

변수 선택 및 샘플링 기법을 적용한 조류 경보 단계 예측 모델의 정확도 개선

Environmental variable selection and synthetic sampling methods for improving the accuracy of algal alert level prediction model

김진휘*, 이한규**, 변서현***, 신재기****, 박용은*****

Jin Hwi Kim, Hankyu Lee, Seohyun Byeon, Jae-Ki Shin, Yongeun Park

요 지

현재 우리나라에서는 4대강 및 주요 호소 29지점을 대상으로 조류경보제가 시행되고 있으며 조류 경보 단계는 실시간 모니터링지점에서 측정되는 유해 조류의 셀농도를 기반으로 발령 단계가 결정된다. 상수원 구간은 관심, 경계, 조류 대발생, 해제 또는 미발생 총 4구간으로 구성되며, 친수 활동 구간의 경우 조류 대발생을 제외한 3구간으로 구성된다. 현재 시행되는 조류 경보제의 목적은 유해 조류 발생 시 사후 대응 방안 마련에 보다 초점이 맞춰져 있으며 특히, 모니터링 주기 확대 여부, 오염원 관리 방안 마련, 조류 제거 여부 등의 의사 결정 수단으로 사용되고 있다. 하지만 조류 경보 단계에 대한 사전 예측이 가능한 경우 유해 조류의 성장을 억제할 수 있으며 이를 통해 안전하고 깨끗한 수자원을 확보할 수 있다. 본 연구에서는 조류 경보 단계의 사전적 예측을 위해 국가 실시간 측정망에서 제공하는 전국 보 모니터링 종합 정보 자료, 기상측정망 자료, 실시간 보 현황 자료를 활용하여 예측 모델을 구축하였다. 또한, 단계 예측의 정확도를 개선하기 위해 변수 선택 기법을 활용하여 조류 경보 단계에 영향을 미치는 환경변수를 선정하였으며 자료의 불균형으로 인해 모델 학습 과정에서 발생하는 예측 오류를 최소화하기 위해 다양한 샘플링 기법을 적용하여 모델의 성능을 평가하였다. 변수 선택 및 샘플링 기법을 고려하지 않은 원자료를 사용하여 예측 모델을 구축한 결과 관심 단계(Level-1) 및 경보 단계(Level-2)에 대해 각각 50%, 62.5%의 예측 정확도를 보인 반면 비선형 변수 선택 기법 및 Synthetic Minority Over-sampling Technique-Edited Nearest Neighbor(SMOTE-ENN) 샘플링 기법을 적용하여 구축한 모델에서는 Level-1은 85.7%, Level-2는 75.0%의 예측 정확도를 보였다.

핵심용어 : 조류경보제, machine learning, synthetic data sampling, feature selection

감사의 글

본 연구는 한강수계관리위원회 환경기초조사사업과 농림축산식품부의 재원으로 농림식품기술기획평가원의 농업기반및재해대응기술 개발 사업(320049-5)의 지원을 받아 수행된 연구입니다.

* 정회원 · 건국대학교 공과대학 사회환경공학부 연구교수 · E-mail : jinhwi25@gmail.com

** 정회원 · 건국대학교 공과대학 사회환경플랜트공학과 박사과정 · E-mail : haeckel@konkuk.ac.kr

*** 정회원 · 건국대학교 공과대학 사회환경플랜트공학과 석사과정 · E-mail : shbyeon1@gmail.com

**** 일반회원 · 한국수자원공사 부산권지사 수석연구원 · E-mail : jaekishin@kwater.or.kr

***** 정회원 · 건국대학교 공과대학 사회환경공학부 교수 · E-mail : yepark@konkuk.ac.kr