

머신러닝을 이용한 국내산 플라이애시의 반응성 예측

Prediction of the Reactivity of Fly Ashes Using Machine Learning

박 우 영* 문 주 혁**

Park, Woo Young Moon, Ju hyuk

ABSTRACT

In this study, a machine learning(ML) model was built from domestic fly ashes' chemical compositions(wt%) and sum of amorphous aluminosilicate(wt%) calculated by X-ray fluorescence(XRF) analysis and quantitative X-ray diffraction(QXRD) analysis. The objective is to predict the amount of amorphous aluminosilicate from their chemical compositions using the built ML model. An ensemble technique is applied to the ML model that is re-trained by a meta model. The meta model is the combined model of Boosting, Random Forest and multi-layer perceptron(MLP). The performance of model for some samples, except for outlier samples, has a coefficient of determination of 0.867 in train dataset and 0.801 in test dataset.

요 약

본 연구에서는 X-선 형광 분석법(XRF)과 정량 X-선 회절 분석법(QXRD)을 통해 계산한 국내산 플라이애시의 화학 조성값(wt%)과 비정질 알루미늄노실리케이트 함량(wt%)에 대한 머신러닝 모델을 구축하였다. 플라이애시의 화학 조성값으로 그것의 비정질 알루미늄노실리케이트 함량을 예측하는 모델이다. 해당 기계학습 모델에서는 부스팅, 랜덤 포레스트, 다층 퍼셉트론(MLP) 모델을 합친 메타모델로 재학습시키는 앙상블 기법을 사용하였다. 이상치에 해당하는 샘플을 제외한 일부 샘플에 대한 모델의 성능은 학습 데이터에서 0.867, 테스트 데이터에서 0.801의 결정계수를 갖는다.

1. 서 론

플라이애시는 비정질 알루미늄노실리케이트 함량이 높을수록 반응성이 높기 때문에 콘크리트 품질을 예측하기 위해서는 플라이애시의 비정질 알루미늄노실리케이트 함량을 아는 것이 중요하다. 하지만, 국내 화력발전소에서 생산된 플라이애시의 정량적 x-선 회절 분석에 관한 연구가 부족한 실정이다.

* 정회원, 서울대학교, 멀티스케일구조재료연구실, 석사과정

** 정회원, 서울대학교, 멀티스케일구조재료연구실, 부교수, juhyukmoon@snu.ac.kr

2. 사용 재료 및 실험 방법

본 연구에서 머신러닝 모델을 구축하기 위해 사용된 데이터는 62종의 국내산 플라이애시로, 17개 국내의 논문에서 수집하였다. 이상치 값을 제외하여 13개 논문에서 선별된 43종의 플라이애시만을 대상으로 머신러닝 모델을 새로 구축하였다. 플라이애시의 화학 조성값 (1) Al_2O_3 , (2) CaO , (3) Fe_2O_3 , (4) SiO_2 , (5) MgO , (6) $\text{Na}_2\text{O} + 0.658\text{K}_2\text{O}$ 을 입력값으로 받아 머신러닝 모델의 출력값인 비정질 알루미늄노실리케이트 함량을 예측하였다. 62개의 데이터는 계층적 샘플링을 통해 훈련 데이터 34개, 테스트 데이터 9개로 나누었다.

3. 결과 및 고찰

실제값-예측값이 $y=x$ 선 위에 모두 분포할 때 완벽한 모델이 구축되었음을 의미한다. 그림 2에서 알 수 있듯이, 이상치를 제거한 새로운 모델은 플라이애시의 비정질 알루미늄노실리케이트 함량값을 비교적 정확하게 예측한다.

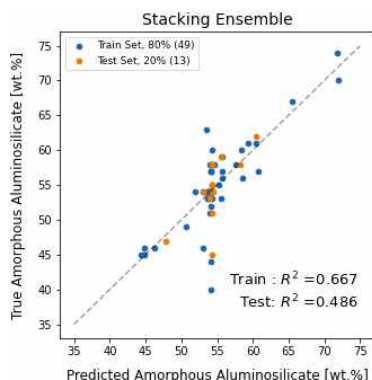


그림1. 앙상블 메타모델의 실제값-예측값 산점도

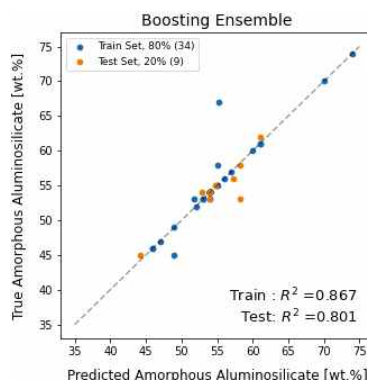


그림2. 이상치를 제거한 새로운 모델의 산점도

4. 결론

앙상블 알고리즘을 이용해 플라이애시의 화학 조성에 따른 반응성을 예측하였으며, 본 연구의 결론은 다음과 같다.

1) 다중 머신러닝 모델을 결합한 메타모델을 통해 이상치 샘플을 제외하여 새로운 모델을 구축하였으며, 테스트 데이터에 대한 0.801 결정계수로 높은 정확도를 보인다.

2) 이상치 샘플 선별에 대한 설명력이 떨어지므로, 해당 모델은 일부 샘플에만 국한된 모델이다.

감사의 글

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2021R1A2C4001944)

참고문헌

1. Song, Y et al. (2021). Machine Learning Enables Rapid Screening of Reactive Fly Ashes Based on Their Network Topology. *ACS Sustainable Chemistry & Engineering*, 2021 9 (7).