

환경변수를 고려한 머신러닝 기반 도로파손 요인 분석

김기남*, 이민재*
*충남대학교 토목공학과
e-mail:go7140@cnu.ac.kr

Analysis of Deterioration Factors for Highway Pavement Based on Machine Learning

Ki-Nam Kim*, Min-Jae Lee*
*Dept. of Civil Engineering, Chungnam National University

요약

본 연구에서는 교통하중과 환경요인이 도로 포장구간 상태에 미치는 영향을 분석하기 위하여 교통하중과 기후 데이터를 기반으로 분류모델을 제안하고자 한다. 총 12개의 변수를 활용하여 머신러닝모델의 입력 데이터의 형태로 가공하였고, 랜덤포레스트 분류 기법을 이용하여 분류 모형을 구축 하였다. 본 연구 모형을 적용한 결과, 정확도 76.3%, F1-척도 66.7% 그리고 AUC 75.3%로 Crack의 상태분류의 가능성을 확인하였다. 또한 MDA(Mean Decrease Accuracy)와 MDG(MeanDecreaseGini)를 통하여 최저초상온도(J), ESAL(A2), AADT(A1)순으로 분류모델의 가장 큰 영향인자임을 알 수 있었고, Crack의 상태분류를 대표하는 주요 영향인자임을 확인하였다.

1. 서론

현재 우리나라에 새로운 대규모 사회기반시설을 건설하기에는 많은 어려움이 있다. 또한 경제 발전기에 건설된 사회기반시설들의 노후화가 진행되고 있어 그 유지·보수의 필요성이 대두되고 있다. 수많은 사회기반시설의 건설 및 관리에 대한 필요성은 높아지고 있는 반면 예산은 한정적이기 때문에 보다 새로운 시설물의 확충보다는 기존 시설물의 유지·보수의 필요성이 대두되고 있으므로 사회기반시설물 관리의 효율성을 위한 비용편익분석 방안이 중요시되고 있다. 사회기반시설물 중 대중에게 가장 영향력이 큰 시설은 도로라 할 수 있다. 도로의 유지·보수에 있어서는 도로 포장의 수명의 측정이 중요한 사항이다. 따라서 도로를 이용함으로써 생기는 교통하중과 환경요인에 대해서 도로 포장 수명에 미치는 영향을 알아볼 필요성이 있다. 본 연구에서는 교통하중과 환경요인이 도로 포장구간 수명에 미치는 영향을 분석하기 위하여 교통하중과 기후 데이터를 기반으로 머신러닝기법 중 랜덤포레스트를 활용하여, 도로상태에 영향을 주는 영향인자를 분석하는 것이 이 연구의 목적이다.

2. 문헌고찰

권수안(2018)은 기상청이 발표한 '8월 기상특성'에 따르면 2018년 여름(6~8월)이 평균기온, 폭염일수, 열대야일수 모두

1위를 기록한 것으로 나타났으며, 국회에서는 폭염과 혹한을 재난관리 기본법에 포함시키는 등 도로 인프라를 중심으로 기후변화로 인한 피해를 최소화하기 위한 제언하였다.

한 대석(2013)은 다양한 유지보수공법을 다양한 유지보수기준을 시행하고 있는 포장관리의 현실에 비추어 볼 때 비현실적이며, 중장기 예산수요예측에도 영향을 미칠 수 있다. 이에 4가지 요인을 도출하였고, 1)포장의 파손과정은 파손정도에 따라 이질적인 특성을 갖고, 2)그 포장파손과정에는 다양한 도로환경(예:포장강도, 교통하중 등)의 영향으로 인한 심각한 불확실성을 포함하고 있으며, 3)관리자가 운용하는 다양한 유지보수기준에 따라 기대수명이 달라진다는 현실을 반영하고자 하였다. 여기에 추가적으로 4)각 유지보수기준에 해당하는 기대수명의 통계적 신뢰수준에 따라 최종 기대수명이 달라질 수 있다는 점도 추가적으로 고려하고자 하였다.

배운신(2012)은 서울시는 월별 강수량 및 소파발생량을 조사하여 강수량이 증가하는 7월, 8월에 소파 보수 발생건수가 증가함을 확인하였고, 1월부터 3월은 해빙기 영향으로 인해 포장파손을 가속화시킨 것으로 분석하였으며, 주요 결론으로는 확일적인 덧씌우기 공법으로 인한 도로포장 조기파손의 문제점을 확인하였다.

3. 랜덤포레스트를 활용한 도로포장 분류모델

3.1 랜덤포레스트(Random Forest)

랜덤 포레스트(random forest)는 머신러닝 분류 기법 중의 하나이며, 랜덤 포레스트는 동일한 하나의 데이터 집합에서 여러 개의 훈련용 데이터를 만들어 여러 번의 학습을 통해 여러 개의 의사결정나무를 생성하여 이들을 결합한 결과를 예측한다. 랜덤 포레스트는 학습된 데이터들로 결정된 의사결정나무 모형을 생성하는 과정에서 데이터를 가공 및 제거하지 않고 무작위로 선정된 데이터를 활용하며, 이로부터 데이터 분류에 효과적인 데이터를 선정하고 중요한 변수들을 평가한다. 랜덤포레스트 기법은 의사결정나무 모형을 활용하기 때문에 속도가 빠르고, 반복하여 분석되기 때문에 의사결정나무가 과적합(over fitting)되지 않는다는 장점을 가지고 있다. 이러한 랜덤포레스트 기법은 분석에 활용된 데이터에 포함된 다수의 변수를 여러 개의 의사결정나무 모형을 생성하여 각각의 모형을 구성하는 마디에 위치한 변수들의 중요도를 평가하는데 활용할 수 있다(Lee, 2011). 본 연구에서는 Crack의 상태를 분류하였으며, 총 2495개 구간과 환경요인을 이용하여 포장상태에 미치는 영향을 미치는 인자를 평가하기 위해 랜덤 포레스트 기법을 활용하여 12개 인자들의 중요도를 평가하였다.

[표 1] 랜덤포레스트 분류 모델 데이터 구성

구간	전국 (2495개 구간)
포장상태	Crack(8년간의 상태변화)
환경요인	AADT(A1), ESAL(A2), Age(A), 공용년수(B)
	평균 최저기온(C), 평균 최고기온(D), 최저기온(E), 최고기온(F), 평균 이슬점(G), 평균지면 최저온도(H), 평균지면 최고온도(I), 최저초상온도(J)

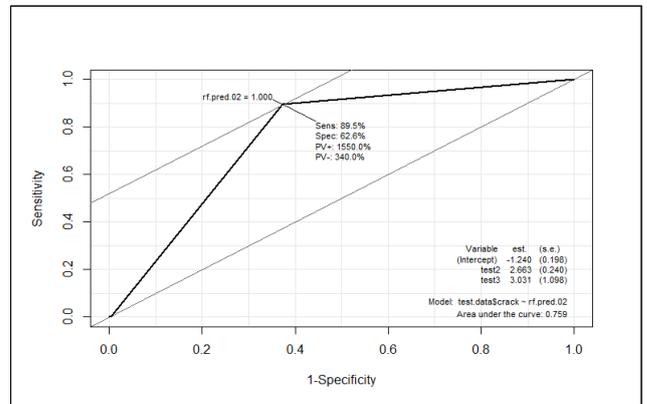
아울러, 8년간의 도로포장상태 모니터링 데이터 중 Crack의 변화를 대상으로 도로상태를 분류하였으며, 분류모델에서는 랜덤포레스트를 구성하는 나무수(Ntree)는 200개이며, 나무마다 분석되는 입력변수(Mtry)는 8개이다. 또한, 크랙의 분류 기준은 표2와같이 3분류로 구분하였다. 모델 구축에 사용된 학습 데이터(80%)와 테스트데이터(20%)는 비율로 구성하여 분류를 시행하였다.

[표 2] Crack 분류 기준

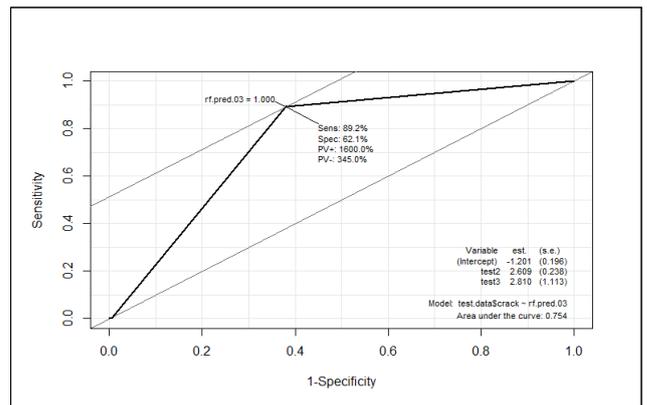
	Crack (%)	분류 기준
1	~0.5	매우 좋음 (육안검사 불가)
2	0.5~15.0	보통 (이용자가 느낄 수 있으며 육안으로 식별가능)
3	15.0~	매우 나쁨 (육안으로 식별가능하며 재포장수준의 유지보수)

3.2 랜덤포레스트 분류 모델 성능평가

분류 모델의 성능을 평가하기 위해 정확도(Accuracy), 오분류율(Misclassification Rate), F1-Score 그리고 AUC(Area Under the Curve)를 사용하였다. 정확도는 실제 1인 것을 1로, 실제 0인 것을 0으로 예측한 비율이고 오분류율은 실제 1인 것을 0으로, 실제 0인 것을 1로 잘못 예측한 비율이다. F1-Score는 재현율(Recall)과 정밀도(Precision)의 조화평균으로 계산되는데 여기서 재현율은 실제 1인 것 중 예측도 1로 한 비율이고 정밀도는 예측된 1중 실제 1인 것의 비율을 의미한다. 분류결과 A 모델과 B모델의 성능지표를 도출하였다. A모델의 경우 정확도와 AUC값이 B모델에 비하여 높았으나, B모델의 F1-Score값이 66.7%로 A모델에 비하여 높은 것으로 분석되었다.



[그림 1] A-MODEL



[그림 2] B-MODEL

[표 3] 랜덤포레스트 분류 모델 성능평가

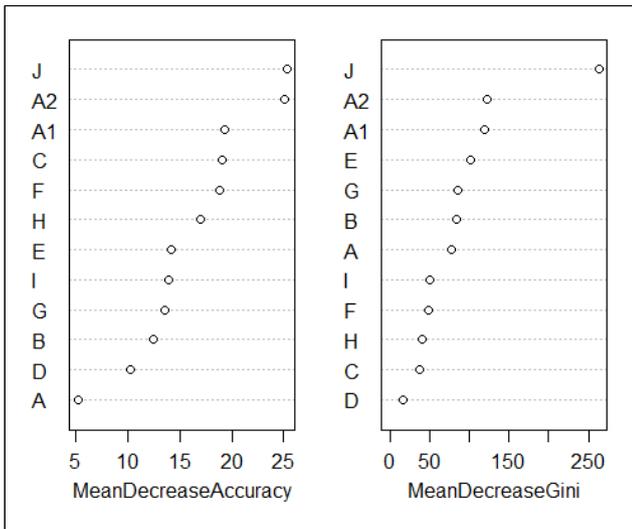
	A-Model	B-Model
Accuracy	0.7655311	0.7635271
F1.score	0.6554992	0.6678948
AUC	0.7588191	0.753753

4. 랜덤 포레스트를 활용한 영향인자 중요도 평가

참고문헌

본 연구에서는 Crack의 상태를 포장상태에 따라 3분류 하였으며, 전국의 국도 2495개 구간과 환경요인을 이용하여 Crack에 미치는 영향을 미치는 인자를 평가하기 위해 랜덤 포레스트 기법을 활용하여 12개 인자들의 중요도를 평가하였다. 분류모델의 영향인자를 평가하기 위하여 MDA(Mean Decrease Accuracy)의 값과 MDG(MeanDecreaseGini)의 값으로 판단하며 변수의 중요도는 J(최저초상온도), A2(ESAL), A(AADT)순으로 나타났다.

[1] 권수안(2018). 도로정책브리프(제131호) “기후변화에 따른 도로 인프라의 관리 제언”, pp. 02-03.
 [2] 배성호, 손덕수, 진태현, 문형철(2017) “가속도 센서를 활용한 도로포장의 주행쾌적성 평가 기초연구”, 한국도로학회 학술대회 발표논문 초록집, 한국도로학회, 19(1), p.141-141.
 [3] 한대석, 이수형, 윤인균(2013) “포장파손과정의 지역적 이질성과 다중유지보수기준을 고려한 기대수명 추정”, 한국도로학회 학술대회 발표논문 초록집, 한국도로학회, 15(1), p.18-18.
 [4] 배운신, 이상엽, 신경엽(2012) “서울시 도로포장 관리에 대한 고찰과 분석”, 서울도시연구 제13권 제2호 2012.6. 논문, PP. 207~219.



[그림 3] 영향인자 중요도

5. 결론

본 연구에서는 랜덤포레스트를 활용하여 Crack의 상태에 따라 3분류를 하였으며, 기상청의 각 지역의 기후 데이터와 교통량 및 공용년수를 이용하여 분류모델에 미치는 주요영향인자를 도출하였다. 데이터구성에 따라 랜덤포레스트 분류기법을 적용한 결과, 모델의 분류 성능은 정확도 76.3%, F1-척도 66.7% 그리고 AUC 75.3%로 비교적 좋은 성능을 보였다. 이러한 모델의 성능을 높이는데 영향을 준 설명변수들을 살펴보기 위해, 랜덤포레스트의 변수중요도 측정 지표인 MDG와 MDA를 사용하여 비교하였고 그 결과 최저초상온도(J), ESAL(A2), AADT(A1)순으로 가장 큰 영향인자임을 알수있었고, 도로포장상태 중 Crack을 분류하는데 보다 중요하게 사용되고 있음을 알 수 있었다.

감사의 글

본 연구는 국토교통부 스마트건설기술개발사업의 연구비 지원(20SMIP-A156374-01)에 의해 수행되었습니다.