

# ANN-XGB를 이용한 수중 산소 농도 예측

조광현<sup>1\*</sup> · 이근영<sup>2</sup>

<sup>1</sup>군산대학교 · <sup>2</sup>세종대학교

## ANN-XGB based predictions of dissolved oxygen

Gwanghyun Jo<sup>1\*</sup> · Keun Young Lee<sup>2</sup>

<sup>1</sup>Kunsan National University · <sup>2</sup>Sejong University

E-mail : gwanghyun@kunsan.ac.kr / defg@effff.kmaritime.ac.kr

### 요 약

하천의 용존 산소량은 어류의 생존과 관계되는 중요한 수질 정보중 하나이다. 안양천에서 수집된 수질 정보 및 기상 정보를 이용하여 artificial neural network - XGBoost (ANN-XGB)를 훈련하였으며, 12 시간 이후의 DO를 예측할 수 있다. 본 발표에서는 ANN-XGB의 performance를 보고한다.

### ABSTRACT

The dissolved oxygen (DO) is one of the factors of ecosystem that affects survival of aquatic life. Artificial neural network - XGboost (ANN-XGB), which was trained by water quality and weather data obtained at Anyang-streamto, was employed to forecast DO after 1 hours. We document the performnce of ANN-XGB.

### 키워드

XGBoost, Artifical Neural Network, Dissolved Oxygen, Anyang Stream

## I. 서 론

하천의 용존 산소량 (dissolved oxygen, DO)는 하천 생태계의 수질을 판단하는 주요한 요소 중 하나이다 [1-3,8]. DO를 예측하는 것은 어류의 집단 폐사를 막는데 사용될 수 있기 때문에, DO를 예측하기 위한 많은 연구들이 진행되고 있다. 본 연구에서는 DO를 예측하기 위한 새로운 방법을 제시한다. Artificial neural network (ANN)을 이용하여, DO의 시계열 데이터를 먼저 예측한다. 그 후 기후 데이터들을 추가하여 XGBoost를 사용하여 최종 DO의 예측을 얻는다. 데이터는 안양천에서 수집된 수질데이터와 기후 데이터를 사용하였다.

## II. 방법

### II A. 데이터

데이터는 안양천에서 2012.1~2012.4 동안 얻어진

수질 데이터와 기상 데이터를 사용하였다. 수질 데이터는 수온, turbidity, DO, conductivity를 포함하며, 기상 데이터는 풍속, 기온, 방사량, 수위를 포함한다.

### II B. ANN-XGBoost

ANN-XGBoost는 두 단계로 구현된다. 먼저 ANN을 훈련하기 위해, 시계열 데이터인 DO를 기준 시간 t에 대해서 다음과 같은 형태로 나타낸다.

Input:  $DO_{t-168}, DO_{t-167}, \dots, DO_t$

Output:  $DO_{t+1}$

즉, 고정된 시간을 기준으로 7일 전까지의 DO를 이용하여 1시간이후의 DO를 예측하는 형태의 데이터이다. 이렇게 만든 총 5184개의 sample을 ANN으로 훈련한다.

ANN을 통하여 예측된 DO를 임시 예측 DO값으로 하고, 이 임시 예측 DO의 정확도를 높이기 위

\* corresponding author

해서, XGB를 사용한다. XGB는 tree 기반 ensemble 알고리즘으로, 매우 빠르게 학습이 될수 있다는 장점이 있다. 또한 각각의 input variable들을 normalize하지 않아도 된다는 장점이 있기 때문에, 다양한 형태의 data를 다루는 본 연구에 사용하기가 적합하다 [4-7]. XGB에는 input variable로 수온, turbidity, DO, conductivity, 풍속, 기온, 방사량, 수위 등이 포함된다.

### III. 결과

ANN은 두 개의 hidden layer가 있고, 각각의 hidden layer의 node수가 80, 40인 구조를 사용하였다. XGB의 경우, 400개의 tree를 사용하였다. 어패류가 죽게되는 critical DO는 2로 알려져 있다. ANN-XGB를 통하여 1시간 이후의 DO가 2이상인지 이하인지를 구분한 결과 94%의 정확도인 것을 확인할 수 있었다.

또한 XGB를 통하여 tree의 branch-division에서 어떤 feature들이 많이 사용되는지를 count하여 feature importance를 분석하였다. 분석 결과, 수위와 물의 온도가 DO에 큰 영향을 끼친다는 것을 확인할 수 있었다.

### IV. Discussion

본 연구에서 ANN과 XGB를 결합하여 DO를 예측하는 알고리즘을 소개하고 결과를 보고하였다. ANN-XGB를 통하여 DO를 미리 예측하여 어류 폐사를 예방하는 응용분야가 있을 것으로 기대한다.

### Acknowledgement

이 논문은 NRF-2020R1C1C1A01005396의 지원을 받음.

### References

- [1] V. Z. Antonopoulos, and S. K. Gianniu, Simulation of water temperature and dissolved oxygen distribution in lake vegoritis, greece, *Ecological Modelling*, 160 (2003), pp. 39-53.
- [2] C. Bishop, *Neural networks for pattern recognition*: Oxford university press, New York, (1996).
- [3] J. Bowers, and C. Shedrow, Predicting stream water quality using artificial neural networks (ann), *WIT Transactions on Ecology and the Environment*, 41 (2000).
- [4] T. Chen, and C. Guestrin, Xgboost: A scalable tree boosting system, in *Proceedings of the 22nd acm sigkdd international conference on knowledge*

- discovery and data mining, *ACM*, 2016, pp. 785-794.
- [5] Z. Chen, F. Jiang, Y. Cheng, X. Gu, W. Liu, and J. Peng, Xgboost classifier for ddos attack detection and analysis in sdn-based cloud, in *2018 IEEE International Conference on Big Data and Smart Computing (BigComp)*, IEEE, 2018, pp. 251-256.
- [6] J. H. Friedman, Greedy function approximation: a gradient boosting machine, *Annals of statistics*, (2001), pp. 1189-1232.
- [7] M. Gumus, and M. S. Kiran, Crude oil price forecasting using xgboost, in *2017 International Conference on Computer Science and Engineering (UBMK)*, IEEE, 2017, pp. 1100-1103.
- [8] M. I. Hejazi, X. Cai, and B. L. Ruddell, The role of hydrologic information in reservoir operation-learning from historical releases, *Advances in water resources*, 31 (2008), pp. 1636-1650.