

댐 일유입량 예측을 위한 데이터 전처리와 머신러닝&딥러닝 모델 조합의 비교연구

Comparative Study of Data Preprocessing and ML&DL Model Combination for Daily Dam Inflow Prediction

조영식*, 정관수**

Youngsik Jo, Kwansue Jung

요 지

본 연구에서는 그동안 수자원분야 강우유출 해석분야에 활용되었던 대표적인 머신러닝&딥러닝(ML&DL) 모델을 활용하여 모델의 하이퍼파라미터 튜닝뿐만 아니라 모델의 특성을 고려한 기상 및 수문데이터의 조합과 전처리(lag-time, 이동평균 등)를 통하여 데이터 특성과 ML&DL모델의 조합시나리오에 따른 일 유입량 예측성능을 비교 검토하는 연구를 수행하였다. 이를 위해 소양강댐 유역을 대상으로 1974년에서 2021년까지 축적된 기상 및 수문데이터를 활용하여 1) 강우, 2) 유입량, 3) 기상자료를 주요 영향변수(독립변수)로 고려하고, 이에 a) 지체시간(lag-time), b) 이동평균, c) 유입량의 성분분리조건을 적용하여 총 36가지 시나리오 조합을 ML&DL의 입력자료로 활용하였다. ML&DL 모델은 1) Linear Regression(LR), 2) Lasso, 3) Ridge, 4) SVR(Support Vector Regression), 5) Random Forest(RF), 6) LGBM(Light Gradient Boosting Model), 7) XGBoost의 7가지 ML방법과 8) LSTM(Long Short-Term Memory models), 9) TCN(Temporal Convolutional Network), 10) LSTM-TCN의 3가지 DL 방법, 총 10가지 ML&DL모델을 비교 검토하여 일유입량 예측을 위한 가장 적합한 데이터 조합 특성과 ML&DL모델을 성능평가와 함께 제시하였다.

학습된 모형의 유입량 예측 결과를 비교·분석한 결과, 소양강댐 유역에서는 딥러닝 중에서는 TCN모형이 가장 우수한 성능을 보였고(TCN>TCN-LSTM>LSTM), 트리기반 머신러닝중에서는 Random Forest와 LGBM이 우수한 성능을 보였으며(RF, LGBM>XGB), SVR도 LGBM수준의 우수한 성능을 나타내었다. LR, Lasso, Ridge 세가지 Regression모형은 상대적으로 낮은 성능을 보였다. 또한 소양강댐 댐유입량 예측에 대하여 강우, 유입량, 기상계열을 36가지로 조합한 결과, 입력자료에 lag-time이 적용된 강우계열의 조합 분석에서 세가지 Regression모델을 제외한 모든 모형에서 NSE(Nash-Sutcliffe Efficiency) 0.8이상(최대 0.867)의 성능을 보였으며, lag-time이 적용된 강우와 유입량계열을 조합했을 경우 NSE 0.85이상(최대 0.901)의 더 우수한 성능을 보였다.

핵심용어 : Machine Learning, Deep Learning, Daily Inflow Prediction, Multipurpose Dam

* 정희원 · 한국수자원공사 수자원운영처, 충남대학교 토목공학과 박사과정 · E-mail : ijslord@kwater.or.kr

** 정희원 · 충남대학교 토목공학과 교수 · E-mail : ksjung@cnu.ac.kr