Transformer를 이용한 유해남조 발생 예측 모델 구축
Building of cyanobacteria forecasting model using transformer

이한규*, 김진휘**, 변서현***, 신재기****, 박용은*****
Hankyu Lee, Jin Hwi Kim, Seohyun Byeon, Jae-Ki Shin, Yongeun Park

요 지

팔당호는 북한강과 남한강이 합류하여 생성된 호소로 수도인 서울과 수도권인 경기도 동부지역의 물 공급을 담당하는 중요한 상수원이다. 이러한 팔당호에서 유해남조 발생은 상수원수 활용과 직접적으로 연관되어 있어 신속하고 정확한 관리 및 예측이 필요하다. 본 연구에서는 안전한 상수원 활용을 위해, 딥러닝 기법을 이용하여 유해남조 사전 예측 모델을 구축하고자 하였다. 모델 입력 변수는 2012년부터 2021년까지 10년 동안의 주간 팔당호 수질(수온, DO, BOD, COD, Chl-a, TN, TP, pH, 전기전도도, TDN, NH4N, NO3N, TDP, PO4P, 부유물질)과 수문(유입량, 총방류량), 기상 정보(평균기온, 최저기온, 최고기온, 일 강수량, 평균풍속, 평균 상대습도, 합계일조량), 그리고 북한강과 남한강 유입지점의 남조 세포 수를 사용하였다. 모델 출력 변수는 수질, 수문, 기상 요인으로 인한 남조의 성장 발현 시기를 고려하여 1주 후의 댐앞 남조 세포수를 사용하였다. 사용한 딥러닝 기법은 최근 주목받고 있는 Temporal Fusion Transformer (TFT)를 사용하였다. 모델 훈련용 데이터와 테스트용 데이터는 각각 8:2의 비율로 나누었으며, 검증용 데이터는 훈련용 데이터 내에서 훈련 데이터와 검증 데이터를 6:4 비율로 분배하였다. Lookback은 5로 설정하였고, 이는 주단위 데이터로 구성된 데이터세트의 특성을 반영한 것이다. 모델의 성능은 실측값과 예측값을 토대로 R-square와 Root Mean Squared Error (RMSE)를 계산하여 평가하였다. 모델학습은 총 154번 반복 진행되었으며, 이 중 성능이 가장 준수한 시점은 54번째 반복 시점으로 훈련손실 대비 검증손실이 가장 양호한 값을 나타냈다(훈련손실:0.443, 검증손실 0.380). R-square는 훈련단계에서 0.681, 검증단계에서 0.654였고, 테스트 단계에서 0.606으로 산출되었다. RMSE는 훈련단계에서 0.614(μg/L), 검증단계에서 0.617(μg/L), 테스트 단계에서 0.773(μg/L)였다. 모델에 사용한 데이터세트가 주간 데이터라는 특성을 고려하면, 소규모 데이터를 사용하였음에도 본 연구에서 구축한 모델의 성능은 양호하다고 평가할 수 있다. 향후 연구에서 데이터세트를 보강하고 모델을 업데이트한다면, 모델의 성능을 더욱더 개선할 수 있을 것으로 기대된다.

핵심용어 : Transformer, 딥러닝, 남조세포수 예측, Attention, 팔당호

감사의 글

본 연구는 한강수계관리위원회 환경기초조사사업의 지원을 받아 수행되었습니다.

* 정회원 · 건국대학교 사회환경플랜트공학과 박사수료 · E-mail : haeckel@konkuk.ac.kr
** 정회원 · 건국대학교 사회환경공학부 · E-mail : jinhwi25@naver.com
*** 정회원 · 건국대학교 사회환경플랜트공학과 석사과정 · E-mail : shbyeon@konkuk.ac.kr
**** 일반회원 · 한국수자원공사 부산권지사 수석연구원 · E-mail : shinjaeki@gmail.com
***** 정회원 · 건국대학교 사회환경공학부 부교수 · E-mail : yepark@konkuk.ac.kr