMA모형을 이용하여 환경오염 지표인 BOD 및 COD, DO, 총질소를 사용하여 수질의 시계열적 특성을 분석하는 연구를 수행하였는데, DO의 경우 실측치와 비교한 결과 ARIMA 모형의 적합성이 높은 것으로 나타났으나, 계절적 요인이 많이 작용하는 것으로 나타났다. Kim et al. (1999)은 ARIMA 모형을 사용하여 진양호의 생물학적 산소요구량(BOD), 부유물질(SS, suspended solids), 총 질소(T-N, total nitrogen)등의 예측을 수행하여 ARIMA 모형의 검증을 실시하였는데, 가까운 시점의 경향을 더 많이 반영하는 ARIMA 모형의 한계점에 의해 장기예측보다는 단기예측의 정확성에 도움이 되는 것을 밝혀냈다. 확정론적 방법, 추계학적 방법에 있어 단기예측에 한정되거나, 기상적인 문제 등이 발생하여 모델 정확도를 높이기 위해서는 지속적인 기상, DO데이터 자료 보충이 필요하다.

인공지능망을 이용한 국내 물관련 연구에서는 수위예측, 강수 예측, 수질예측 등 다양한 분야에서 연구가 수행되었다. Seo et al. (2017)은 기계학습모델인 ANN (artificial neural network), GRNN (generalized regression neural network), ANFIS (adaptive neuro-fuzzy inference system) 및 RF (radio frequency) 모델을 이용하여 저수지 수위 예측 연구를 진행하였는데, 1일 후 수위 예측 결과 ANN, ANFIS 모델이, 5일과 10일 후의 수위예측 결과 ANN, RF모델이 우수한 예측 결과를 나타내면서 종합적으로 판단할 경우 ANN모델이 저수지 예측 모델이 가장 적합하다고 결론지었다. Tran and Song (2017)은 RNNs (recurrent neural networks) 계열의 딥러닝 방법론인 LSTM (long short-term memory)을 활용하여 시계열 침수센서 데이터를 이용하여 수위예측을 수행한 결과 높은 정확도를 얻어낼 수 있었다. Kim et al. (2020)은 미세측 지역이 많은 습지 수위자료에 대하여 다중회귀분석, 주성분회귀분석, 인공지능망, DNN (deep neural network)을 활용하여 수위예측을 하였으며, DNN을 활용한 수위 예측 모형이 일반 인공지능망에 비해 예측성능이 우수한 것으로 나타났다.

Ryu and Jun (2004)은 신경망 모형을 이용하여 갑천유역의 수질예측모형을 구축하고 모형의 적용성을 확인한 결과, 신경망 모형에 의해 구성된 수질예측모형은 비교적 우수한 예측을 하는 것을 확인하였으며, 실시간 수질예측에도 활용 가능할 것으로 판단하였다. Han et al. (2004)은 수년간 관측된 월평균 용존산소(DO)와 수온을 ARIMA 모형, 상태공간모형 및 신경망 모형으로 재현한 결과 신경망 모형의 성능이 가장 우수하였고, 상태공간모형, ARIMA 모형 순으로 높은 정확도를 나타냄을 보인다. 이와 같이 수질분야에 있어서 기계학습의 활용 가능성에 대한 연구 및 검증은 타 물분야와 비교해서 제한적으로 이루어져 왔으며 대부분은 월자료를 이용해서 이루어져 왔다. 이는 타 분야에 비해 비교적 데이터의 생산이 제한적으로 이루어지는 수질 분야의 특성에서 기인한 것으로 판단되나, 최근에는 지속적으로 일별 수질 데이터를 생산하고자 하는 시도를 통해 이러한 한계를 극복할 수 있을 것으로 보인다. Lim et al. (2019)은 시계열 학습에 유리한 RNN 알고리즘과 일자료를 활용하여 하루 후의 BOD, COD, SS를 성공적으로 예측할 수 있음을 보였다. 하지만 Lim et al. (2019)에서는 일자료를 기반으로 하루가 아니라 수일 후의 예측 정확도에 대한 고찰이 없다는 점이 한계로 지적된다.

본 연구에서는 일자료 뿐만 아니라 시자료를 기반으로 시계열 데이터 학습에 특화된 RNN-LSTM 알고리즘을 활용하여 DO예측을 하고자 하였다. 시단위 및 일단위 자료를 기반으로 하여 수시간에서 수일 후 수질예측에의 기계 학습 활용성을 종합적으로 검토하였다. 시단위 및 일단위 DO자료는 부산광역시 보건환경정보 공개시스템(<http://heis.busan.go.kr/>)으로부터 온천천 유역의 부곡교 지점의 자료를 활용하여 연구를 진행하였다.

Materials and methods

기계학습(machine learning)

기계학습은 인공지능 연구의 한 분야로 컴퓨터를 인간처럼 학습시켜 스스로 규칙을 형성하게 하여 학습하고 패턴을 찾아내는 기술 및 기법이다. 기계학습을 구현하는 학습방식에는 지도학습(분류, 회귀 등), 비지도학습(군집화, 압축 등) 등 다양한 학습방식이 있는데, 딥러닝은 기계학습의 한 모델로 기존의 인공신경망의 한계를 뛰어넘은 인공신경망으로 기존 기계학습보다 뛰어난 성능을 보이고 있다. 딥러닝의 핵심은 수많은 데이터 속에서 스스로 규칙을 형성하여 학습하고 패턴을 발견하여 객체를 분별하는 것으로, 지도 학습, 비지도 학습, 강화 학습으로 패턴을 발견해 나가는 과정을 거친다. 딥러닝의 핵심은 학습 데이터의 양으로, 데이터의 양이 많아질수록 학습의 성능이 높아지는 특징을 가지고 있다. 순환신경망(RNN)은 순차적인 데이터 학습에 특화된 인공신경망의 한 종류로, 순환구조가 들어있다는 특징을 가지고 있으며 시계열 학습에 우수한 성능을 가지고 있는 것이 특징이다. Lim et al. (2019)는 순환신경망과 LSTM을 활용한 금강유역의 옥천지점 BOD, COD, SS의 일 예측 연구에서 순환신경망, LSTM 모두 우수한 예측력을 나타내었다. 또한 인공신경망을 활용한 영산강의 수질예측에 대해 우수한 예측력을 나타내었으며(Oh et al, 2002), 인공신경망 이론을 이용한 충주호 수질예측에서도 이상점(out lier)을 제외하면 신경망의 예측력은 전체적으로 양호한 것으로 나타났다(Jeong et al., 2002).

LSTM

순환신경망은 순차적인 데이터 학습에 특화된 인공신경망으로 시계열 학습에 최적화된 우수한 성능을 나타내는데, RNN은 이전의 정보를 현재의 문제 해결에 활용하여 순차적으로 데이터를 처리를 하는 구조로 현재의 문제를 해결하는 중요한 정보를 가진 데이터가 오랜 시간 전에 발생하였을 때 오래전 기억이 소실되어 이를 반영하기 힘든 문제가 있다. 이러한 문제를 장기 의존성(long-term dependency) 문제라고 하는데 이러한 문제점을 해결하기 위해 Hochreier and Schmidhuber (1997)가 LSTM 기법을 제안하였으며, 현재까지 다양한 분야에서 시계열 자료 예측에 활용되고 있다. LSTM은 RNN의 주요 모델 중 하나로 LSTM의 가장 큰 핵심은 장기 의존성 문제를 해결할 수 있는 것이다. 직전 데이터뿐만 아니라 좀 더 과거의 데이터를 고려하여 미래의 데이터를 예측하기 위해 LSTM 알고리즘을 활용하고 있다. LSTM은 Fig. 1과 같이 구조를 나타내는데 특정 시점의 상태(h_t)를 업데이트하기 위해 셀(cell, C_t)이라는 개념을 도입하여 가중치 반영과 활성화함수 변환을 통하여 입력값(input gate, i_t)에서 출력값(output gate, o_t)으로 변환하는 단계를 하나의 셀로 보고 셀 내부의 상태량인 셀 스테이트를 입력, 망각(forget gate, f_t), 출력 게이트를 통하여 가중치 소실문제가 발생하지 않도록 조절한다(Jung et al., 2018). Fig. 1에서 forget gate layer은 다음 식(1)과 같다(Olah, 2015).

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \tag{1}$$

Input Gate Layer는 식(2)와 같이 쓸 수 있고, \tilde{C}_t 는 식(3)과 같이 쓸 수 있다.

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \tag{2}$$

$$\tilde{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c) \tag{3}$$

Update Cell State는 식(4)와 같이 쓸 수 있다.

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \tag{4}$$

Output Gate Layer는 식(5), h_t 는 식(6)과 같이 쓸 수 있다.

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \tag{5}$$

$$h_t = o_t \cdot \tanh(C_t) \tag{6}$$

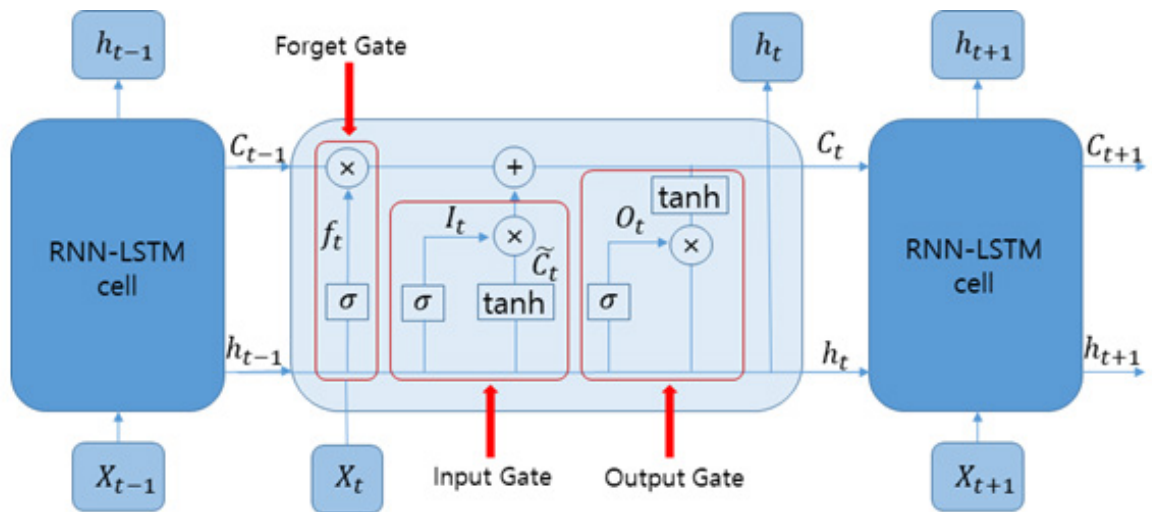


Fig. 1. Structure of recurrent neural network long-short term memory (RNN-LSTM). C_t , cell state; f_t , forget gate; h_t , hidden state; i_t , input gate; o_t , output gate.

Analysis of research data

온천천 유역

부산광역시의 도시하천인 온천천은 수영강의 제1지류로서 부산광역시 금정구 청룡동 금정산에서 발원하여 수영강으로 합류하고, 유로연장은 15.62 km, 유역면적은 55.97 km²이다. 유역의 위치는 동경 129-02-40, 북위 35-09-20 ~ 35-17-30 사이에 위치하고 북쪽으로는 양산시와 경계하면서 양산천 유역과 접하고 있고 서쪽으로는 부산광역시 북구 및 부산진구와 경계를 하면서 덕천천 및 동천유역과 접해 있으며 동쪽으로는 수영강 본류 유역과 접하고 있다. 온천천은 부산광역시의 대표적인 도심하천으로 실시간 자동측정망을 설치하여 수온, 전기전도도, 용존산소, 수소이온농도 등 수질 정보를 실시간으로 제공하고 있다. 본 연구는 부산광역시 금정구 부곡동 79번지에 위치한 부곡교 측정소를 대상으로 DO농도를 예측하였다.

데이터 구축

본 연구에는 부산광역시 보건환경정보 공개시스템으로부터 온천천 유역의 부곡교 지점(Fig. 2)의 2014년 1월 1일 - 2018년 12월 31일까지의 DO농도 자료를 활용하였다. 부산광역시 보건환경정보 공개시스템으로부터 받은 자료 중에서 교정, 보수 중, 비사용, 장비전원단절 등으로 인해 누락데이터가 1,732개의 데이터가 누락되어 있었고, 2014년 1월 1일 - 2018년 12월 31일의 데이터 전수조사 후 이상데이터를 확인하여 다음의 전처리 과정을 통해 보간하였다. 시간단위의 예측 연구를 위해 시간단위로 구축되어있는 데이터에서 12시간 이하 데이터가 누락되었을 때는 선형 보간 하여 데이터를 사용하였고, 12시간 초과되는 데이터는 기간에서 삭제하여 연구를 진행하였다. 연속된 자료 중에서 2015년 3월부터 12월까지의 데이터가 가장 많아 연구에 DO농도 시간 예측 연구에 활용하였고 연구에 있어 기상자료(기온, 상대습도, 풍속, 강수량)를 활용하였는데 기상자료는 국가수자원관리종합정보시스템(<http://www.wamis.go.kr>)에서 수집한 시간단위 자료를 활용하였다. 일 예측 연구를 위해서는 1시간 단위로 되어있는 데이터 중 12개 이상의 데이터를 확보했을 때 12개 이상의 데이터의 평균값을 활용하였고 12개미만의 경우 데이터를 구축하지 않았다. 구축한 1일 데이터 중 10일 이하 데이터가 누락되었을 때 선형 보간 하여 데이터를 사용하고 10일 초과 데이터가 누락되었을 때는 그 기간을 삭제하여 DO농도 일 예측 연구를 진행하였다. 시간 예측과 일 예측 분석을 위해서 input data의 기상자료 유무로 나누어 연구를 진행하였는데 시간 예측을 위해서 input data가 1시간 단위(DO데이터)만 들어간 경우에는 Case H1, 1시간 단위(기온, 상대습도, 풍속, DO데이터)가 들어간 경우는 Case H2, 일예측에서는 1일 단위(DO데이터)만 들어간 경우 Case D1, 1일 단위(기온, 상대습도, 풍속, 강수량 DO 데이터)가 들어간 경우 Case D2로 나누어서 연구를 진행하였다.

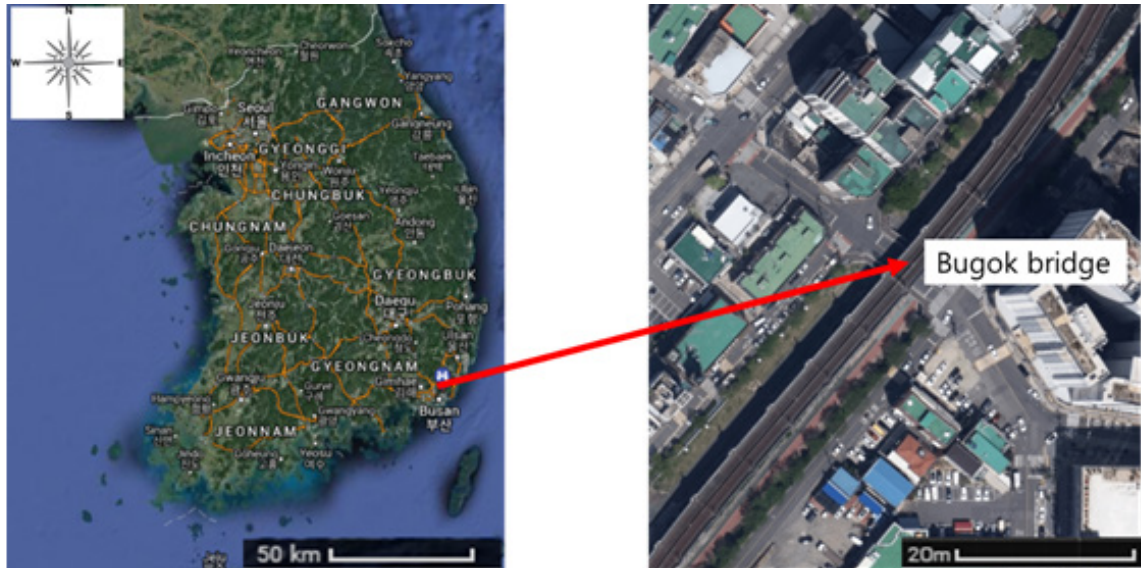


Fig. 2. Study area of Bugok bridge (Oncheoncheon, Busan, Korea).

연구자료 분석

본 논문의 알고리즘은 Anaconda3, TensorFlow를 기반으로 구축되었다. RNN-LSTM 모형의 입력자료로는 DO농도, 기온, 상대습도, 풍속, 강수량자료를 사용하였으며, 본 실험을 위해 LSTM 알고리즘 구성에서 손실함수로는 평균제곱오차를 사용하였고, 최적화 함수로는 AdamOptimizer를 사용하였다. 활성화 함수로는 tanh를 사용하였다. RNN-LSTM 구성에 있어 sequence length는 3, 5, 7로 연구를 수행하였고, Iteration은 3,000, 5,000, 10,000으로 연구를 수행하였다. Hidden layer은 10으로, learning rate는 0.01로 연구를 수행하였다.

Results and Discussion

RNN-LSTM 시간 예측 결과 분석

LSTM 알고리즘으로 Case H1, Case H2를 사용하여 1시간 예측의 경우 Table 1과 같이 Case H1, Case H2 R²은 0.9262 - 0.9396, 0.9234 - 0.9427의 범위를 나타내었으며 높은 정확도를 나타냈고, Case H1, Case H2에서 많은 차이를 나타내지 않았다. 1, 3, 6, 12, 18, 24, 30, 36, 48시간 예측 비교를 위해 예측 성능이 평균적으로 잘 나왔던 sequence length 3, iteration 3,000일 때 1, 3, 6, 12, 18, 24, 30, 36, 48시간 예측을 비교 분석하였다. 1시간 예측의 경우 Case H1, Case H2 R² 0.9331, 0.9427을 3, 6, 12, 18시간 예측에서 3시간 R² 0.6668, 0.6992, 6시간 R² 0.4679, 0.6056, 12시간 R² 0.5424, 0.5868, 18시간 R² 0.5500, 0.5797을 나타내면서 3시간부터는 예측 정확도가 낮아지는 현상을 볼 수 있었다. 그러나 24시간 예측에서는 R² 0.8561, 0.7376을 나타내면서 예측의 정확성이 높아지고 30시간 R² 0.4739, 0.5950, 36시간 R² 0.5220, 0.4801로 점점 예측의 정확성이 떨어지다가 48시간 R² 0.7708, 0.6759로 예측이 정확성이 높아지는 경향을 보였다. 24시간, 48시간 기준으로 예측의 정확도가 상승하였다가 떨어지는 현상을 볼 수 있었는데 예측 시간이 길어질수록 input data의 개수 관계없이 예측 정확도가 낮아지는 것을 확인하였다(Fig. 3).

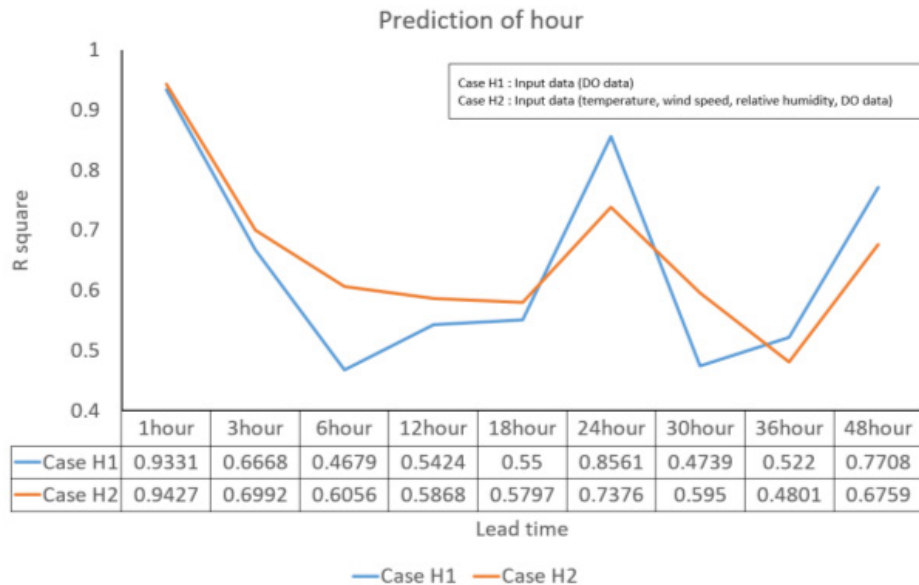


Fig. 3. Hourly prediction results with different lead times. DO, dissolved oxygen.

Table 1. Model performance results of the DO (dissolved oxygen) LSTM (long short-term memory) (prediction 1 hour).

Sequence length	Iterations	Case H1 input data (DO data)		Case H2 input data (temperature, wind speed, relative humidity, DO data)	
		case	R square	case	R square
3	3,000	PH111-1	0.9331	PH114-1	0.9427
	5,000	PH111-2	0.9363	PH114-2	0.9234
	10,000	PH111-3	0.9352	PH114-3	0.9382
5	3,000	PH111-4	0.9396	PH114-4	0.9376
	5,000	PH111-5	0.9383	PH114-5	0.9393
	10,000	PH111-6	0.9335	PH114-6	0.9402
7	3,000	PH111-7	0.9276	PH114-7	0.9366
	5,000	PH111-8	0.9288	PH114-8	0.9372
	10,000	PH111-9	0.9262	PH114-9	0.9375

RNN-LSTM 일 예측 결과 분석

시간 예측과는 다르게 일 예측에서는 연구 대상 지점인 부곡교에서 가장 가까운 지점에서 관측된 강우량 데이터를 Case D2에 추가하여 1일 예측의 경우 Table 2와 같이 Case D1, Case D2 R²은 0.7513 - 0.8284, 0.6989 - 0.8406의 범위를 나타내었으며 Case D1, Case D2에서 뚜렷한 차이를 나타내지 않았다. 시간 예측과 동일하게 sequence length 3, iteration 3000일 때 1, 3, 5, 7, 15일 예측을 비교 분석 하였다. 1일 예측의 경우 Case D1, Case D2 R² 0.8284, 0.8406, 3, 5, 7, 15일 예측에서는 3일 R² 0.6168, 0.6380, 5일 R² 0.5833, 0.6170, 7일 R² 0.5833, 0.6170, 15일 R² 0.5164, 0.5346을 나타내면서 3일부터는 예측 정확도가 낮아지는 현상을 볼 수 있었다(Fig. 4). DO농도 일 예측을 종합적으로 분석해 보면 sequence length에 따른 결과 차이는 없었고, iteration 횟수에 따른 결과는 iteration 횟수가 많아질수록 예측 정확도가 낮아지는 것을 볼 수 있었다.

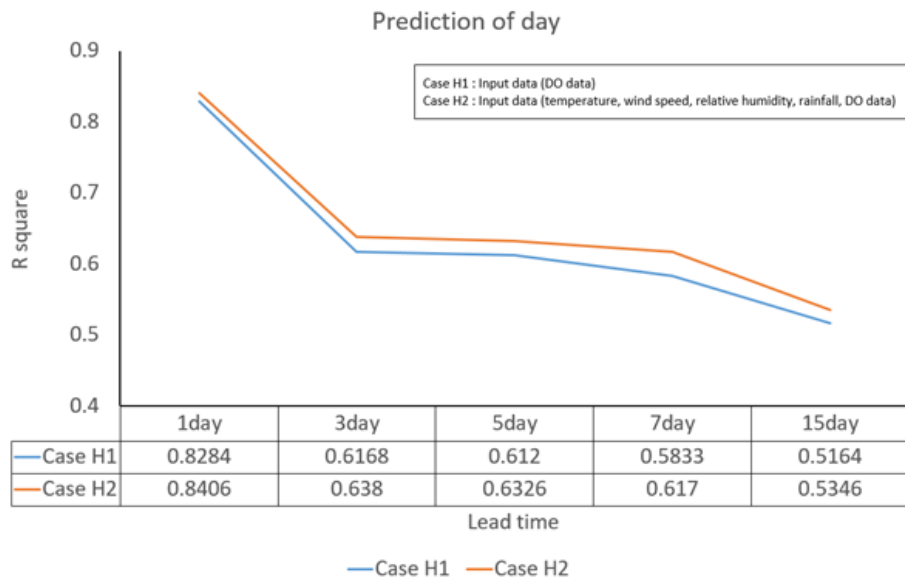


Fig. 4. Daily prediction results with different lead times. DO, dissolved oxygen.

Table 2. Model performance results of the DO (dissolved oxygen) LSTM (long short-term memory) (prediction 1 day).

Sequence length	Iterations	Case D1 input data (DO data)		Case D2 input data (temperature, wind speed, relative humidity, rainfall DO data)	
		case	R square	case	R square
3	3,000	PD511-1	0.8284	PD511-1	0.8284
	5,000	PD511-2	0.8171	PD511-2	0.8171
	10,000	PD511-3	0.7532	PD511-3	0.7532
5	3,000	PD511-4	0.8144	PD511-4	0.8144
	5,000	PD511-5	0.7995	PD511-5	0.7995
	10,000	PD511-6	0.7612	PD511-6	0.7612
7	3,000	PD511-7	0.8036	PD511-7	0.8036
	5,000	PD511-8	0.8030	PD511-8	0.8030
	10,000	PD511-9	0.7513	PD511-9	0.7513

Conclusion

본 연구에서는 RNN-LSTM 알고리즘을 사용하여 시간위 및 일단위 DO 예측에 대한 적용성을 분석하였다. RNN-LSTM 알고리즘을 구현하기 위하여 Anaconda3, TensorFlow를 활용하여 모형을 구성하였고, 시간단위 예측을 위해서는 2015년 3월 - 12월까지의 1시간 단위 데이터를, 일 단위 예측을 위해서는 2014년 1월 1일 - 2018년 12월 31일 1일 단위 데이터를 활용하여 연구를 수행하였다. 모형의 학습과 검증을 위해서 시간단위 예측에는 2015년 3월 - 9월까지(7개월)의 자료를 모형의 학습에 활용하였고, 2015년 10월 - 12월까지(3개월)의 자료에 대하여 검증을 실시하였다. 일 예측에는 2014년 1월 - 2017년 6월(42개월)의 자료를 모형의 학습에 활용하였고, 2017년 7월 - 2018년 12월(18개월)의 자료에 대하여 검증을 실시하였다. 분석한 결과를 정리하면 다음과 같다. RNN알고리즘의 중요한 매개변수인 sequence length가 예측 정확도에 미치는 영향은 시간 예측 및 일별 예측의 경우 모두 크지 않은 것으로 나타났다. Iteration횟수는 시간 예측의 경우에는 정확도에 영향을 크게 미치지 않았으나, 일 예측 overfitting 현상을 나타냈다. 시간 예측과 일 예측을 종합적으로 분석해본 결과 시간 예측, 일 예측 모두 input data 개수에 따른 예측의 정확도는 차이를 보이지 않았다. 시간 예측, 일 예측 모두 예측 시간이 길어질수록 input data의 개수와 관계없이 예측 정확도가 낮아지는 것을 확인하였는데, 시간 예측에서는 24, 48시간 예측에 대해 24시간 주기로 예측 정확도가 높은 것으로 나타났는데, 24시간 예측이 1일 예측보다 정확도가 높았다. 향후 연구에는 시간단위 자료를 활용한 일 예측을 진행할 필요가 있을 것으로 판단되고, 부산 온천천 부곡교 지점만이 아닌 다른 지점의 분석을 통해 비교분석이 필요할 것으로 판단되어, 이에 대한 검증을 수행할 예정이다.

Acknowledgements

본 연구는 농림축산식품부의 재원 농림식품기술기획평가원의 농업기반 및 재해대응기술 개발사업(과제번호:320004-1)의 지원으로 수행되었습니다.

Authors Information

Heesung Lim, Chungnam National University, Ph.D. student

Hyunuk An, <https://orcid.org/0000-0002-4566-5159>

Eunhyuk Choi, Korea Rural Community Corporation, Researcher

Yeonsu Kim, <https://orcid.org/0000-0003-0893-672X>

References

- Ahn SJ, Yeon IS, Han YS, Lee JK. 2001. Water quality forecasting at Gongju station in Geum River using neural network model. *Journal of Korea Water Resources Association* 34:701-711. [in Korean]
- Box GEP, Jenkins GM. 1970. *Time series analysis, forecasting and control*. pp. 109-118. WILEY, San Francisco, USA.
- Cho DH, Um HS. 2005. A study on the forecasting of water quality of Nakdong River - focusing on the Gorung measurement point using the monthly time series data. *Korea Environmental Policy and Administration Society* 13:5-30. [in Korean]
- Han YS, Lee WH, Lee JK, Cho YJ. 2004. Time series analysis using black box model in dissolved. *Proceedings of Korean Geo-Environmental Society* 1:355-362. [in Korean]
- Hochreiter S, Schmidhuber J. 1997. Long short-term memory. *Neural Computation* 9:1735-1780.

- Huck PM, Farquhar GJ. 1974. Water quality models using the Box-Jenkins method. *Journal of the Environmental Engineering. Division* 100:733-751.
- Jeong HJ, Lee SJ, Lee HK. 2002. Water quality forecasting of Chungju Lake using artificial neural network algorithm. *Journal of the Environmental Sciences* 11:201-207. [in Korean]
- Jung SH, Cho HS, Kim JY, Lee GH. 2018. Prediction of water level in a tidal River using a deep-learning based LSTM model. *Journal Korea Water Resour* 51:1207-2016. [in Korean]
- Kim DH, Kim JW, Kwak JW, Kim JS. 2020. Development of water level prediction models using deep neural network in mountain wetlands. *Journal of Wetlands Research* 22:106-112. [in Korean]
- Kim JO, Yoo HH, Kim OS, Park JS. 1999. Forecasting of water quality in Chinyang Reservoir using ARIMA model. *Journal of Wetlands Research* 1:17-28. [in Korean]
- K-water. 1996. A study on the prediction of pollutant advection and dispersion and on the reduction countermeasures in a large river: Focusing on the main stream of Nakdong River. K-water, Daejeon, Korea.
- Lim HS, An HU, Kim HD, Lee JJ. 2019. Prediction of pollution loads in the Geum River upstream using the recurrent neural network algorithm. *Korean Journal of Agricultural Science* 46:67-78. [in Korean]
- Oh CR, Park SC, Lee HM, Pyo YP. 2002. A forecasting of water quality in the Youngsan River using neural network. *Journal of the Korean Society of Civil Engineers* B 22:371-382. [in Korean]
- Olah C. 2015. Understanding lstm networks. GITHUB blog. Accessed in <http://colah.github.io/posts/2015-08-Understanding-LSTMs> on 27 August 2015.
- Ryu BR, Jun KW. 2004. Forecasting of water quality using neural network at Gabcheon. *Journal of Korean Society of Environmental Technology* 5:231-240. [in Korean]
- Seo YM, Choi EH, Yeo WK. 2017. Reservoir water level forecasting using machine learning models. *Journal of the Korean Society of Agricultural Engineers* 59:97-110. [in Korean]
- Streeter HW, Phelps EB. 1925. A study of the pollution and natural purification of the Ohio River. pp. 1-75. Public Health Service, Washington, D.C., USA.
- Tran QK, Song S. 2017. Water level forecasting based on deep learning: A use case of Trinity River-Texas-The United States. *Journal of Korean Institute of Information Scientists Engineers* 44:607612. [in Korean]
- Wetzel. 1984. *Limnology* 2nd edition. Harcourt Brace Jovanovich Colleague Publishers, San Diego, USA.