

기계학습을 이용한 지하매설물 특성에 따른 지반함몰 발생등급 분류 예측모델 비교

이성열, 강재모, 김진영
한국건설기술연구원 지반연구본부
e-mail:leesy@kict.re.kr

Comparison of ground subsidence prediction models using machine learning according to the characteristics of underground facilities

Sung-Yeol LEE, Jae-Mo Kang, Jin-Young Kim
Department of Geotechnical Engineering Research, Korea Inst. of Civil Engr. and Building Tech.

요약

도심지 중심에서 발생하는 지반함몰은 인명 및 물질적 피해, 사회 인프라 마비 등 심각한 피해를 입히기 때문에, 이에 대한 원인 규명 및 관리가 반드시 이루어져야 한다. 지반함몰의 주요 원인은 하수관, 상수관과 같은 지하매설물의 손상으로 알려져 있으나 복합적인 원인에 의해 발생하는 현상이기 때문에 지반함몰을 정확히 예측하는 것은 매우 어렵다. 따라서 지반함몰을 예측하여 사고를 방지하고자 다양한 기법을 활용한 연구가 꾸준히 수행되고 있다. 본 연구에서는 〇〇시 〇〇구의 지하매설물 특성 데이터와 해당 지역을 500mx500m 규모로 나눈 그리드 내에 발생한 지반함몰 개수를 등급화(3,5단계)하여 데이터 셋을 구축하였으며, 지하매설물 특성 데이터는 매설년수와 환경이 활용되었다. 기계학습 모델 중 Random Forest(RF), Gradient Boosting(GBDT), XGBoost, LightGBM 모델에 적용하여 지반함몰 발생 등급 분류 모델을 비교하였다. 그 결과, 3등급으로 지반함몰 위험도를 구분했을 때, 대부분 모델에서 우수한 성능의 지표가 도출되었으며 특히 XGBoost 모델이 가장 적합한 것으로 나타났다.

1. 서론

인구 밀집도가 높은 도심지를 중심으로 발생하는 지반함몰은 사고 발생 시 커다란 피해를 발생시키므로 사고를 미연에 방지하여 피해를 최소화 하여야 한다. 지반함몰의 주요 원인으로서는 지하매설물의 손상으로 알려져 있으나 지하매설물 손상 하나의 원인으로 발생하지 않고 복합적인 원인에 의해 발생하는 현상이므로 그 원인을 찾아 사고를 예측하긴 매우 어려운 실정이다. 이에 따라 지반함몰 발생 예측을 위해 다양한 기법을 적용한 연구가 꾸준히 수행되고 있는 추세이며, 본 연구에서는 기계학습 기법을 활용하여 〇〇시의 〇〇구 지하매설물 특성에 따른 지반함몰 예측 모델을 제안하고 비교하고자 한다.

2. 데이터셋 구축

지반함몰 예측 모델을 제안하기 위해 기계학습에 필요한 데이터셋을 구축해야하며, 본 연구에서는 〇〇시 〇〇구를 ArcGIS를 활용하여 500mx500m크기의 그리드로 분할하여

해당 그리드 내의 지반함몰 위험도를 예측하고자 하였다.

지하매설물(하수, 상수, 전기, 가스, 통신)을 1개의 관으로 가정하여 그리드내에 매설되어 있는 관의 매설년수 및 환경을 영향인자로 선정하였다. 매설년수의 단위는 관로의 길이(m)로 선정하였으며, 환경은 100mm 당 관로 길이(m)를 단위로 선정하였다. 또한, 지반함몰 발생 위험도는 그리드내의 지반함몰 개수를 고려하여 3단계와 6단계로 구분하였고, 지하매설물 특성 중 매설년수는 5년과 10년 단위로 병합하였으며, 환경은 100mm 단위로 세트를 구축하였다.

3. 지반함몰 발생등급 분류 예측모델 결과

지반함몰의 발생등급을 예측하기 위한 모델은 RF, GBDT, XGB, LightGBM을 활용하였으며, Python의 Scikit-learn 패키지를 활용하였다. 또한, GridSearch CV를 활용하여 각 모델의 Hyper-parameter를 최적화 하였으며, 모델의 평가지표는 Accuracy(Test, Train), F1-Score로 선정하였다.

분석에 활용된 구분된 데이터셋의 경우의 수는 4가지 이며,

매설년수 5년, 10년에 각각 위험등급 3단계, 6단계이다. 표 1~4는 각 데이터셋을 모델에 적용한 결과의 평가지표를 요약한 표이다.

[표 1] 5Year_3grade 데이터셋의 평가결과

5Year_3grade	Test Score	Train Score	F1 Score
RF	0.688	0.705	0.500
GBDT	0.744	0.786	0.710
XGB	0.750	0.783	0.720
LightGBM	0.719	0.758	0.660

[표 2] 5Year_6grade 데이터셋의 평가결과

5Year_6grade	Test Score	Train Score	F1 Score
RF	0.744	0.798	0.390
GBDT	0.756	0.797	0.400
XGB	0.744	0.766	0.390
LightGBM	0.750	0.797	0.400

[표 3] 10Year_3grade 데이터셋의 평가결과

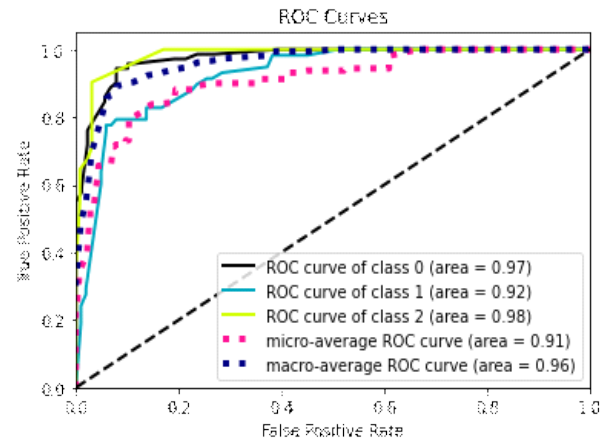
10Year_3grade	Test Score	Train Score	F1 Score
RF	0.775	0.817	0.750
GBDT	0.725	0.806	0.680
XGB	0.788	0.839	0.770
LightGBM	0.644	0.709	0.510

[표 4] 10Year_6grade 데이터셋의 평가결과

10Year_6grade	Test Score	Train Score	F1 Score
RF	0.756	0.808	0.400
GBDT	0.781	0.816	0.410
XGB	0.769	0.887	0.410
LightGBM	0.781	0.861	0.450

위의 표에 나타난 바와 같이, 모델 결과 위험등급을 3단계로 구분한 데이터셋의 정확도 및 F1 Score의 값이 우수하게 도출되었으며, 가장 좋은 성능을 발휘하는 모델은 10Year_3grade의 XGboost 모델로 나타났다. 따라서 해당 모

델의 ROC Curve를 통한 모델 신뢰도를 분석하였다. 그림 1은 해당모델의 AUC 곡선을 나타낸 그래프이며, 그래프를 살펴보면 3단계(0,1,2)를 구분하는 모든 AUC값이 0.9 이상으로 나타났다. 일반적으로 AUC의 값이 0.8이상이면 매우 좋은 성능을 발휘하는 모델로 평가하기 때문에, 해당 모델은 지반함몰 발생 등급을 분류하는데 적절한 것으로 판단된다.



[그림 1] 10year_3grade XGB 모델의 AUC

해당 연구를 통해 제시된 지반함몰 발생등급 분류 모델은 넓은 그리드(500mx500m)를 대상으로 위험도를 예측하기 때문에 정확한 사고 예측 지역을 선정하기 어렵다. 하지만 지중 탐사레이더와 같이 공동을 사전에 탐사하는 장비를 투입할 경우, 본 연구에서 제시한 모델을 활용하여 위험 지역을 우선적으로 투입하여 지반함몰 사고를 사전에 예방할 수 있을 것으로 기대된다.

참고문헌

- [1] 김진영, 강재모, 최창호, 박두희, “하수관로 건전도와 도로 함몰 발생 상관관계 분석”, 한국지반환경공학회 논문지, 제 18권 6호, pp. 31-37, 2017년.
- [2] 서울시, “도로함몰 예방을 위한 동공발생 메커니즘과 대응방안”, 2015년.
- [3] 이성열 “하수관로 지반함몰 위험도 분석 최적화를 위한 기계학습 기법 선정 및 영향인자 중요도 분석”, 전남대학교 박사학위논문, 2022년.