머신러닝 기법을 이용한 유량 자료 생산 방법 Estimation of River Flow Data Using Machine Learning

강노을*, 이지훈**, 이정훈***, 이충대**** Noel Kang, Ji Hun Lee, Jung Hoon Lee, Chungdae Lee

요 지

물관리의 기본이 되는 연속적인 유량 자료 확보를 위해서는 정확도 높은 수위-유량 관계 곡선식개발이 필수적이다. 수위-유량 관계곡선식은 모든 수문시설 설계의 기초가 되며 홍수, 가뭄 등물재해 대응을 위해서도 중요한 의미를 가지고 있다. 그러나 일반적으로 유량 측정은 많은 비용과시간이 들고, 식생성장, 단면변화 등의 통제특성(control)이 변함에 따라 구간분리, 기간분리와 같은비선형적인 양상이 나타나 자료 해석에 어려움이 존재한다. 특히, 국내 하천의 경우 자연적 및인위적인 환경 변화가 다양하여 지점 및 기간에 따라 세밀한 분석이 요구된다.

머신러닝(Machine Learning)이란 데이터를 통해 컴퓨터가 스스로 학습하여 모델을 구축하고 성능을 향상시키는 일련의 과정을 뜻한다. 기존의 수위-유량 관계곡선식은 개발자의 판단에 의해 데이터의 종류와 기간 등을 설정하여 회귀식의 파라미터를 산출한다면, 머신러닝은 유효한 전체데이터를 이용해 스스로 학습하여 자료 간 상관성을 찾아내 모델을 구축하고 성능을 지속적으로 향상 시킬 수 있다. 머신러닝은 충분한 수문자료가 확보되었다는 전제 하에 복잡하고 가변적인수자원 환경을 반영하여 유량 추정의 정확도를 지속적으로 향상시킬 수 있다는 이점을 가지고 있다.

본 연구는 머신러닝의 대표적인 알고리즘들을 활용하여 유량을 추정하는 모델을 구축하고 성능을 비교·분석하였다. 대상지역은 안정적인 수량을 확보하고 있는 한강수계의 거운교 지점이며, 사용자료는 2010~2018년의 시간, 수위, 유량, 수면폭 등 이다. 프로그램은 파이썬을 기반으로 한 머신러닝라이브러리인 사이킷런(sklearn)을 사용하였고 알고리즘은 랜덤포레스트 회귀, 의사결정트리, KNN(K-Nearest Neighbor), rgboost을 적용하였다. 학습(train) 데이터는 입력자료 종류별로 조합하여 6개의 세트로 구분하여 모델을 구축하였고, 이를 적용해 검증(test) 데이터를 RMSE(Roog Mean Square Error)로 평가하였다. 그 결과 모델 및 입력 자료의 조합에 따라 3.67~171.46로 다소 넓은 범위의값이 도출되었다. 그 중 가장 우수한 유형은 수위, 연도, 수면폭 3개의 입력자료를 조합하여 랜덤포레스트 회귀 모델에 적용한 경우이다. 비교를 위해 동일한 검증 데이터를 한국수문조사연보(2018년) 내 거운교 지점의 수위별 수위-유량 곡선식을 이용해 유량을 추정한 결과 RMSE가 3.76이 산출되어, 머신러닝이 세분화된 수위-유량 곡선식과 비슷한 수준까지 성능을 내는 것으로 확인되었다.

본 연구는 양질의 유량자료 생산을 위해 기 구축된 수문자료를 기반으로 머신러닝 기법의 적용가능성을 검토한 기초 연구로써, 국내 효율적인 수문자료 측정 및 수위-유량 곡선 산출에 도움이될 수 있을 것으로 판단된다. 향후 수자원 환경 및 통제특성에 영향을 미치는 다양한 영향변수를 파악하기 위해 기상자료, 취수량 등의 입력 자료를 적용할 필요가 있으며, 머신러닝 내 비지도학습인 딥러닝과 같은 보다 정교한 모델에 대한 추가적인 연구도 수행되어야 할 것이다.

핵심용어: 머신러닝, 빅데이터, 유량자료 생산, 수위-유량곡선

^{*} 정회원·한국수자원조사기술원 연구원·E-mail: noelkang@kihs.re.kr

^{**} 정회원·한국수자원조사기술원 전임연구원·E-mail: net0710@kihs.re.kr

^{***} 정회원·한국수자원조사기술원 선임연구원·E-mail: ljh0817@kihs.re.kr

^{****} 정회원·한국수자원조사기술원 책임연구원·E-mail: chungdea@kihs.re.kr