

지진 데이터를 이용한 건물 피해 예측 모델의 성능 분석

채송화¹, 임유진²

¹ 숙명여자대학교 인공지능공학부 학생

² 숙명여자대학교 인공지능공학부 교수

watermelon97@sookmyung.ac.kr, yujin91@sookmyung.ac.kr

Performance Analysis of Building Damage Prediction Models using Earthquake Data

Songhwa Chae¹, Yujin Lim²

¹Div. of Artificial Intelligence Engineering, Sookmyung Women's University

²Div. of Artificial Intelligence Engineering, Sookmyung Women's University

요 약

내진 설계가 되어있지 않은 건물의 경우, 지진으로 인해 건물 붕괴 가능성이 높아지며 이로 인해 많은 인명 피해가 발생할 수 있다. 지진으로 인한 건물의 피해를 예측하고 이를 기반으로 취약점을 보완한다면 인명 피해를 줄일 수 있으므로 건물 피해 예측 모델에 대한 연구가 필요하다. 본 논문에서는 2015 년 네팔 대지진으로 인해 손상된 건물 데이터를 활용하여 Random Forest 와 Extreme Gradient Boosting 기계학습 분류 알고리즘을 사용하여 지진 피해 예측 모델의 정확도를 비교하였다.

1. 서론

2015 년 4 월 25 일, 네팔의 남정지역에 규모 7.8 의 강력한 지진으로 인해 수십만명이 집을 잃고 많은 문화재가 상당수 파괴되었다. 뿐만 아니라, 최소 8 천여명의 사망자와 22,000 여명의 부상자가 발생했다. 이와 같이 피해가 극심했던 이유로 지진의 규모가 워낙 컸다는 점을 들 수 있으나 근본적인 이유로는 네팔 건축물의 대부분이 내진 설계를 감안하여 건축되지 않았고 대부분의 건물이 흙벽돌로 지어졌다는 점을 지적할 수 있다 [1].

그러나 많은 건물들을 내진 설계를 통해 재건축하거나, 보강하는 것은 현실적으로 어려운 일이다. 따라서 건물의 특성 정보를 통해 지진에 의한 피해를 예측하여 사전에 약점을 보완하거나 지진 발생 이후 피해 복구의 우선 순위를 정할 수 있다면 건물 및 인명 피해 경감에 도움을 줄 수 있을 것으로 예상된다.

2. 데이터 수집 및 전처리

본 논문에서는 2015 년 네팔 대지진으로 인해 피해를 입은 11 개 지역의 건물 데이터와 해당 건물들의 지진 피해 정도를 나타내는 데이터 세트를 활용하여 지진으로 인한 건물 피해를 예측하는 적절한 모델을

찾고자 한다. 데이터 세트는 총 762,106 개 건물의 바닥, 토대, 지붕, 나이 등 모든 건물의 특성 정보와 피해 정도를 나타내는 Damage Grade 로 구성되어 있다 [2]. 피해 정도는 Grade 1 부터 Grade 5 까지 다섯 단계로 구분되며, 숫자가 증가함에 따라 피해 정도가 증가함을 의미한다.

각 데이터 세트에 원-핫 인코딩을 적용하고 K-Fold 교차 검증을 위해 데이터를 다섯 개로 분할한 후 나눈 데이터의 80%를 학습 데이터로, 20%를 테스트 데이터로 구분하였다.

3. 모델 학습 및 성능 평가

본 논문에서는 Random Forest(RF), Extreme Gradient Boosting(XGB), 두 가지의 기계학습 분류 알고리즘을 사용하여 건물 지진 피해 예측을 수행하였다.

RF 는 이진, 숫자, 범주형 특징을 쉽게 처리하며 단일 트리에 비해 높은 성능을 제공하고 낮은 오류율, 뛰어난 정밀도 및 효율성과 함께 탁월한 노이즈 저항성을 제공하기 때문에 예측 모델에 적합하다는 장점을 가지고 있다 [3]. XGB 는 기존의 Gradient Tree Boosting 알고리즘의 약점인 과적합을 방지하기 위해 파라미터를 추가한 지도 학습 알고리즘이다. 해당 알

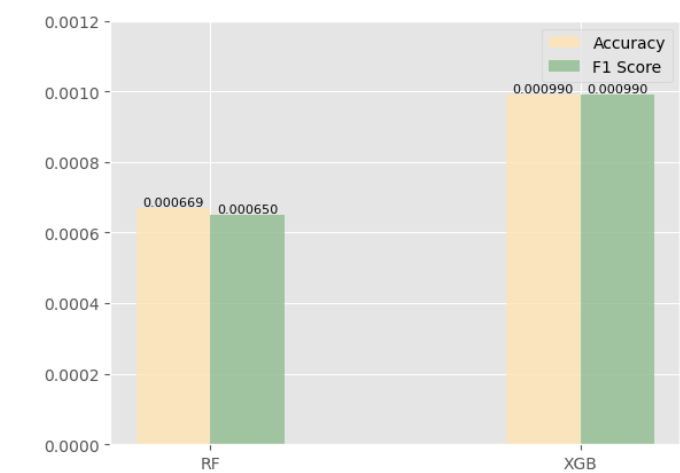
고리즘은 모델의 성능과 처리 속도를 높이는 기술 중 하나이다. 이 때, 데이터의 양이 적을 경우 과적합 가능성이 있으므로 주의하여야 한다 [4].

이 두 알고리즘은 앙상블 학습의 대표적인 알고리즘에 속하지만, 결과 예측 방식에서 차이를 보인다. RF 는 표본을 여러 번 추출하여 각 모델을 학습시켜 결과를 집계하는 병렬 학습 방식의 배경 앙상블이고, XGB 는 서로 독립적인 결정 트리가 각각 값을 예측한 뒤 결과를 집계하여 최종 결과값을 예측하는 순차적 학습 방식의 부스팅 앙상블이다 [5]. 건물의 피해 정도 예측을 위한 적합한 알고리즘을 탐색하기 위해 RF 와 XGB 의 성능 평가를 수행한다.

Validation Set	Accuracy		F1 Score	
	RF	XGB	RF	XGB
1	0.8971	0.8935	0.8973	0.8935
2	0.8953	0.8913	0.8958	0.8913
3	0.8961	0.8929	0.8961	0.8929
4	0.8971	0.8943	0.8973	0.8943
5	0.8961	0.8931	0.8961	0.8931
Average	0.8963	0.8930	0.8965	0.8930

(표 1) 모델 정확도

RF 알고리즘으로 학습을 수행했을 경우 약 89.63%의 정확도와 89.65%의 F1 Score 로 건물의 지진 피해를 예측했고, XGB 알고리즘의 경우에는 약 89.30%의 정확도와 F1 Score 를 나타냈다. 두 알고리즘의 성능이 유사한 것을 알 수 있다.



(그림 1) 성능 표준편차

그러나 정확도 측면에서의 성능은 비슷하다 하더라도 표준편차 측면에서 RF 알고리즘이 약 33.43% 더 나은 성능을 보임을 알 수 있다.

4. 결론

본 논문에서는 지진 데이터를 통해 Random Forest 와 Extreme Gradient Boosting 과 같은 기계학습 분류 알고리즘을 활용하여 건물 지진 피해를 예측하는 모델의 정확도를 평가하였다. 향후 지진으로 인한 인명 및 자산 피해를 개선하기 위해 보다 다양한 모델을 활용한 추가적인 연구가 필요하다.

사사문구

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. 2021R1F1A1047113).

참고문헌

[1] 강익범, "네팔국가와 2015 년 4 월 25 일 네팔지진," vol. 62, pp. 70-75, 한국방재학회지, 2015.

[2] 2015 Nepal Earthquake: Open Data Portal, (<http://eq2015.klldev.org/#/>)

[3] P. M. Chanal and M. S. Kakasageri, "Random Forest Algorithm based Device Authentication in IoT," IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), Bangalore, India, 14-16 July 2023.

[4] J. A. Sabejon, J. B. Rejas, and G. S. Limacad, "XGBoost-Based Analysis of the Early-Stage Diabetes Risk Dataset," International Conference in Advances in Power, Signal, and Information Technology (APSIT), Bhubaneswar, India, 09-11 June, 2023.

[5] A. Mokhtar et al, "Estimation of SPEI Meteorological Drought Using ML Algorithms," IEEE Access (Volume: 9), 20 April 2021.