

머신러닝 알고리즘을 이용한 결측 강우 데이터 추정에 관한 연구

Imputation of missing precipitation data using machine learning algorithms

한희찬*

Heechan Han

요 지

강우 데이터는 수문기상, 환경, 농업, 자연재해, 그리고 수자원 시스템 분야에서 가장 필수적인 기본 요소 중 하나이다. 또한 강우 데이터는 수문학적 분석에서 활용되는 필수 입력 자료 중 하나로 관측 데이터의 품질에 따라 수문 모형을 이용한 모의 결과물의 정확도가 결정된다고 할 수 있다. 따라서, 강우 관측소별로 강우 데이터의 품질을 어떻게 관리하느냐에 따라 수문 모형의 활용 범위 및 수자원 관리의 효율성이 결정될 수 있다.

강우의 시공간적 변동성은 수 많은 인자들과 직간접적으로 연계되어 있기 때문에 미계측 강우 자료에 대해 직접 관측이 아닌 수치 모형을 이용하여 강우의 발생과 강우량을 산정하는 것은 매우 복잡한 과제 중 하나이다. 현재 국내에서 운용되고 있는 강우 관측소의 경우에도 미계측된 강우 데이터가 존재함으로써 강우 데이터의 활용에 제한이 생기는 경우가 있다. 따라서, 이러한 미계측 데이터의 추정 및 보완은 보다 효과적인 수재해 방지, 수자원 관리를 위한 필수 과제 중 하나이다. 일반적으로, 미계측 강우를 산정하기 위해서 Kriging, Thiessen, 등우선법, 그리고 역거리 관측법 등 다양한 수문학적 방법들이 적용되고 있다. 이러한 방법들은 산악효과나 강우 관측소의 분포 상태 등을 고려하지 못하기 때문에 측정하는 지역에 따라 강우 추정 오차가 커질 수 있다는 한계가 있다.

최근에는 데이터 관측 시스템과 빅데이터 기술의 발전과 활용 가능한 데이터의 양이 증가함에 따라 머신러닝을 활용한 사례가 증가하고 있다. 머신러닝은 데이터 사이의 관계를 기반으로 분류, 회귀, 그리고 예측 문제에 주로 사용되는 기법 중 하나이다.

따라서, 본 연구에서는 광주광역시 지역에 위치한 주요 강우 관측 지점들을 대상으로 미계측된 시강우 데이터를 추정 및 복원하고자 한다. 여기서 데이터 추정 기술이란 미계측 강우의 발생 유무 및 강우량을 추정할 수 있는 기술을 의미한다. 이를 위해 대표적인 머신러닝 알고리즘인 인공신경망(Artificial Neural Network) 및 랜덤포레스트(Random Forest)를 적용하였다.

핵심용어 : 강우추정, 머신러닝, 인공신경망, 랜덤포레스트

감사의 글

본 연구는 2023년도 조선대학교 학술연구비의 지원을 받아 연구되었음. 이에 감사드립니다.

* 정회원 · 조선대학교 공과대학 토목공학과 조교수 · E-mail : heechan@chosun.ac.kr