랜덤 포레스트를 활용한 도로 및 교통시설 개선방향 추정 연구

A Study on Estimation of Road and Transportation Facility Improvement Direction Using Random Forest

황 재 성* · 김 도 경** · 김 남 선*** · 이 철 기****

* 주저자 : 아주대학교 교통공학과 석박사통합과정
** 공저자 : 아주대학교 교통공학과 석사과정
*** 공저자 : 경찰대학 치안정책연구소 자율주행 치안솔루션 연구센터 책임연구관
**** 교신저자 : 아주대학교 교통시스템공학과 교수

Jae-seong Hwang* · Do-kyeong Kim** · Nam-sun Kim*** · Choul-ki Lee****

* Dept. of Transportation Eng., Univ. of Ajou
** Dept. of Transportation Eng., Univ. of Ajou
*** Police Science Institue, Autonomous Driving Police solution Center
**** Dept. of Transportation Systems Eng., Univ. of Ajou

† Corresponding author : Choul ki Lee, cklee@ajou.ac.kr

Vol.20 No.6(2021)
December, 2021
pp.37~46

pISSN 1738-0774
eISSN 2384-1729
https://doi.org/10.12815/kits.2021.20.6.37

Received 25 November 2021
Revised 8 December 2021
Accepted 8 December 2021

요 약

교통사고 예방을 위해 경찰 및 지자체 등 정부기관에서는 교통시설 및 도로시설의 개선사업을 추진하여 교통 위해 요소를 제거하고 편안한 도로 환경을 조성하는데 노력하고 있다. 이를 위해 도로 및 교통시설을 개선 및 조정하며, 교통사고 잦은 지역의 개선사업이 대표적인 사업이다. 교통사고 잦은 지역의 개선사업은 담당자와 관계자의 주관에 따라 사업별, 지역별 편차가 발생하고 있으며, 우선순위 도출 등에 민원 및 주관성이 반영되어 사업의 효율성에 한계가 발생하고 있다. 이를 위해 교통사고 잦은 곳 개선사업의 효과가 높은 대표사업을 대상으로 도로여건, 교통여건, 사고여건 등을 종합적으로 고려하여 사업 대상지의 개선방향을 추정하는 연구를 진행하였다. 연구결과 개선사업 추정 정확도가 88% 수준으로 분석되었으며, 개선방향을 추정하는데 교통량, 사고율, 사고심각도 순으로 높은 관계가 있는 것으로 분석되었다.

핵심어 : 교통안전, 교통시설물, 교통체계개선, 랜덤 포레스트

ABSTRACT

Government agencies, such as police and local governments, strive to prevent traffic hazards and create a comfortable road environment by pormoting transportation and road facilities. To this end, roads and transportation facilities are enhanced and adjusted, and improvement projects in areas with frequent traffic accidents are carried out. Usually, improvement projects in areas with frequent traffic accidents vary by projects and region. Moreover, these projects are carried out under the supervision of a person in charge and related parties. Hence, civil complaints and subjectivity are reflected in deriving priorities for the improvement projects, limiting the efficiency of the project. To this end, a study was conducted to estimate the direction of improvement of the project target site. This study comprehensively considered road, traffic, and accident conditions of representative projects with high effectiveness in

handling traffic accidents. The results of the study state that the accuracy of estimating the improvement project was around 88%. In addition, the study found that there was a strong relationship between traffic volume, accident rate, and accident severity in estimating the improvement direction.

Key words : Traffic Safety, Transportation facilities, TSM, Random Forest

Ⅰ. 서 론

1. 연구 배경 및 목적

정부는 교통사고 사망자 감소를 위해 교통안전 종합대책, 안전속도 5030 등 다양한 정책을 추진하고 있으며, 특히, 1988년부터 지금까지 매년 교통사고 잦은 곳을 대상으로 교통사고 잦은 곳 개선사업을 시행하며 도로상 교통사고를 유발하는 시설을 제거하고 도로 기하구조와 안전시설 등 도로교통환경을 개선하여 안전하고 편안한 교통환경을 조성하는데 노력하고 있다.

교통사고 잦은 곳은 교통사고 건수를 기준으로 선정하는데, 특별·광역시는 7건/년, 일반시 5건, 기타지역 3건을 기준으로 하고 있다. 개선사업의 우선순위는 교통사고 잦은 곳 중에서 도로 기하구조 및 안전시설 측면에서 문제점이 부각되어 개선시 뚜렷한 사고감소 효과가 기대되는 지점을 우선적으로 선정한다.

다만, Kim(2017)은 사고감소 효과가 기대되는 지점을 판단하기 어렵고, 사업 개선방향을 담당자와 관계자의 주관성이 짙어 사업의 우선순위를 초기에 판단하기 어렵다는 문제가 있다. 이에 따라 지역주민의 민원 등이 작용하여 사업 우선순위가 정량적인 기준보다 정성적인 판단으로 변경되기도 하며, 주관적인 개선안의 적용으로 지역별 개선방향 및 사업의 효과가 서로 상이하다는 문제가 발생하고 있다.

교통사고 잦은 곳의 개선방향은 대상지의 교통여건, 사고발생원인 등을 종합적으로 고려하여 판단하지만, Data로 구축된 교통여건과 도로여건, 사고 발생 유형을 사전에 분석하여 사업시행 전 개략적인 사업범위와 방향을 도출하고 명확한 기준에 의한 사업 대상지의 우선순위를 도출하면 개선사업의 업무효율성 및 효과가 더욱 향상될 것으로 판단된다.

따라서, 본 연구는 과거 7년(2012년 ~ 2018년)의 사고 잦은 곳 개선사업의 시행 결과를 바탕으로 교통여건, 도로여건, 사고 발생유형을 분석하여 특정 대상지의 최적 도로 및 교통시설 개선(안)을 추정하는 연구를 수행하였다. 개선항목을 사전에 예측하여 개선방향 설정 및 교통 관련 정책에 활용됨을 목표로 한다.

2. 연구수행 절차

본 연구는 도로 및 교통시설의 개선방향을 예측하기 위해 도로 및 교통 관련 시설의 범위를 사전에 검토하고, 교통사고 잦은 곳 개선사업의 현황과 시행결과를 검토하여 사업시행 항목 및 시행범위를 도출하여 종속변수를 선정하였다. 이후 교통사고 잦은 곳 개선사업에 관련된 선행연구를 검토하여 연구의 차별성을 도출하고, 다양한 변수를 활용하여 변수간 상관도가 낮아도 종속변수의 예측에 적용할 수 있는 의사결정 모형인 랜덤 포레스트에 관하여 이론적 고찰을 수행하여 연구 방법론을 도출하였다.

또한, 본 연구는 사전에 사업의 방향을 설정하는 것을 목표로 하기 때문에 현장조사를 최소화하도록 사전에 조사되었거나 시스템으로 데이터를 활용할 수 있도록 데이터 항목을 선정하였다.

연구방법론을 적용하여 도로여건, 교통여건, 사고발생 유형을 활용한 도로 및 교통시설 개선방향을 도출하는 연구를 수행하였다.

<Fig. 1> Research Procedure

Ⅱ. 현황 및 이론적 고찰

1. 도로 및 교통안전시설 설치 관련 현황

1) 교통안전시설 설치 및 관리 기준

교통안전시설은 도로교통법과 동법 시행규칙에 따라 종류 및 설치 · 관리 기준이 정의되어 있다. 안전표지(주의표지, 규제표지, 지시표지, 보조표지)와 노면표시, 신호기로 구성된 교통안전시설의 구체적인 표시방식과 색상 등을 규정하고 설치 필요지점(구간)을 명시하고 있다. 도로교통법에서는 무인 교통단속용 장비 설치기준을 명시하고 있으며, 동법 시행규칙 [별표 8의2], 무인 교통단속용 장비 설치기준을 제시하고 있다. 교통사고 위험지수(ARI), 사고유형 · 원인, 도로 및 교통조건 등을 종합적으로 고려하여 설치장소를 선정한다.

2) 도로안전시설 설치 및 관리 기준

도로안전시설은 시선유도시설(갈매기표지, 시선유도봉 등), 조명시설, 차량방호 안전시설(방호울타리, 충격흡수시설 등), 기타안전시설(미끄럼방지포장, 과속방지턱 등)로 구성되며, 도로의 안전과 소통을 확보하고 교통사고 및 사고심각도를 감소시키기 위해 설치한다. 도로안전시설은 도로안전시설 설치 및 관리지침에 따라 설치장소, 구조 및 성능, 시공방법 및 유지관리 등이 규정되어 있다.

3) 교통사고 잦은 곳 개선사업

교통사고 잦은 곳 개선사업은 도로에서 일정기준 이상의 교통사고가 발생하는 지점을 교통사고 잦은 곳으로 선정하고, 사고요인 분석과 현장조사를 통해 개선대책을 수립하는 저비용 고효율 교통안전개선 사업으로 1988년부터 지금까지 매년 사업을 추진하여 교통사고 발생건수와 교통사고 인명피해를 줄이는데 커다란 효과를 보이고 있다. 교통사고 잦은 곳 개선사업은 ① 교통안전시설물 조정 및 신설, ② 도로부대시설물 설치 및 개선, ③ 교통운영체계 개선, ④ 도로구조 개선, ⑤ 교차로 복합개선 5가지 항목으로 나눌 수 있다.

행정안전부가 주관하여 도로교통공단이 시행하고 있으며, 지자체 및 경찰서와 같이 추진하고 있다. 사고요인 및 대상지의 복합적인 검토로 개선을 수행하고 있지만, 개선사업 담당자와 관련시설 설치 담당자의 경험과 노하우를 중심으로 추진하고 있는 실정으로 지역간 개선효과에 편차가 발생하고 있는 실정이다.

2. 교통사고 잦은 곳 개선사업 관련 연구 고찰

Yoon.(2017)은 기존 국내외 사례를 참고하거나 설계 담당자의 주관적인 판단과 경험에 의해 시설물을 설치하는 방식에서 본 연구결과를 교차로에서 교통안전사업의 객관적인 근거 확보와 향후 사업추진방향에 참

고할 수 있도록 개별 도로교통안전시설물의 교통사고 감소효과를 추정하는 연구를 수행하였다. 교통사고 잦은 곳 개선사업 4,172개 지점을 대상으로 비교그룹방법을 활용하여 5가지 시설물(교통섬, 표지명, 과속단속카메라, 무단횡단금지시설, 미끄럼방지포장)의 교통사고 감소효과를 분석하였다. 분석결과 교통섬 4.45%, 표지병 32.17%, 과속단속카메라 24.13% 사고감소효과가 분석되었으며, 무단횡단방지시설과 미끄럼방지포장은 각각 0.61%, 1.67% 사고가 증가한 것으로 분석되었다.

Kim(2017)은 교통사고 잦은 곳 개선사업이 지자체 인구규모, 재정상황 등을 반영하지 않고 사고 유형에 따른 획일적인 개선 공사방식으로 개선사업의 효과가 감소되고 있다고 하였다. 사업의 효율성을 높이기 위해 기 수행된 높은 효율성을 가진 지자체가 수행한 개선사업의 유형 및 특성을 분석하여 효율성이 낮은 지자체에 적용하여 최대의 요율을 이끌어내도록 유도할 필요가 있으며, 이를 위해 지자체별 개선사업에 진행된 개선공사 방식을 세분화하여 DEA(Data Envelopment Analysis)모형을 활용하여 사업의 효율성을 추정하였다. 분석결과로 인구수 30만명 이상의 지자체의 경우 토목과 교통공사 지점수를 16%, 62% 정도로 실시하고, 30만명 미만 지자체는 32%, 50% 정도로 실시하여야 한다고 제시하였다.

Jeong(2009)은 교통사고 잦은 곳 선정에 있어 단순 사고건수와 심각도를 고려하여 선정하는 방식보다 사고예측 정보를 이용하여 교통사고 잦은 곳 개선사업 대상지를 선정하는 방법을 제시하였다. 잠재적 위험요소들을 고려하여 개선사업 우선순위를 결정하는 것이 바람직하다는 의견으로 EB(Emperical Bayes Method) 기법을 활용하여 사고예측모델을 개발하였다. EB 기법이 비선형 회귀모형보다 예측력이 좋은 것으로 분석되었고, 분석 대상지 중 몇 개 지점의 개선 우선순위에 변동이 발생하는 것으로 분석되었다. 연구결과를 활용하여 사업 개선 지점 선정에 비용적인 측면과 사고감소 측면으로 기존의 문제점을 극복할 것으로 바라보았다.

3. 연구분석 이론 고찰

1) 분류 및 판별모형

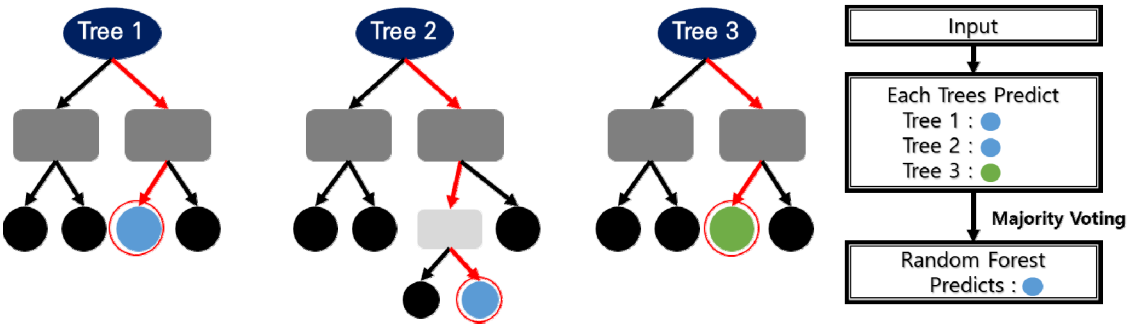
분류 및 판별모형에 다양한 기법들이 적용되고 있다. K-NN(K-Nearest Neighborhood), 선형모델(Logistic Regression, Support Vector Machine), 나이브베이즈, 랜덤포레스트 등이 대표적인 기법으로 각각의 특징과 장단점을 가지고 있다. K-NN은 이해하기 쉬운 장점이 있지만 특성이 다양할수록 예측능력이 저하되어 활용성이 낮은 단점이 있으며, 선형모델은 수식을 통해 예측 결과에 대한 이해가 쉽지만 기본적으로 이진분류만을 처리하는 단점이 있다. 나이브베이즈 기법은 훈련과 예측이 빠른 반면 일반적인 성능이 다른 모델에 비해 낮다는 단점을 가진다. 랜덤포레스트는 예측능력이 우수하고 변수타입에 자유롭다는 장점을 가지며, 긴 학습시간과 데이터가 적을 경우 정확도가 떨어지는 단점을 가진다.(dohk DOKH, 2021)

본 연구에서는 종속변수가 여러 개인 다중분류를 수행하며, 변수별 특성 연속형과 명목이 포함되어 다양한 변수의 특성을 가지기 때문에 이를 만족할 수 있는 랜덤포레스트를 분석기법으로 선정하였다.

2) 랜덤 포레스트(Random Forest)

랜덤 포레스트(Random Forest)는 의사결정나무(Decision tree) 분석 중 CART 알고리즘과 배깅(Bagging) 알고리즘을 적용한 알고리즘으로 CART 알고리즘을 기반으로 하고 있어 분포에 대한 가정이 없고 목표변수와 입력변수 타입에도 자유로워 제약조건이 거의 없다. 또한, 의사결정나무의 약점인 과대적합(Over-fitting) 문제를 해결하고 앙상블 모형의 장점인 예측 정확도를 높인 알고리즘이다.

다만, 모형 산출의 근거를 제시하지 못하는 데이터마이닝 기법의 일반적인 단점과 데이터셋에 레코드와 변수가 많지 않다면 선택 레코드와 변수가 중복되어 모형 적합도가 높지 않을 수 있다



<Fig. 2> Random Forest Algorithm Concept Diagram

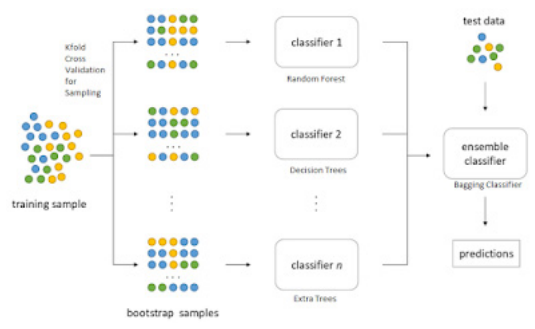
CART 알고리즘은 의사결정 나무분석 중 하나로 의사결정 규칙을 나무구조로 나타내어 전체 데이터를 분류하여 예측하는 기법이며 불순도를 계산하여 불순도를 낮추는 방향으로 예측 및 결정하는 기법이다. 불순도를 낮추는 과정인 이진분리(Binary split)를 수행하는 방법은 지니계수(Gini Index)와 분산의 감소량을 사용하며, 지니계수는 목표변수가 범주형일 때 사용하며, 분산의 감소량은 목표변수가 연속형일 때 사용한다. 지니계수와 분산을 계산하는 수식은 아래와 같다.

$$Gini\ Index = G(S) = 1 - \sum_{i=1}^{C} p_i^2,\ S = \text{이미 발생한 사건의 모음}, C = \text{사건의 갯수} \quad \cdots\cdots (1)$$

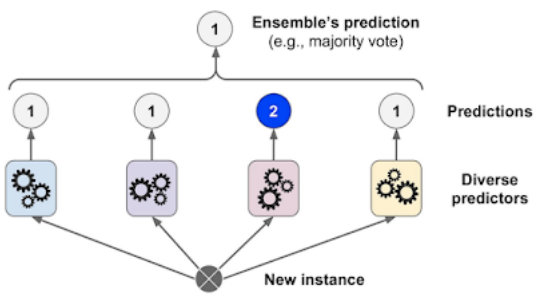
$$Variance = \sum (X - \mu)^2 / n \quad \cdots\cdots (2)$$

배깅 알고리즘은 앙상블 모형의 한 종류로 앙상블 모형은 여러 가지의 모형을 만든 후 각각의 모형별로 예측결과를 산출한 뒤 가장 좋은 예측결과를 선택하는 모형이다. 배깅 알고리즘의 주요 특징은 부트스트랩(Bootstrap)과 어그리게이팅(Aggregating)의 과정을 거쳐 예측모형을 생성하는 것이다. 부트스트랩은 하나의 데이터에서 여러 개의 데이터셋을 추출하는 것을 말하며, 어그리게이팅은 부트스트랩으로 K개의 데이터셋이 만들어졌을 때 분류자(Classifier)를 이용하여 결과를 결합하는 것을 말한다.

랜덤포레스트는 배깅 방법을 이용하여 K개의 샘플 데이터셋을 구축하고, CART 알고리즘을 분류자로 하여 모형을 생성한 뒤 검증용 데이터에 모형을 적용하여 결과를 예측하는 방법이다.



<Fig. 3> Bagging Classifier Process Flow



<Fig. 4> Prediction Methods of Ensemble Models

Ⅲ. 자료수집 및 모형 개발

1. 자료 수집

연구에 활용된 자료는 2012년부터 7년간(2012 ~ 2018)의 전국 교통사고 잦은 곳 개선사업 개선내용의 자료를 활용하였다. 도로교통공단에서 관리하고 있는 교통사고 잦은 곳 개선사업에 대한 개선항목, 대상지의 도로종류, 교차로 형태, 사고 건수, 사고심각도, 교통량 등의 데이터를 활용하였다.

교통사고 잦은 곳 개선사업은 대상지 현황과 사고원인 등을 종합적으로 분석하여 개선안을 마련하지만, 교통사고 잦은 곳 선정에 있어 단순 사고건수 만을 기준으로 하고 있어 지점별 사고원인을 파악하기 어려운 한계가 있다. 또한, 대상지 현황에 대한 구체적인 기하구조의 개별적인 확인이 어렵기 때문에 대상지 교통현황을 도로종류(도로법에 따른 도로 분류), 도로유형(삼거리, 사거리, 단일로 등), 교통량으로 분류할 수 있도록 하였다. 사고건수, 인명피해(사망 · 부상자수), EPDO, 사고종류(차대사람, 차대차, 차량단독)를 사고 변수로 설정하여 사고에 대한 종합적인 분류를 수행하도록 하였다. 위처럼 가능한 많은 변수를 적용한 이유는 랜덤 포레스트 기법에서는 변수 항목이 많으면 많을수록 부트스트랩에 따른 다양한 데이터셋이 생성되기 때문에 가능한 많은 변수를 활용하고자 하였다.

또한, 분석에 활용한 변수는 유관 시스템에서 관리하고 있는 정보로 교통량을 제외한 변수들은 별도의 현장조사 없이 수집 및 활용한 데이터로 확인되었다. 교통특성 및 도로특성 변수는 국토교통부에서 운영하는 국가교통정보센터의 표준노드링크에서 수집이 가능하며, 사고요인은 도로교통공단에서 운영하는 교통사고 분석시스템(TAAS)를 통해 수집이 가능하다.

분석에 사용된 사고요인 중 사고건수와 사고율, EPDO는 아래의 수식에 의해 산출되었다.

$$\textit{사고건수} = G(S) = \sum(\text{3년 사고건수}) \div 3 \quad \cdots\cdots (3)$$

$$\textit{사고율(차량 100만대당 사고건수)} = \textit{사고건수} \div \textit{1,000,000} \quad \cdots\cdots (4)$$

$$\textit{EPDO(대물피해환산법)} = [(\textit{사망사고 건수} \times 12) + (\textit{부상사고 건수} \times 3) + (\textit{물적피해 건수} \times 1)] \quad \cdots\cdots (5)$$

<Table 1> Data used for Analysis

Traffic Characteristics	Road Characteristics	Accident Characteristics
Traffic Volume	Road type (Local roads, City roads, etc.)	The number of Accidents
		The number of casualties
Region	Road Shape (3-way, 4-way, Single way, etc.)	EPDO
		Accident type (Car to person, Car to car, Vehicle only)

총 데이터는 2083개의 대상지를 대상으로 11개의 독립변수와 1개의 종속변수를 활용하였으며, 독립변수의 기술통계는 아래 표와 같다.

<Table 2> Characteristics of variables used and descriptive statistics

Classification	Road Type	Road Shape	volume	Number_ Accident	Accident rate	Number_ Casualty	EPDO	Accident Type		
								Car-human	Car-car	Car only
Mean	Categorical	Categorical	45,151	9.85	0.90	16.30	33.17	1.41	8.15	0.26
min			286	0	0	0	0	0	0	0
Max			229,532	135.30	46.93	210.70	429.90	21.30	116.0	4.30
StDev			35,773	9.52	2.0	15.34	30.47	2.13	7.91	0.46

2. 랜덤 포레스트 모형 개발

렌덤 포레스트를 이용하여 모형을 개발하기 앞서 예측의 결과로 도출되는 종속변수에 대한 검토를 선행하였다. 교통사고 잦은 곳 개선사업은 앞서 언급한 듯이 5개의 항목으로 구분되어 있으며, 각 항목별로 시설물의 종류에 따라 세부 항목으로 구분되어 사업이 진행되고 있다. 5개 개선항목 중 교통운영체계 개선(신호운영 체계 개선)과 교차로 복합분석을 제외한 나머지 3개 항목(교통안전시설 개선, 도로부대시설 개선, 도로구조 개선)을 종속변수로 선정하였다.

신호운영체계는 교차로의 이동류와 신호운영 효율성을 변수로 고려해야 할 필요성이 있지만, 본 연구에서는 교차로 신호운영에 관한 변수를 사용하고 있지 않아 제외하였으며, 교차로 복합개선은 앞서 4개의 항목을 해당 대상지에 맞춰 복합적으로 적용하였기 때문에 분류의 결과에 오차를 유발할 것으로 판단하였다.

종속변수가 범주형이기 때문에 범주형 결과를 반환할 수 있도록 Classification Tree를 적용하였으며, 부트스트랩의 개수(트리의 개수)를 100개, 200개, 300개, 500개 총 4번을 수행하여 각각의 예측값을 비교하여 최적 모형을 도출하였다.

Ⅳ. 도로 및 교통시설 개선항목 예측

부트스트랩 및 변수 샘플 데이터셋을 100개, 200개, 300개, 500개로 증가시켜 분류하여 예측 정확도 분석으로 모형의 성능을 판단하였다. 그 결과 100개의 샘플데이터 셋을 적용한 모형이 88.8%로 가장 높은 예측 정확도를 보였다.

<Table 3> Evaluation Model performance according to bootstrap

Comparison of Prediction results	Number of bootstrap and sample datasets			
	100	200	300	500
Overall error rate	0.112	0.115	0.115	0.113
Prediction Accuracy	88.8%	88.5%	88.5%	88.7%

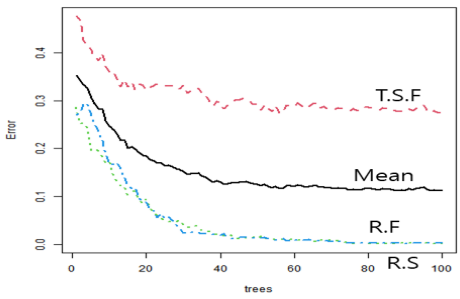
정확도가 가장 높은 100개의 샘플데이터를 적용한 예측값에서 예측의 결과(종속변수) 각각의 정확도는 도로부대시설 개선 > 도로구조 개선 > 교통안전시설 개선 순으로 분석되었다. 교통안전시설의 경우 종류가 다

양하고 직접적인 교통사고 개선 효과가 도로부대시설 및 도로구조 개선에 미치지 않기 때문에 사고유형과 사고영향 요인들이 변수로 들어간 모형의 예측 정확도가 다른 개선항목보다 낮은 것으로 판단된다.

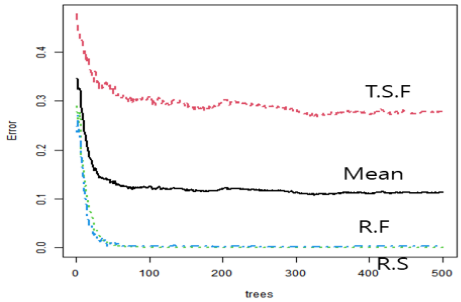
또한, 도로부대시설, 도로구조개선의 경우 사고요인의 영향이 크게 작용하고, 각각의 특성이 명확하게 구분되어 예측의 결과가 높게 도출된 것으로 판단된다.

<Table 4> Evaluation Model performance according to bootstrap

Improve	Traffic safety Facility Improvement	RoadFacility Improvement	Road Structure Improvement	Prediction Accuracy
Traffic safety Facility	609	97	133	72.6%
Road Facility Improvement	1	571	0	99.8%
Road Structure Improvement	0	2	670	99.7%



<Fig. 5> Error rate Graph(100 Bootstrap)



<Fig. 6> Error rate Graph(500 Bootstrap)

<Fig. 5>와 <Fig. 6>은 부트스트랩의 증가에 따른 오차율의 변화를 나타낸 그래프이다. 부트스트랩 및 변수 샘플 데이터셋의 개수가 적을수록 오차율이 높지만, 80개 이상부터는 일정한 오차율로 안정화되는 것을 볼 수 있다.

랜덤 포레스트를 적용하여 노드 불순도 감소량에 따른 변수의 중요도를 분석하였다. 노드 불순도는 트리의 노드가 분류될 때 지니계수(GINI Index)를 얼마만큼 감소 시킬 수 있느냐를 판단하여 노드 불순도가 큰 변수일수록 노드를 잘 분류하였다고 볼 수 있다. 또한, 얼마만큼 예측에 중요한 변수인지를 판단할 수 있는 지표로 활용된다.

<Table 5> Node Impurity Reduction

Boot strap	volume	Accident rate	Number_ Casualty	EPDO	Car-car	Number_ Accident	Car-human	Car only	Road Type	Road Shape
100	222.06	190.68	190.94	174.75	140.35	135.72	114.60	70.06	64.56	57.93
200	221.05	196.75	193.58	168.48	146.98	137.60	108.50	68.04	62.44	59.15
300	223.48	192.25	189.92	167.84	145.91	140.58	113.31	69.62	62.17	58.01
500	221.47	197.65	191.34	169.56	143.60	138.16	112.36	68.20	62.41	58.25

노드 불순도 감소량으로 판단한 개선항목 예측에 사용되는 변수의 중요성은 교통량이 가장 중요한 변수로 분석되었으며, 다음으로 사고율, 사상자수, EPDO, 차대차 사고건수, 교통사고 건수, 차대사람 사고건수, 차량단독 사고건수, 도로유형, 도로형태 순으로 분석되었다.

Ⅴ. 결론 및 향후과제

본 연구는 도로 및 교통시설의 개선방향을 사전에 판단하여 사업의 우선순위 및 교통안전 계획에 반영하기 위해 의사결정 기법인 랜덤 포레스트 기법을 활용하여 개선방향을 추정하는 연구를 수행하였다.

교통사고 잦은 곳 개선사업은 개선방안을 마련하는 담당자 및 관계자의 주관성이 반영되어 사업의 효과에서 지역적 편차가 나타나고 있는 실정으로 데이터 기반 분석체계를 활용하여 지역별 편차를 감소시켜 사업의 효과를 높이고자 하였다. 교통사고 잦은 곳 개선사업의 시행 결과 데이터를 바탕으로 다양한 개선방안을 5개의 항목으로 취합하고, 연구의 목적에 맞게 3개의 개선항목(교통안전시설 개선, 도로부대시설 개선, 도로구조 개선) 축약하여 도로요인, 교통요인, 사고요인의 변수를 활용하여 개선항목을 추정하는 연구를 수행하였다. 연구 결과 최적의 모형 도출을 위해 부트스트랩 및 변수 샘플 데이터셋을 100개, 200개, 300개, 500개로 구분하여 분석을 수행하였고, 예측 정확도는 88.8%로 분석되었다. 랩덤 포레스트가 80개의 부트스트랩 및 데이터 셋 이상이면 예측결과는 안정화되어 최소 80개 이상으로 설정하면 안정적인 예측결과를 가져올 수 있을 것으로 분석되었다.

도로 및 교통시설물은 해당 지점 및 구간의 다양하고 복합적인 요소로 개선 및 조정을 수행하고 있지만, 본 연구를 통해 데이터가 충분히 갖춰진다면 사전에 데이터를 기반으로 사업의 범위 및 방향을 예측하여 정책을 추진할 수 있음을 확인할 수 있다. 본 연구결과를 바탕으로 사고 잦은 곳 개선사업 및 교통체계 개선사업에서 연간 가용 가능한 예산을 종합적으로 검토하여 사업 대상지 우선순위 선정에 활용할 수 있으며 개략적인 사업 범위를 사전에 예측하여 관리 및 사업추진의 효율성을 높일 수 있을 것이다.

본 연구는 도로부대시설 개선, 도로구조 개선, 교통안전시설 개선 3개 항목을 예측하였지만, 신호운영이 포함된 교통체계 개선항목 예측에 한계가 있다. 교통체계 개선을 위해서는 신호시간에 따른 교차로 처리 용량과 점유율 등 신호운영 효율성에 관련된 변수와 방향별 교통량, 방향별 차로수 등 신호운영과 연관된 변수들이 조사 및 자료수집에 한계가 있어 예측 범위에 제외된 한계점과 제약사항이 존재한다.

또한, 현장 적용 및 구체적인 개선항목 제시를 위해서는 본 연구의 종속변수를 다양하게 정의하고 예측할 필요가 있으며, 이를 위해 더 많은 지점의 데이터를 수집하며, 신호운영 등 교통체계의 복합성을 반영하기 위해 더욱 다양하고 풍부한 변수를 적용할 필요가 있다. 마지막으로 향후 연구로 다른 분석기법를 활용한 교통시설 개선방안 추정연구를 수행하여 비교분석 결과로 최적의 분석모형 분석이 필요할 것으로 판단된다.

ACKNOWLEDGEMENTS

이 논문은 2021년도 정부(경찰청)의 재원으로 과학치안진흥센터의 지원을 받아 수행하였습니다.
(No.092021C28S01000, 자율주행 혼재시 도로교통 통합관제시스템 및 운영기술 개발)

REFERENCES

DOHK, *Comparing the pros and cons of machine learning algorithms*, https://dohk.tistory.com/170, 2021.11.08.

Jeong S. B.(2009), "Development of Evaluation Model for Black Spot Improvement Priorities by using Emperical Bayes Method," *Journal of Korea Society of Transportation*, vol. 27, no. 3, pp.81-90.

Jo Y. J.(2018), *Big Data SPSS Latest Analysis Techniques*, Seoul: Hannarae, pp.124-146.

Kim H. K.(2017), "A Study on Measuring Efficiency Improvement of Improvement Project at Black Spot by DEA," *Journal of the Korean Society of Safety*, vol. 32, no. 5, pp.141-148.

Korea Ministry of Government Legislation, www.law.go.kr, 2021.11.02.

Ministry of Land, Infrastructure and Transport(2002), *Accident-prone area improvement project work manual*.

Superzzangzzang, https://leedakyeong.tistory.com/entry/%EC%9D%98%EC%82%AC%EA%B2%B0%EC%A0%95%EB%82%98%EB%AC%B4Decision-Tree-%EB%8F%85%EB%A6%BD%EB%B3%80%EC%88%98%EA%B0%80-%EC%97%B0%EC%86%8D%ED%98%95-%EC%9D%BC-%EB%95%8C, 2021.11.02.

Yoon Y. I.(2017), "Estimating Traffic Accident Reduction Effect of Road Safety Facilities in Intersections," *Journal of Korean Society of Transportation*, vol. 35, no. 2, pp.129-141.