빅데이터 처리 활용 및 머신러닝 기법 적용으로 인한 도로 손상 예측 모형 개발

Predictive Modeling of Pavement Damage Using Machine Learning and Big Data Processing

김도완* · 전진호** · 김담룡***

Kim, Dowan*, Jeon, Jinho**, and Kim, Damryung***

Abstract

Potholes, soil settlement, and road subsidence have become major road safety hazards in South Korea. Such problems not only impede driver and pedestrian safety but also cause secondary accidents, economic losses, and damage the nation's image. To this end, we developed local predictive models that can be extrapolated to national estimation models. These models were developed from a specific area (Seoul Metropolitan City) that has the highest occurrences of potholes and road subsidence. This research utilized big data and artificial intelligence techniques to develop these models. The first step involved the dimensional reduction of independent variables using a mechanical-statistical approach. A data standardization process was then used for reducing the uncertainty of these variables. A total of 19 machine learning optimization methods were used to train the standardized variables. The optimized models were finally determined by an error comparison. As a result, the optimized prediction models for potholes, soil settlement, and road subsidence were found to be multiple regression analysis that showed an accuracy of 70% and robust regression analysis that showed an accuracy of 73%.

Key words : Pothole, Soil Settlement, Road Subsidence, Big Data, Machine Learning

요 지

본 연구는 운전자 및 보행자의 안전성을 확보하기 위해, 최근 사회적 중점사항으로 부상하고 있는 포트홀, 지반침하 및 도로함몰에 대한 예측모형을 개발하는 것에 그 목적을 두고 있다. 포트홀, 지반침하 및 도로함몰은 운전자의 안전성을 저해할 뿐만 아니라 2차 사고를 발생시킬 수 있으며, 나아가 경제적 손실, 국가적 이미지 실추 등의 다양한 문제를 야기시킬 수 있다. 이와 관련하여 본 연구에서는 국가적 예측모형의 확장을 위한 방안으로 최근 도로 파손이 가장 빈번하게 발생하는 지역을 대상으로 예측모형을 개발했다. 예측모형 개발에 있어서 빅데이터의 활용과 인공지능기술(AI, Artificial Intelligence)의 적용에 중점을 두었다. 세부적인 예측 모형을 개발하는 과정에서는 구축된 빅데이터에 역학적·확률적 접근방법을 적용하여 독립변수의 차원을 축소시켰으며, 이 데이터의 불확실성을 저감시킬 목적으로 데이터 표준화를 실시했다. 표준화과정을 거친 인자들을 이용하여 19가지의 알고리즘으로 구성된 머신러닝의 학습을 실시했으며, 최소 오차비교로 최적의 알고리즘을 구축했다. 그 결과, 다중회귀분석으로 수행된 포트홀 예측모형과 로버스트 회귀분석을 통한 지반침하 & 도로함몰 예측모형을 개발했다. 이 예측 모형은 각각 70% 및 73%의 정확성을 가지고 있는 것으로 판단되었다.

핵심용어 : 포트홀, 지반침하, 도로함몰, 빅데이터, 머신러닝

*정회원, 서울과학기술대학교 건설시스템공학과 박사과정(E-mail: kimdw@kunhwaeng.co.kr)

Member, Doctorate Course, Department of Civil Engineering, Seoul National University of Science and Technology

**서울과학기술대학교 건설시스템공학과 석사과정(E-mail: jino@seoultech.ac.kr)

Master's Course, Department of Civil Engineering, Seoul National University of Science and Technology

***교신저자, 정회원, 서울과학기술대학교 건설시스템공학과 석사과정(Tel: +82-2-6938-7141, Fax: +82-2-555-1402, E-mail: drkim@seoultech.ac.kr)

Corresponding Author, Member, Master's Course, Department of Civil Engineering, Seoul National University of Science and Technology

1. 서 론

최근 도로함몰 및 포트홀이라는 단어는 ‘도로의 지뢰’, ‘도로의 구멍’이라 불리며 교통안전과 관련된 사회적인 이슈로 부각되고 있다. 이는 교통사고의 직접적인 원인이 되기도 하며, 이로 인해 교통안전에 대한 국민의 안전 불감증에 악영향을 미친다. 최근 2011년부터 2015년까지 전국에서 총 363,255건의 포트홀이 발생했으며, 이로 인해 1,653건의 교통사고가 발생한 것으로 조사되었다. 또한, 2009년부터 2015년을 대상으로 한 재난안전사고 조사에서는 월 평균 172,603건의 재난사고가 발생했으며, 그 중 131,265건이 교통사고로 추산되어 도로교통사고가 차지하는 비율이 76%로 달해 우리나라에서 발생하는 재난 중 가장 높은 비율을 차지하고 있다(MOIS, 2017). 이는 결국 도로교통에 대한 국민의 안전 불감증과 직접적인 관계를 갖는다(Kim et al., 2015). Figs. 1 및 2는 이에 대한 조사 자료이다.

Fig. 1. Pothole and Accident Occurrences (2011~2015)

Fig. 2. Safety Accident Occurrence Types (2009~2015)

올해 우리나라에서는, 부산시에서 140여개의 동공이 발생했으며, 올해 7월에 대형 싱크홀(지름 4 m, 깊이 3.5 m)이 발생한 후 닷새 만에 같은 지자체에서 다시 지름 0.6 m,

깊이 0.6 m의 싱크홀이 발생해 국가적 이슈로 부상한 바 있다. 한편, 경기도의 조사에 따르면 2015년에 3만6천 여 개의 포트홀이 발생했으며, 이듬해인 2016년에 4만1천 여 개의 포트홀이 발생하면서 그 수가 급격히 증가했다. 그럼에도 불구하고, 작년 역시 5만2천 여 개의 포트홀이 발생해 해가 지날수록 그 수가 증가하고 있다. 이 뿐만이 아니라, 평택시에서는 올해 6월 포트홀이 원인으로 추정되는 교통 사망사고가 발생하면서 이에 대한 운전자의 불감증은 계속해서 커지고 있다.

이러한 파손 중 도로함몰이라는 단어는 최근에 들어서야 그 의미가 정립되고 있다. 과거에는 도로에서 발생하는 포장의 함몰을 크게 포트홀(Pothole)과 싱크홀(Sinkhole)로 구분했지만, 서울시에서 도로함몰 관리대책을 수립(2014.08)하면서 그 의미가 세분화 및 재정립되었다. 이와 관련하여, 서울시에서는 도로에서 발생할 수 있는 파손의 형태를 싱크홀, 지반침하, 포트홀, 도로함몰로 구분하여 대책을 수립하고 있다.

최근 도로파손으로 발생하는 재난을 저감시키기 위해 각 지자체 및 도로관리소에서 다양한 노력을 기울이고 있다. 서울시에서는 다채널 투과레이더(3D Ground Penetrating Radar, GPR)와 Road Survey Vehicle (RSV)장비를 도입하여 서울시 내 동공탐사를 수행하고 있으며, 서울시 자체의 포장 상태 지수(Seoul Pavement Index, SPI)를 개발하여 적용하고 있다. 한국도로공사에서는 고속도로의 감시 및 순찰체계를 강화하여 유지보수를 실시하고, 최근 ICT센터에서는 포트홀을 자동 탐지하도록 하는 Artificial Intelligence (AI)기술 중 딥러닝(Deep Learning)기법을 적용하여 신기술을 개발하고 있다. 또한, 경기도와 서울시에서는 택시와 같은 대중교통 수단을 이용해 도로 손상을 신고하여 이에 대해 포상하는 제도를 정책적으로 활용하고 있다. 이와 같은 지자체 및 관리사업소의 노력에도 불구하고, 이러한 접근방법은 국가적 표준화 혹은 운영 및 예측에 한계가 있으며, 실제 도로파손은 우리나라에서 꾸준히 발생하고 있는 상황이다. 이러한 포장의 파손은 직접적인 사고로 이어지며, 이 사고는 다시 2차 사고를 유발하는 원인이 되기도 하기 때문에 도로 파손에 대한 대책이 시급하다고 할 수 있다.

한편, 정보기술면에서 보면 최근 우리나라는 정보개방으로 인해 도로 및 환경에 대한 정보를 누구나 쉽게 취득할 수 있는 사회가 되었다. 이러한 빅데이터로 인해 운전자는 쉽게 자신의 목적지까지 도달하는 다차원(거리, 시간, 날씨, 도로상황, 상태 등)의 정보를 접할 수 있다.

이와 관련하여 본 연구에서는 지도학습적 머신러닝을 적용하기 위해, 빅데이터의 처리 과정에서 역학적-확률적 접근방법(Mechanical-Statistical Approach)을 적용하여 도로에서 발생할 수 있는 파손에 영향을 미칠 수 있는 주성분을 파악했다. 또한, 예측 모형을 전국적인 범위로 보다 정확하게 확장시키기 위한 일환으로 우리나라에서 도로함몰과 포트

홀의 발생이 가장 빈번한 지역을 선정하여 예측 모형을 개발했다. 예측모형을 개발하는 과정에서는 인공지능 기법의 한 종류인 머신러닝(Machine Learning)을 적용하여 도로 파손 예측모형을 개발했다.

2. 도로 파손 형태 및 원인

교통이 개방된 도로포장에서 발생할 수 있는 손상은 소성 변형, 거북등균열, 종방향균열, 반사균열 등 다양한 형태로 발생될 수 있다. 그러나 본 장에서는 최근 사회적 문제로 부상하고 있는 균열에 대하여 중점적으로 설명하고자 한다. 도로에서 발생하는 손상 중 포장 표층에서 발생하는 파손은 차량의 주행능력을 저해하며, 차선도색과 같은 교통안전시설에 영향을 미치기 때문에 사고와 직·간접적으로 관련이 깊다. 이와 관련하여, 최근 서울시에서는 이러한 표면에서 발생하는 포장의 파손에 관련된 용어를 우리나라의 환경에 맞춰 재정립했다(Bae et al., 2016). 재정립된 용어로는 포트홀, 싱크홀, 지반침하, 도로함몰이 있다. 이 포장 파손형태들은 발생원인이 각기 다르며, 그에 대한 유지보수방안 또한 상이하다.

2.1 포트홀(Pothole)

포트홀은 표층에서부터 포장 층 내부까지 아스팔트 포장이 국부적으로 손상되어 포장의 부재가 발생하는 것을 의미한다. 포트홀은 다양한 원인에 의해 발생하는데 그 중 가장 중요하게 작용하는 것이 기상현상이다. 우기 시에는 아스팔트 포장 내부가 포화상태가 되고, 이 포화된 포장에 교통하중이 가해져 발생하는 간극수압으로 인해 포트홀이 발생할 수 있다. 여름철 우기에는 고온으로 인한 소성변형으로 성능이 저하된 경우 더 극심하게 발생할 수 있다. 또한, 동절기에는 반복적으로 일어나는 동결융해 작용으로 골재간의 결합력이 저하되며, 골재간의 결합력이 저하된 포장에 교통하중이 작용해 포트홀이 발생될 수 있다. 동절기 역시 제설제에 의해 빙설(氷雪)이 녹아 포장이 포화되는 경우 역시 포트홀의 발생가능성이 높아진다. 이 뿐만이 아니라 포장공사에서 다짐이 적절하지 않은 경우에는 재료분리 및 포장 내 침투수로 인해 교통하중으로부터 포트홀이 발생할 수 있다.

2.2 싱크홀(Sinkhole) 및 지반침하(Soil Settlement)

싱크홀과 지반침하는 유사한 용어로 활용되지만, 공학적으로 엄연히 다른 의미를 가지고 있다. 싱크홀은 석회암으로 구성된 지반에 지하수가 유입됨으로 인해 지반이 용해되어 동공을 생성하고, 이 동공의 범위가 확장되어 포장이 교통하중에 저항하는 것이 불가능할 정도로 충분히 상부층을 지지할 수 없는 경우 발생한다. 혹은, 화산재가 주된 지반에서는 지하수나 우수로 인한 지반 내 침투로 인해 순간적으로

하부층에 세굴현상이 일어나는 경우 싱크홀이 발생한다. 반면, 지반침하는 포장층 하부 토사나 암반층의 용해나 성분으로 인한 현상이 아닌, 다양한 원인에 의해 동공이 발생하는 것을 의미한다. 이러한 지반 침하는 도로 성토부의 다짐이 적절하지 못한 경우, 우수의 침투흐름으로 인해 토사가 유실되는 경우 발생한다.

2.3 도로함몰(Road Subsidence)

도로함몰이라는 용어는 2014년 8월 서울시의 ‘도로함몰 관리대책 수립’시 정립되면서 새롭게 사용되기 시작했다. 이는 기반시설의 영향으로 인해 도로 포장이 손상된다는 점에서 싱크홀이나 지반침하와 다른 의미를 갖는다. 도로함몰은 상·하수도과 같은 지하매설물이 노후되어 손상된 경우, 이 시설물 내부로 포장 하부 토사가 유입되어 교통하중에 저항하는 토사가 유실되어 발생한다. 또한, 지하시설물을 신설하는 공사의 굴착과정에서 발생한 동공으로 인해 지지력을 잃고 포장층이 동공 내부로 가라앉아 도로함몰이 발생할 수 있다.

최근 이러한 포장 파손현상이 국내에서 빈번하게 발생하고 있다. Fig. 3은 이에 대한 발생현상을 나타낸다.

Fig. 3. Pavement & Soil Damage Types

(a)와 같은 포트홀에 대해 유지보수를 실시하는 경우, 파손의 형태 및 크기에 따라 패칭, 오버레이, 절삭 후 덧씌우기, 일부단면 및 전단면 재포장 등의 공법을 선정하여 적용한다. 2012년 인천에서 발생한 싱크홀(b)이나 2006년 중국의 북부고속도로에서 발생한 지반침하(c)와 같은 경우, 연약지반 혹은 동공의 내부를 양질의 토사로 치환 및 다짐을 실시한 후 상포 포장층을 보수하여 대응한다. 2014년 서울 교대역 근방에서 발생한 도로침하(d)의 유지보수는 기존 지하시설물을 보강·보수하고, 내부 토사를 치환한 다음 포장층에

대한 보수를 실시한다.

3. 도로 파손의 발생현상

국도교통부와 서울시의 조사에 따르면, 최근 2014년부터 2016년 사이에 발생한 지반침하 현상 중 평균 79%가 서울시에서 발생한 것으로 나타났다. 이러한 현상이 서울시에 집중되는 것을 근거로 하여, 본 연구에서는 서울시를 목표로 하여 예측모형을 개발했다. Figs. 4 및 5는 지반침하 발생에 대한 최근 3년간 국도교통부의 조사 결과와 서울시에서 최근 9년간 발생한 포트홀 및 지반침하 & 도로함몰 발생량을 나타낸 것이다.

Fig. 4. Soil Settlement Occurrence in South Korea

Fig. 5. Road Failure Occurrence in Seoul

포트홀과 지반침하의 발생이 서울시에 집중된다는 것은 인구 집중화현상, 도심지의 열섬현상 등의 다양한 현상과 복합적인 환경요인에 의한 것으로 판단할 수 있다. 이에 대한 현상을 명확하게 파악하기 위해 Fig. 6과 같이 2009년부터 2017년까지 매 달 서울시에서 발생한 포트홀에 대하여 발생빈도를 분석했다. Fig. 6 및 Table 1은 본 연구에서 수행한 발생빈도 분석결과를 나타낸다. Table 1은 가장 최저 포트홀 발생을 기준으로 매달 발생하는 포트홀의 발생률을 파악했으며, 각 해마다 최저 포트홀 발생수를 기준으로 매달 발생하는 포트홀 발생률을 산정했다.

그 결과, 최근 서울시에서는 여름철에 포트홀이 가장 많이 발생하는 경향을 보였으며, 연간 총 포트홀 발생량은 2010년에 가장 많이 발생했다. 사계절의 경향으로 봤을 때, 여름철의 우기에 포트홀 발생빈도가 가장 높았으며, 그 다음으로 겨울, 해빙기인 봄인 것으로 나타났다.

기존 연구에서는 서울시 내 하부의 지반이 대부분 충적층, 호상편마암층, 화강암층으로 구성되어 싱크홀의 발생가능성이 적은 것으로 판단했다(Bae et al., 2016).

Table 1. Pothole Occurrence Rates

	Month											
	1	2	3	4	5	6	7	8	9	10	11	12
2009	1.6 (1.6)	2.1 (2.1)	3.1 (3.1)	2.9 (2.9)	3.0 (3.0)	2.4 (2.4)	6.6 (6.6)	3.9 (3.9)	1.8 (1.8)	1.5 (1.5)	1.0 (1.0)	1.0 (1.0)
2010	11.3 (7.0)	11.0 (6.8)	11.8 (7.3)	5.1 (3.2)	4.2 (2.6)	3.6 (2.2)	6.2 (3.9)	9.6 (6.0)	13 (8.0)	2.9 (1.8)	1.7 (1.1)	1.6 (1.0)
2011	2.3 (1.8)	2.9 (2.4)	4.3 (3.5)	4.2 (3.4)	5.0 (4.1)	7.8 (6.4)	17.5 (14)	4.6 (3.8)	2.6 (2.1)	1.5 (1.2)	1.2 (1.0)	1.5 (1.2)
2012	1.5 (1.0)	1.7 (1.2)	2.8 (1.9)	3.9 (2.7)	2.8 (1.9)	2.0 (1.4)	4.9 (3.4)	7.7 (5.3)	5.5 (3.8)	1.6 (1.1)	1.9 (1.3)	7.3 (5.0)
2013	12.0 (9.5)	15.1 (12)	6.4 (5.0)	7.3 (5.8)	4.9 (3.8)	3.5 (2.72)	16 (12)	3.9 (3.1)	4.0 (3.1)	1.8 (1.4)	1.3 (1.0)	1.9 (1.5)
2014	2.8 (2.1)	3.6 (2.6)	1.9 (1.4)	2.4 (1.8)	2.6 (1.9)	2.7 (2.0)	4.2 (3.1)	4.2 (3.1)	2.9 (2.1)	2.0 (1.5)	1.4 (1.0)	1.7 (1.3)
2015	2.4 (1.1)	3.0 (1.4)	3.8 (1.8)	4.9 (2.3)	2.9 (1.4)	2.9 (1.4)	5.9 (2.8)	3.4 (1.6)	2.3 (1.1)	2.1 (1.0)	2.7 (1.3)	2.7 (1.3)
2016	2.9 (1.6)	2.5 (1.4)	3.7 (2.1)	3.2 (1.8)	3.7 (2.0)	2.9 (1.6)	6.5 (3.6)	3.7 (2.0)	2.9 (1.6)	2.7 (1.5)	1.8 (1.0)	2.4 (1.3)
2017	2.3 (2.0)	4.1 (3.6)	3.0 (2.6)	3.0 (2.6)	2.0 (1.8)	2.0 (1.7)	6.5 (5.6)	6.6 (5.7)	2.0 (1.8)	1.5 (1.3)	1.2 (1.0)	1.6 (1.4)

Fig. 6. Pothole Occurrence Analysis

기존 연구(Kim, Han et al., 2014)에서는 포트홀 발생에 대한 예측은 기상 및 교통량에 관련된다는 결론을 내렸다(Lee et al., 2014; Lee et al., 2017).

2014년 도로함몰이 정의되기 전 서울시 현장 조사에서는 지반침하와 도로함몰을 한 종류의 균열로 간주하여 관측했다. Fig. 7은 서울시에서 발생한 도로함몰 수를 나타낸다.

Fig. 7. Road Subsidence & Soil Settlement occurrence in Seoul (2010~2015)

4. 역학적-확률적 접근을 활용한 빅데이터 처리

전 세계의 급속한 기술의 발전으로 인해 현대사회에서는 정보의 홍수라고 할 정도로 개인이 많은 정보를 쉽게 습득할 수 있게 되었다. 최근 우리나라에서도 국가적 데이터를 국민에게 보다 효율적으로 제공하고, 정립하려는 목적으로 Open API (Application Programming Interface)를 개설해 다양한 정보를 제공하고 있다. 본 연구에서는 우리나라에서 발생하는 포트홀과 지반침하 및 도로함몰에 대한 예측모형을 확장시키려는 목적으로 서울시에서 발생하는 포트홀 및 지반침하 & 도로함몰 예측 모형을 개발했다.

이와 관련하여 국가적 정보 중 포트홀 발생과 관련된 데이터를 재정립할 필요가 있기 때문에 본 연구에서는 역학적-확률적 접근을 적용하여 데이터해석을 실시했다.

한국에서는 도로를 설계하고 유지관리를 실시하는 경우, 한국형 포장설계법(Korea Pavement Design Guideline, KPDG)에 기초하여 한국형포장설계 프로그램을 사용해야 한다. 이 한국형포장설계프로그램은 역학적-경험적 설계법(Mechanical-Empirical Pavement Design Guide, M-EPDG)에 기초하고 있다. M-EPDG는 복잡한 포장의 거동 및 파괴 메커니즘에 영향을 미치는 요소들에 대해 역학적으로 접근하여 공용성을 평가하고, 경험적인 원시 과거 데이터의 분석을 통해 주인자를 파악하는 과정으로 개발되었다. 이처럼 도로공학에서는 역학적인 공학적 접근과 다른 접근방법을 접목하여 기술을 개발한다. 이와 관련하여 본 장에서는 과거에 서울시에서 발생한 포트홀 및 지반침하 & 도로함몰에 영향을 미치는 주요인을 파악하기 위해 역학적-확률적 접근을 시도했다. 빅데이터를 처리하기 위해 먼저 역학적으로 접근하여 데이터 중 일부를 정제했으며, 이를 다시 주인자 분석을 통해 확률적으로 가장 적합하다고 판단되는 인자를 다시 재분류했다.

4.1 역학적 접근법(Mechanical Approach)

아스팔트 혼합물은 점탄성(Visco-Elasticity)물질로써, 그 성능을 나타내는 동탄성계수가 하중과 온도에 따라 유동성을 갖는다. 이러한 이유로 동탄성계수시험(Dynamic Modulus Test)를 실시한 후, 시간-온도 중첩의 원리(Time-Temperature Superposition Principle, TTSP)를 적용하여 아스팔트 혼합물의 실제 성능을 나타내는 Sigmoidal 함수의 동탄성계수 마스터커브(Master Curve)를 결정한다. 이는 아스팔트 포장이 기온과 교통하중에 민감하다는 것을 의미한다. 또한, 아스팔트 포장은 수분에 매우 민감하여 실제 투수 포장에 적용되는 아스팔트는 별도의 KS규격에 적합한 수분저항성 시험을 실시해야 한다. 이와 관련하여, 강수량과 겨울철 제설제로 인한 강설량은 아스팔트 포장의 성능과 매우 밀접한 관계를 갖는다.

더불어, 우리나라에서는 크게 3가지의 포장에 대한 상태 지수를 사용하고 있다. 이 지수들은 포장의 평탄성(m/km), 피로 파괴률(%), 소성변형(mm)의 크기에 따라 평가되며, 이는 손상된 포장의 유지관리에 직접적으로 사용된다.

우리나라에서는 일반적으로(National Highway Pavement Condition Index, NHPCI)의 함수를 사용하고 있으나, 서울시에서는 SPI, 한국도로공사에서는 Highway Pavement Condition Index (HPCI)라는 별도의 포장상태지수를 사용한다(Son et al., 2013). 이 포장의 상태지수는 현재 포장의 상태를 나타내기 때문에 포장의 파손과 직접적으로 관계가 있다고 할 수 있다. Eqs. (1), (2) 및 (3)은 우리나라에서 사용하고 있는 아스콘 포장의 NHPCI, SPI 및 HPCI 결정식을 나타낸다.

$$NHPCI = \frac{1}{(0.33 + 0.003 \times Cr + 0.004 \times RD + 0.0183 \times IRI)^2} \quad (1)$$

$$SPI = 10 - \left[\left\{ 10 - (10 - 2.23 \times Cr^{0.3}) \right\}^5 + \left\{ 10 - (10 - 0.2 \times RD) \right\}^5 + \left\{ 10 - (10 - 0.667 \times IRI) \right\}^5 \right]^{1/5} \quad (2)$$

$$HPCI = 4.564 - 0.348 \times IRI - 0.36 \times RD - 0.01 \sqrt{5(Cr^* + A)} \quad (3)$$

여기서, Cr , Cr^* , RD , IRI 및 A 는 각각 균열, 온도균열, 소성변형, 종단 평탄성 및 단위 구역 당 거북등 균열과 패칭 적용 면적을 의미한다.

결론적으로, 기상에 대한 빅데이터 중 온도, 수분 등과 관련된 데이터를 별도로 수집했으며, 서울시에서 제공하고 있는 교통량 데이터, 평균 주행속도 및 SPI를 일괄적으로 수집했다.

4.2 확률적 접근법(Statistical Approach)

역학적으로 아스팔트 혼합물의 상태에 영향을 주는 인자

들을 분류한 후, 이를 이용하여 본 장에서는 확률적 접근방법을 적용하여 포트홀 발생에 영향을 미치는 요인을 파악했다. 역학적인 접근방법으로 교통량, SPI, 온도, 강수 및 강설이 아스팔트 포장에 영향을 미치는 것으로 판단했으나, 실제 확률적으로 영향을 미치는 요인으로 작용하는지 확인할 필요가 있다. 또한, 온도, 강수 및 강설에 대한 공공자료 역시, 다양한 종류로 세분화되어 제공되고 있기 때문에 실제 영향을 주는 인자를 파악할 필요가 있다. 이와 관련하여 최대 우도 추정법(Maximum Likelihood Estimation method, MLE method)에 기초한 탐색적 요인분석(Exploratory Factor Analysis)을 실시하여 종속변수에 대한 독립변수의 영향성을 검토했다.

기상과 관련된 오픈데이터는 총 47종류이며, 수치적으로 사용될 수 있는 실제 데이터는 34가지(월평균기온, 월평균최고기온, 월평균최저기온, 월최고기온, 월최저기온, 월평균수증기압, 월최고수증기압, 월최저수증기압, 월평균이슬점온도, 월평균상대습도, 월최소상대습도, 월합강수량, 일최대강수량, 1시간최대강수량, 10분최대강수량, 소형충증발량, 소형일최대증발량, 대형충증발량, 대형일최대증발량, 최심적설, 최심신적설, 일적설량합, 평균최저초상온도, 최저초상온도, 평균지면온도, 0.05 m 평균지중온도, 0.1 m 평균지중온도, 0.2 m 평균지중온도, 0.3 m 평균지중온도, 0.5 m 평균지중온도, 1.0 m 평균지중온도, 1.5 m 평균지중온도, 3.0 m 평균지중온도, 5.0 m 평균지중온도)이다. 그러나 이 변수들 사이에서 강한 상관성을 가져 공통요인으로 작용하는 변수가 존재할 수 있으며, 종속변인과 유의성을 갖지 않는 독립인자가 존재할 수 있기 때문에 이에 대한 차원축소를 실시했다.

독립변수 간의 선형적 구조($[x_i] = [l_{ij}][f_j]$)와 오차(η_i)항으로 가정하여 확률적 변수의 내재 관계를 파악하는 탐색적 요인분석은 각각의 상호 독립적인 f_k 와 η_p 는 모두 평균이 0인 동일 분포를 따르지만 공통인자는 분산이 1인 분포를 갖으며, 특정인자는 ψ_j 를 갖는 분산으로 분포한다($k=1,2,...,i$, $p=1,2,...,j$). 이는 MLE 방법으로 Eq. (4)를 이용하여 도출할 수 있다.

$$\max_{L, \Psi} L(\mu, \Sigma|x) = L(\mu, LL' + \Psi|x) \quad (4)$$

여기서, f_p 및 η_p 가 서로 독립적이며, x_p 가 다변량의 정규분포를 따른다고 가정된다.

이에 대한 탐색적 요인분석 결과의 스크리 도표(Scree Plot)는 Fig. 8과 같다.

탐색적 요인분석 결과로부터 5개의 군집 이상에서는 독립변수들의 설명력이 줄어들며, Kaiser-Meyer-Olkin 측도 및 Bartlett 검정으로부터 기온, 수증기압, 적설량, 초상온도 및 강수량의 5개 군집으로 차원을 감소시켰다. 이 결과를

토대로 실제 포트홀과 지반침하 및 도로함몰에 영향을 미치는 변수를 파악하기 위해 군집 별 선형회귀에 대한 다중공선성, 잔차의 독립성, 분산분석의 유의수준 및 상관관계를 확인했다. 그 결과, 기상자료 중 포트홀과 지반침하 및 도로함몰에 통계적으로 영향을 미치는 요소는 Table 2와 같다.

Fig. 8. Scree Plot for Exploratory Factor Analysis

Table 2. Determination of Independent Variables

	x_1 ($^{\circ}\text{C}$)	x_2 (hPa)	x_3 (mm)	x_4 ($^{\circ}\text{C}$)	x_5 (mm)
Pothole	Max Temp	Min Vapor Pres	Extreme Snowdrifts	Avg. Min Grass Temp	Daily Max Rainfall
Soil Settlement & Road Subsidence	Min Temp	Max Vapor Pres	Extreme Snowdrifts	Avg. Min Grass Temp	Minute Max Rainfall

결론적으로 본 장에서는 역학적인 접근방법을 사용하여 도로의 파손에 영향을 미치는 3가지 요인(교통량, 평균주행속도 및 SPI)을 결정했으며, 통계적인 접근방법을 이용해 5가지 요인(Table 2)을 추가적으로 파악했다.

5. 예측모형 결정을 위한 머신러닝의 적용

최근 AI기술은 다양한 분야를 아울러 적용되고 있는 빅데이터 처리 및 자동 학습 접근 방법이다. AI기술의 일종인 머신러닝 기법은 단어 그대로 기계(Machine)가 빅데이터를 수렴하는 동시에 이를 자동적인 확률적 학습(Learning)을 통해 최적의 해(Solution)나 방법(Method)을 결정하는 것에 그 목적을 두고 있다. 머신러닝 기법은 비지도학습(Unsupervised Learning)과 지도학습(Supervised Learning)으로 구분되는데, 비지도학습은 클러스터링(Clustering)기법을 사용하여 빅데이터 자체가 내포하고 있는 고유 구조를 해석하고 추론하는 과정을 포함한다. 그러나 본 연구는 수집한 데이터와

반응변수 간의 불확실성을 가지고 있기 때문에 지도학습과 정의 머신러닝을 적용시켰다. 입·출력 데이터에 기반하여 예측모형을 개발하는 지도학습법을 적용하는 과정에서 데이터 자체의 차원에서 발생할 수 있는 불확실성을 제어하기 위해 Eq. (5)와 같이 데이터 표준화(Data Standardization)작업을 실시했다.

$$x_s = \frac{x_r - x_{\min}}{x_{\max} - x_{\min}} \quad (5)$$

여기서, x_s , x_r , $x_{\min}$ 및 $x_{\max}$ 는 각각 표준화 값, 데이터 실측값, 최소값 및 최대값을 의미한다.

본 연구에서 활용한 머신러닝의 기본적 메카니즘은 Fig. 9와 같다.

Fig. 9. Flowchart for Machine Learning

5.1 머신러닝의 구성 및 적용

본 연구에서 활용한 머신러닝은 오차비교를 통한 최적화 알고리즘 결정기법을 적용했다. 개별적 알고리즘을 구성하는 과정에서도 기본적으로 오차최소화 기법이 적용되었다. 본 연구에서는 머신러닝 적용에 총 19가지 알고리즘을 구현하여 이들을 평가했다. 본 연구에서 구축한 머신러닝의 19가지 알고리즘은 다중선형회귀(op_1), 상호작용 다중선형회귀(op_2), 로버스트회귀(op_3), 단계적 선형회귀(op_4), 조밀 회귀나무(op_5), 중간 회귀나무(op_6), 성긴 회귀나무(op_7), 선형 SVM(Support Vector Machine)(op_8), 2차 SVM(op_9), 3차 SVM(op_{10}), 조밀 SVM(op_{11}), 중간 SVM(op_{12}), 성긴 SVM(op_{13}), 부스팅의사결정나무(op_{14}), 배깅의사결정나무(op_{15}), 제곱지수 GPR (Gaussian Process Regression)(op_{16}), 매턴 5/2 GPR(op_{17}), 지수 GPR(op_{18}) 및 유리2차 GPR(op_{19})이다. 본 연구에서는 포트홀 발생수와 도로함몰 및 지반침하 발생수에 대한 2가지 예측모형을 구성하는 것으로써 Fig. 9의 과정을 2회 반복했다. 본 연구에서 적용한 알고리즘은 크게 다중회귀,

SVM, 의사결정나무 및 가우스과정회귀의 4가지로 구성되었다. 본 장에서는 이를 설명하고자 한다.

5.1.1 다중회귀(Multiple Regression)

다중회귀는 2개 이상의 독립변수(X)와 반응변수(Y)간의 관계를 설명하는 분석방법이다. 일반적으로 독립변수수와 데이터 수가 같은 경우, 모수의 추정값($\hat{\beta}$)은 Eq. (6)과 같이 행렬간의 Determinant를 계산하여 구할 수 있다.

$$\min \sum_{i=1}^N \hat{\epsilon}_i^2 = \sum_{i=1}^N (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^N [Y_i] - [\hat{\beta}_i][X_i]^T)^2 \quad (6)$$

$$\Rightarrow \hat{\beta} = (X'X)^{-1}X'Y$$

여기서, N 및 $\hat{\epsilon}$ 은 독립변수의 수($i=1 \sim N$)와 잔차제곱합(Regression Sum of Squares, RSS)을 의미한다.

그러나, 본 연구와 같이 데이터의 수가 독립변수보다 많은 경우 오차를 최소화시키기 위해 추정 모수를 찾는 학습과정이 필요하며, 상호작용 회귀, 로버스트 회귀 및 단계적 선형회귀를 수행하는 경우 새로운 독립변수에 대한 추가적인 학습과정을 수행해야 한다. 이 학습과정에서 유클리드 거리(Euclidean Distance)척도에 기초한 K-Nearest Neighbor (K-NN)을 사용하여 최근방 데이터를 정렬한 뒤 잔차제곱합을 최소화하는 방법으로 모수에 대한 추정을 실시했다.

5.1.2 SVM (Support Vector Machine)

SVM은 초평면(Hyperplane)을 이용하여 수집된 데이터를 하나의 군집으로 분류한 뒤, 최대 마진(Margin)을 갖는 서포트벡터(Support Vector)를 파악하는 최적화 방법이다. 즉, 이는 데이터로부터 초평면까지의 거리벡터를 양과 음으로 구분하여 벡터상의 거리를 최대로 하는 서포트벡터를 탐색하는 것이다. 이는 비선형인 경우 변수공간을 확장시키기 위해 확장방법에 따라 특정 커널(Kernel)을 사용한다. Eq. (7)은 단순 SVM에 대한 함수이다.

$$\text{Maximize } \beta_0, \dots, \beta_p, \epsilon_1, \dots, \epsilon_n \quad M = |d_+| + |d_-| = \frac{2}{|\omega|} \quad (7)$$

$$\text{Subject to } \sum_{i=1}^p \beta_j^2 = 1$$

$$Y_i(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}) \geq M(1 - \epsilon_i)$$

$$\epsilon_i \geq 0, \sum_{i=1}^n \epsilon_i \leq C$$

$$\text{where, } d_+ = \frac{\omega^T x_+ + \omega_0}{\omega} = \frac{1}{|\omega|}$$

$$d_- = \frac{\omega^T x_- + \omega_0}{\omega} = \frac{-1}{|\omega|}$$

여기서, C, M 및 ϵ 은 각각 조율파라미터, 마진폭 및 초평면

에 대한 슬랙변수(Slack variable)이다.

또한, k차 SVM에 대한 함수는 선형 Y_i 의 함수가 β 와 X 에 대한 k차의 함수를 대포하게 된다. p 개의 독립변수를 k승수의 변수 $k \cdot p$ 개의 변수로 사용하게 될 경우 서포트벡터 분류기를 Eq. (8)과 같이 적합 시킬 수 있다.

$$\begin{aligned} & \text{Maximize} \\ & \beta_0, \beta_{1,1}, \dots, \beta_{p,k}, \epsilon_1, \dots, \epsilon_n \quad M \quad (8) \\ & \text{Subject to } Y_i(\beta_0 + \sum_{j=1}^p \beta_{j,1}x_{i,j} + \dots + \sum_{j=1}^p \beta_{j,k}x_{i,j}^k) \geq M(1 - \epsilon_i) \\ & \epsilon_i \geq 0, \sum_{i=1}^n \epsilon_i \leq C, \sum_{j=1}^p \sum_{o=1}^k \beta_{j,o}^k = 1 \end{aligned}$$

위 식의 관측치의 벡터 내적에 대한 선형, 가우스 및 다항에 대한 서포트벡터 분류기의 커널함수는 Table 3과 같다.

Table 3. Kernel Functions for SVM

Kernel	Function
Linear	$f(x_j, x_k) = x_j' x_k$
Gaussian	$f(x_j, x_k) = e^{(-\ x_j - x_k\ ^2)}$
Polynomial	$f(x_j, x_k) = (1 + x_j' x_k)^q$ where, $q = 2, 3, \dots$

5.1.3 의사결정 및 회귀나무

회귀나무는 뿌리(Root)로부터 Child Node와 Parent Node를 구성하여 leaf Node를 결정해 나가는 의사결정기법이다. 이에 대한 분기수와 가중치 값은 반복되는 훈련에 의하여 결정된다. 이 과정은 순도(Homogeneity)가 최대가 되며, 불순도(Impurity) 또는 불확실성(Uncertainty)이 최소가 되는 방향으로 의사결정을 수행한다. 이에 대한 지표로 엔트로피(Entropy)를 사용하는데, d개의 데이터 중 특정 A영역에 속하는 비율(R_i)로 나타내며, 다시 이 영역내의 i영역을 새로 이 분할하여 특정 데이터 m개가 이 영역에 속하는 비율(p_k)로 표현한다. 이 과정에서 엔트로피를 저감시키는 방향으로 의사결정을 수행한다. Eq. (9)는 i영역에 대한 엔트로피 함수를 나타낸다.

$$Entropy(A) = \sum_{i=1}^d R_i (-\sum_{k=1}^m p_k \log_2(p_k)) \quad (9)$$

이 엔트로피 기준이 저감될수록 불확실성이 감소하며 순도가 증가됨을 의미한다. 이 값은 회귀나무를 결정할 때 정보획득(Information Gain, IG)값이라는 상대 엔트로피로 적용하며, IG는 Parent에 대한 엔트로피에서 Child의 엔트로피 값의 차로 정의된다.

배깅트리는 독립변수에서 동일한 크기로 랜덤으로 추출

하여 구성된 변수를 개별적으로 훈련시켜 트리를 구성하는 방식으로 의사결정하며, 부스팅트리는 반복적으로 취약한 학습을 중심으로 이에 대한 가중치를 조정하여 전체적으로 학습을 강화시키는 방법이다. 이 두 가지 방법은 앙상블(Ensemble)기법으로 불리며 개별적인 러닝머신의 알고리즘을 보완하는 다중 학습 알고리즘기법이다.

회귀나무에서는 조밀(fine), 중간(Medium), 성긴(Coarse) 나무들로 구성할 수 있는데, 이는 반응함수에 대한 Leaf Nodes (leaves)의 유연도(flexibility)에 따라 구분된다. 여기서, 조밀나무(Fine Tree)는 최소 잎의 수를 4개로 제한하며, 중간나무(Medium Tree)는 12개, 성긴나무(Coarse Tree)는 36개로 제한했다.

5.1.4 가우스과정회귀(Gaussian Process Regression, GPR)

가우스과정회귀 분석은 연속 종속변수를 예측하기 위해 사용되는 비모수적 예측모형이다. 가우스과정을 회귀함수에 적용한 사전확률(GP)을 표현하면 입력(x)에 대한 출력($f(x)$)함수는 평균함수($m(x)$)와 커널($k(x, x')$)로 구성되어 Eq. (10)과 같이 나타낼 수 있다.

$$\begin{aligned} f(x) & \sim GP(m(x), k(x, x')) \\ m(x) & = E[f(x)] \\ k(x, x') & = E[(f(x) - m(x))(f(x') - m(x'))^T] \end{aligned} \quad (10)$$

결국, 데이터 내의 결합 가우스과정을 다시 표현하면 Eq. (11)과 같이 표현할 수 있다.

$$p(f|x) = N(f|\mu, K_{ij}) \quad (11)$$

여기서, $K_{ij} = k(x_i, x_j)$ 의 커널함수이며, $\mu = (m(x_1), \dots, m(x_N))$ 을 나타낸다.

위 식으로부터 커널의 적용성에 따라 가우스과정회귀는 다른 출력을 나타내게 된다. 본 연구에서는 가우스 과정 회귀에 2차 지수, 지수, 매턴 5/2 및 유리 2차 커널을 적용했다. 이에 대한 커널함수는 Table 4와 같다.

Table 4. Kernel Functions for GPR

Kernel	Function
Exponential	$k(x_i, x_j \theta) = \sigma_f^2 e^{\frac{-r}{\sigma_\ell}}$
Squared Exponential	$k(x_i, x_j \theta) = \sigma_f^2 e^{\frac{(x_i - x_j)^T(x_i - x_j)}{-2\sigma_\ell^2}}$
Matern 5/2	$k(x_i, x_j \theta) = \sigma_f^2 (1 + \frac{\sqrt{5}r}{\sigma_\ell} + \frac{5r^2}{3\sigma_\ell^2}) e^{\frac{-\sqrt{5}r}{\sigma_\ell}}$
Rational Quadratic	$k(x_i, x_j \theta) = \sigma_f^2 (1 + \frac{r^2}{2\alpha\sigma_\ell^2})^{-\alpha}$

여기서, ℓ 은 함수변이의 수평길이를 조정하며 σ_f 및 σ_ℓ 은 각각 함수의 수직 및 수평적 변이를 조절한다. α 는 양수의 스케일을 조정하는 파라미터로 작용하며, 2차 지수 커널을 제외한 다른 커널함수에서 사용한 r 은 x_i 와 x_j 에 대한 유클리드 거리($r = \sqrt{(x_i - x_j)^T(x_i - x_j)}$)를 나타낸다.

5.2 머신러닝의 적용 결과

본 연구에서 반응 데이터의 학습은 오차에 기준을 두어 수행되었다. 따라서 각 알고리즘의 적용에서 훈련된 최적의 오차를 이용하여 최적 알고리즘을 도출했으며, 이 오차의 지표는 RMSE를 사용했다. 머신러닝은 다양한 알고리즘을 해석해야하며 빅데이터를 처리해야하기 때문에 AI기술에서 정확도만큼 중요한 것이 데이터에 대한 학습속도이다. 이에 대한 머신러닝의 적용결과는 Figs. 10, 11, Tables 5 및 6과 같다. 여기서, MSE 및 MAE는 각각 Mean Squared Error 및 Mean Absolute Error이다. Figs. 10 및 11의 가로와 세로축은 관측 값과 반응 값에 대한 결과이며, 그래프의 대각선은 예측과 실측에 대해 완벽하게 일치하는 절대선(Absolute Line)을 나타낸다.

Fig. 10. Application Results of Machine Learning (Pothole Number)

Fig. 10. (Continued)

Fig. 11. Application Results of Machine Learning (Soil Settlement & Road Subsidence)

Fig. 11. (Continued)

Table 5. Optimization Results (Pothole Number)

	RMSE	MSE	MAE	Learning Velocity (Data/sec)	Learning Time (sec)
op_1	0.1318	0.017372	0.094139	850	7.6916
op_2	0.13923	0.019385	0.099417	580	7.4308
op_3	0.15966	0.02549	0.09236	870	7.2333
op_4	0.14194	0.020148	0.095866	2400	20.914
op_5	0.15648	0.024487	0.10171	1200	7.0202
op_6	0.1726	0.029791	0.11197	1200	6.8695
op_7	0.19419	0.03771	0.12832	1100	6.6994
op_8	0.13638	0.0186	0.085162	1000	6.5188
op_9	0.21106	0.044546	0.10442	1000	6.3419
op_{10}	0.21472	0.046104	0.11938	860	6.1716
op_{11}	0.19453	0.037841	0.11407	1100	5.9747
op_{12}	0.16615	0.027606	0.09421	1000	5.7589
op_{13}	0.17096	0.029226	0.096948	2400	8.1285
op_{14}	0.15615	0.024383	0.097755	1200	7.9908
op_{15}	0.15743	0.024785	0.099482	1500	7.198
op_{16}	0.14049	0.019738	0.090859	2500	7.0497
op_{17}	0.14034	0.019694	0.088326	2600	6.8964
op_{18}	0.1362	0.018549	0.084752	2600	6.736
op_{19}	0.1412	0.019939	0.089129	2800	7.3268

Table 6. Optimization Results (Soil Settlement & Road Subsidence)

	RMSE	MSE	MAE	Learning Velocity (Data/sec)	Learning Time (sec)
op_1	0.1559	0.024306	0.12241	710	7.2834
op_2	0.24087	0.05802	0.17911	440	8.1673
op_3	0.13520	0.018279	0.101458	580	7.9963
op_4	0.15496	0.024012	0.12048	1700	15.916
op_5	0.19967	0.039869	0.14785	770	6.5165
op_6	0.18144	0.032919	0.14231	830	6.3412
op_7	0.25915	0.06716	0.21558	760	6.177
op_8	0.15612	0.024374	0.11898	620	5.9863
op_9	0.1682	0.02829	0.13065	610	5.8063
op_{10}	0.17746	0.031492	0.13624	700	5.6351
op_{11}	0.24572	0.060376	0.19936	600	5.4525
op_{12}	0.17057	0.029094	0.13065	670	5.2268
op_{13}	0.16027	0.025685	0.12382	1800	6.5679
op_{14}	0.16562	0.027431	0.13109	900	7.4963
op_{15}	0.17274	0.029838	0.14184	830	6.7067
op_{16}	0.15389	0.023684	0.12247	1500	6.5411
op_{17}	0.15483	0.023971	0.12261	1700	6.396
op_{18}	0.16505	0.027241	0.12765	1600	6.2602
op_{19}	0.15389	0.023684	0.12247	1300	6.9513

5.3 최적 예측모형의 결정

본 연구에서 수행한 러닝머신의 최적화 알고리즘 기법 적용에서 최적의 예측모형을 결정하기 위해서 RMSE, MSE 및 MAE에 대한 비교 검토를 실시했다. 그 결과, 포트홀 발생수를 예측하기 위한 최적의 모델은 다중회귀(op_1)를 사용한 모형으로 결정되었으며, 지반침하 및 도로함몰 발생수를 예측하기 위한 최적의 모델은 로버스트 회귀분석(op_3)을 사용한 모형으로 결정했다. 이 모형은 Eq. (12)와 같이 구성되어 있으며, 이에 대한 결과는 Table 7과 같다.

$$Y = \varepsilon_o + \beta_1 x_1 + \cdots + \beta_8 x_8 \quad (12)$$

이에 대한 관측값과 예측값에 대한 추세를 확인하기 위해 결정계수를 파악한 결과, 다중회귀 및 로버스트 회귀에 의한 예측모형은 각각 69.73% 및 72.49%의 동등 경향성을 갖는 것으로 나타났다.

이와 더불어 예측 모형의 정확성을 평가하기 위해 Mean Absolute Percentage Error (MAPE)를 도출한 결과, 최적 포트홀 예측모형은 약 0.132의 오차를 가지고 있는 것으로 나타났으며, 지반침하 및 도로함몰 예측 최적 모형은 약

0.135의 오차를 가지고 있는 것으로 판단되었다. 이에 대한 실측 및 예측 비교결과는 Fig. 12와 같다.

Table 7. Optimized Variables of Prediction Function

	Multiple Reg(op_1)		Robust Reg(op_3)	
	Value	P-value	Value	P-value
intercept	-0.19203	0.00391	-0.2248	0.00627
β_1	0.08915	0.62798	0.30992	0.26363
β_2	-0.00117	0.99049	-0.01901	0.94464
β_3	0.84550	7.99E-16	0.06932	0.55064
β_4	0.74171	2.25E-11	0.03423	0.79452
β_5	-0.00316	0.98806	0.35735	0.36443
β_6	0.05799	0.16150	0.00581	0.91978
β_7	0.17491	0.00144	0.37765	5.68E-06
β_8	0.04216	0.22693	0.03852	0.55114

Fig. 12. Comparative Analysis between Measurements and Predictions

6. 결 론

본 연구에서는 운전자 및 보행자의 안전성을 높이려는 목적으로 포트홀 및 지반침하 & 도로함몰 발생 예측 모델을 개발했다. 이 모델을 개발하는 과정에서, 빅데이터를 처리하기 위해 역학적·확률적 접근방법을 사용하여 새로운 데이터를 구축했다. 구축된 데이터를 이용해 데이터 표준화 작업을 거쳐 지도학습으로 분류된 인공지능 기법 중 하나인 머신러

닝을 이용하여 최적 알고리즘을 결정했다. 머신러닝 학습과 정에서는 총 19가지 알고리즘을 구현하여 오차 발생의 크기를 중심으로 최적 알고리즘을 탐색하여 예측모형을 구성했다. 이에 대한 본 연구의 결론은 다음과 같다.

- (1) 포트홀 발생에 대한 최적 예측모델은 다중회귀분석에 의해 도출되었으며, 지반침하 & 도로함몰 예측을 위한 최적 모델은 로버스트 회귀분석에 의해 도출되었다. 로버스트 회귀 분석모형에서 현재 데이터로 구성된 독립변수만으로 알고리즘을 구현한 것으로 보아 역학적·확률적 접근방법의 신뢰성이 높은 것으로 판단했다.
- (2) 예측 모형을 구성한 결과, 로버스트 회귀분석에 의한 지반침하 & 도로함몰 예측모형이 포트홀 예측 모형보다 더 높은 경향성과 낮은 오차를 보였다.
- (3) 최적화 기법 중 단계적 선형 회귀분석방법은 학습속도가 다른 기법에 비해 매우 빠름에도 불구하고, 매개변수의 조합처리로 인해 학습시간이 가장 긴 것으로 나타났다. 따라서 이는 빅데이터 처리 시 전체적인 머신러닝의 구동속도를 저하시키는 것으로 평가되었다.
- (4) Fig. 12로부터 이벤트발생에 대한 경향은 우수하지만, 오차를 포함하는 것으로 보아 이를 필터링할 수 있는 별도의 분석기법이 필요한 것으로 판단했다.
- (5) 추후 연구에서는 보완된 인공지능 기법을 적용하기 위해 비지도 학습적 머신러닝과 추론의 접근성을 높이기 위한 방안으로 베이지안 기법을 추가로 적용하여 예측모형을 개발할 것이다.
- (6) 또한, 서울시에 제한된 예측모형을 국가적 예측모형으로 확장한 후, 각 지자체 및 특정 구역에서 발생할 수 있는 사건에 대한 예측을 실시할 것이다.

감사의 글

이 논문은 2018년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. NRF-2018R1A2B6005786).

References

- Bae, Y.S., Shin, S.Y., Won J.S., and Lee, D.H. (2016). *The road subsidence conditions and safety improvement plans in Seoul*. The Seoul Institute Policy Research, 2016-PR-09, The Seoul Institute.
- Kim, D.W., Han, B.S., Kim, Y.J., and Mun, S.H. (2014). A study of reliability of predictive models for permanent deformation and fatigue failure related to flexible pavement design. *International Journal of Highway Engineering*, Vol. 16, No. 6, pp. 105-113.

Kim, K.J., Kim, Y.T., and Hwang, J.H., (2015). Factors and solutions of insensitivity to road traffic safety. *Korean Police Studies Review*, Vol. 14, No. 2, pp. 19-40.

Lee, S.Y., Kim, D.W., and Choi, Y.W. (2014). Development of predictive model for the number of potholes using multi regression analysis. *J. Korean Soc. Hazard Mitig.*, Vol. 14, No. 3, pp. 91-98.

Lee, S.Y., Kim, D.W., and Mun, S.H. (2017). Development of estimated models of the number of potholes with the statistical optimization method. *KSCE Journal of Civil Engineering*, Vol. 21, No. 7, pp. 2683-2694.

Ministry Of the Interior and Safety (MOIS). (2017). *Results*

of disaster and safety situation and disaster types to be managed by priority.

Son, H.J., Kwon, S.A., Lee, J.J., Baek, C.M., Lim, J.K., and Sin, H.J. (2013). A study for determining the rehabilitation method group using NHPPI on asphalt concrete pavement of national highway. *International Journal of Highway Engineering*, Vol. 15, No. 2, pp. 1-9.

Received	■	August 13, 2018
Revised	■	October 27, 2018
Accepted	■	November 27, 2018