

XGBoost를 활용한 시설물의 부재 상태 예측

장태연¹ · 윤시후² · 지석호^{3*} · 임석빈⁴

¹서울대학교 건설환경공학부 석박사통합과정 · ²서울대학교 건설환경공학부 석박사통합과정 · ³서울대학교 건설환경공학부 교수

⁴국토안전관리원 안전성능연구소 수석연구원

Condition Estimation of Facility Elements Using XGBoost

Chang, Taeyeon¹, Yoon, Sihoo², Chi, Seokho^{3*}, Im, Seokbeen⁴

¹Graduate Student, Department of Civil and Environmental Engineering, Seoul National University

²Graduate Student, Department of Civil and Environmental Engineering, Seoul National University

³Professor, Department of Civil and Environmental Engineering, Seoul National University

⁴Principal Research Manager, Research Institute for Safety Performance, Korea Authority of Land & Infrastructure Safety (KALIS)

Abstract : To reduce facility management costs and safety concerns due to aging of facilities, it is important to estimate the future facilities' condition based on facility management data and utilize predictive information for management decision making. To this end, this study proposed a methodology to estimate facility elements' condition using XGBoost. To validate the proposed methodology, this study constructed sample data for road bridges and developed a model to estimate condition grades of major elements expected in the next inspection. As a result, the developed model showed satisfactory performance in estimating the condition grades of deck, girder, and abutment/pier (average F1 score 0.869). In addition, a testbed was established that provides data management function and element condition estimation function to demonstrate the practical applicability of the proposed methodology. It was confirmed that the facility management data and predictive information in this study could help managers in making facility management decisions.

Keywords : Facility Management, FMS, Element Condition Estimation, XGBoost

1. 서론

공용년수가 30년 이상인 고령 시설물의 비중이 10년 후 43.3%로 급증한다는 통계에서 보듯이 고령화로 인한 시설물의 노후화 및 안전문제는 피할 수 없는 현실이다(KALIS, 2021). 현행 시설물 유지관리 방식은 점검진단을 통해 발견한 손상에 대해 유지보수를 진행하기 때문에 급증하는 노후 시설물을 점검·보수하는 데에 많은 비용을 요구한다. 또한, 적기에 발견하지 못한 손상은 시설물 대형사고의 원인이 되기 때문에 안전에 대한 사회적 불안감을 초래한다. 따라서

노후 시설물의 유지관리 비용부담을 줄이고 안전성을 확보하기 위해서는 유지관리 데이터를 활용하여 시설물의 현저한 손상이 나타나기 이전에 상태를 예측하고, 이를 유지관리 의사결정에 활용하는 것이 중요하다.

국토안전관리원은 시설물통합정보시스템(Facility Management System, 이하 FMS)을 운영하여 시설물 관리대장, 설계도서, e-보고서 등 유지관리 데이터를 관리하고 있다(MOLIT, 2021a). 구체적으로 시설물 관리대장은 FMS에 등록된 시설물의 기본제원(예. 종구분, 준공일, 연장)을 축적·관리하며, 설계도서는 구조정보(예. 부재별 재료, 강도, 수량)를 포함하고, e-보고서는 정밀안전점검 및 정밀안전진단 실시결과(예. 부재 및 시설물 전체 상태평가)를 포함한다. 그러나 이러한 정보들은 대부분 비정형 형태(예. PDF, TIFF)로 수집되고 있어 정보의 활용이 용이하지 않은 실정이다.

따라서 본 연구는 FMS 데이터로부터 시설물의 기본제원, 구조정보, 점검정보가 포함된 정형데이터를 구축하고, 이를 바탕으로 향후 시설물의 상태를 예측하여 유지관리 의사결정을 지원하고자 한다. 시설물의 안전 및 유지관리 실시 세

* **Corresponding author:** Chi, Seokho, Department of Civil and Environment Engineering, Seoul National University, 1 Gwanak-Ro, Gwanak-Gu, Seoul 08826, Korea

E-mail: shchi@snu.ac.kr

Received September 30, 2022; **revised** October 27, 2022

accepted December 22, 2022

부지침(MOLIT, 2021b)에 따르면 전체 시설물 상태등급은 a~e등급으로 점검된 부재별 상태평가 결과를 종합하여 결정된다. 이에 구축된 정형데이터로부터 시설물 부재의 상태 등급을 예측함으로써 전체 시설물의 상태를 추정할 수 있다.

본 연구에서는 FMS에 등록된 시설물 중 가장 많은 비중(20.7%)을 차지하는 교량시설물을 분석대상으로 선정했다(KALIS, 2021). 다수의 교량시설물 관련 선행연구는 시설물 전체 또는 주요부재의 향후 상태를 예측하는 모델을 개발해 왔으며, 최근에는 전통적인 예측모델의 한계를 극복하기 위해 다양한 기계학습 알고리즘을 활용하고 있다(Alipour et al., 2017; Bektas et al., 2013; Lim & Chi 2019; Martinez et al., 2020; Morcou, 2005). 본 연구는 대표적인 기계학습 알고리즘인 XGBoost (Extreme gradient boost)를 활용하여 시설물의 기본제원과 부재별 구조정보 및 과거 점검정보 등 다양한 변수로부터 향후 부재 상태를 추정하는 방법론을 제안한다. 방법론의 유효성을 검증하기 위해 교량시설물 일부를 대상으로 샘플 데이터를 구축하고, 기존 기계학습 알고리즘과의 성능 비교를 통해 XGBoost의 우수성을 확인하며, 차기 정밀안전진단 및 정밀안전점검 시 부재 상태등급을 예측하는 모델을 개발 및 평가했다. 또한 테스트베드를 구축하여 개발된 예측모델의 활용 가능성을 실증했다.

2. 문헌고찰

교량시설물 전체 또는 주요 부재의 향후 상태를 예측하기 위해 다수의 선행연구는 전통적으로 회귀 또는 마코비언 모형 기반 예측모델을 개발했다(Bolukbasi et al., 2004; Cesare et al., 1993; Kim & Yoon, 2010; Morcou, 2006). 그러나 이러한 연구들은 점검진단 시 부재 상태평가에 중요한 역할을 하는 해당 부재의 과거 상태등급과 구조적 영향변수 등을 동시에 고려하는 데에 한계가 있었다. 이러한 한계를 극복하고자 최근 연구들은 기계학습을 활용함으로써 많은 영향변수들을 동시에 고려하고 다량의 데이터를 학습하여 예측모델의 성능을 높이려고 시도했다.

그 중에서도 Decision Tree, Random Forest, XGBoost와 같은 트리(Tree) 기반 기계학습 알고리즘이 활발히 활용되고 있다. Decision Tree는 특정 기준(Node)에 따라 데이터를 구분하는 가장 기본적인 트리 기반 알고리즘으로써 다수의 연구에서 교량시설물 전체 및 주요부재의 향후 상태를 예측하는 데에 주목할만한 성능을 보였다(Bektas et al., 2013; Martinez et al., 2020; Morcou, 2005). 또한 Random Forest는 여러 번의 샘플링을 통해 Decision Tree를 학습시키고 결과물을 집계하는 알고리즘으로써 빠른 성능과 우수한 성능이 검증되어 시설물 유지관리 분야에도 활용되었다

(Alipour et al., 2017).

Chen and Guestrin (2016)이 소개한 XGBoost는 Classification and regression trees (CARTs)가 결합된 그래디언트 부스팅 방식에 기초한 알고리즘이다. 부스팅은 여러 개의 성능이 낮은 분류기를 결합해 학습과 예측을 반복하면서 오분류된 데이터에 가중치를 부여해 오류를 개선하는 방식이다(Shmueli et al., 2010). 부스팅 방식 중에서도 그래디언트 부스팅은 비용함수의 기울기를 줄이는 방향으로 가중치를 부여하여 성능을 높이는 방식이다(Friedman, 2001). 특히 XGBoost는 오버피팅의 위험을 줄이고 자원 효율성을 높임으로써 테이블 형태의 정형데이터를 활용한 예측에 있어 높은 성능과 짧은 학습시간을 보여주고 있다(Bansal, 2018). 따라서 본 연구에서는 우선적으로 기존에 활용되었던 기계학습 알고리즘과의 성능 비교를 통해 XGBoost의 우수한 성능을 확인하고, XGBoost를 활용한 시설물 부재 상태 등급 예측모델을 개발 및 평가하고자 한다.

3. 연구 방법론

3.1 FMS 유지관리 샘플 데이터 구축

FMS로부터 시설물의 부재 상태등급 추정을 위해 필요한 정보(기본, 구조, 점검정보)를 추출하여 샘플 데이터를 구축한다. 구체적으로 본 연구는 시설물 관리대상으로부터 교량 시설물의 기본정보(예. 종구분, 준공연도, 관리기관, 교상)를 추출했으며, 설계도서와 e-보고서를 활용하여 상부구조정보(예. 바닥판두께, 주형간격)와 하부구조정보(예. 교대지지 형식, 교각구체높이)를 추출했다. 또한 e-보고서로부터 1~4년 주기로 실시되는 교량시설물의 정밀안전점검 및 정밀안전진단 결과를 취득하여 점검년도, 공용년수, 부재별 상태등급, 직전 점검진단의 부재별 상태등급 등의 구체적인 점검 정보를 추출했다. 추가적으로 교량시설물의 기본정보 중 하나인 교통량정보를 수집하기 위해 교통량정보제공시스템(MOLIT, 2022)을 활용했다. 시스템에 축적된 교통량 데이터를 조사지점을 기준으로 시설물 관리대상과 연계하여 점검진단 전년도 교통량 및 트럭교통량을 수집·연계했다.

3.2 데이터 전처리

본 연구는 예측모델의 성능 향상을 위해 데이터 정제(Cleaning), 통합(Integration), 축소(Reduction) 단계를 거쳐 구축된 샘플 데이터를 전처리했다(Han et al., 2011). 데이터 정제는 노이즈 값과 결측치를 처리하는 단계이다. 예를 들어 a~e 외 노이즈 값(예. x, q)으로 표기된 상태등급 데이터는 제거했으며, 수치형 변수의 결측치는 중간값으로 대체하고 범주형 변수의 경우 결측치가 있는 데이터는 삭제했다.

직전 점검진단이 없는 경우에는 점검정보에서 직전 점검진단의 상태등급을 'f'로 대체했다. 다음으로 여러 개의 테이블 형태로 구성된 샘플 데이터는 모델에 투입되기 위한 하나의 데이터셋으로 통합되었다.

데이터 축소는 분석결과에 영향을 미치지 않는 변수를 제거하는 단계이다. 예를 들어 본 연구에서 바닥판 종류는 모든 데이터에서 같은 범주(Reinforced concrete)를 가지므로 삭제했다. 추가적으로 변수 간 중복성은 설명변수의 차원(Dimension)을 높이고 분석결과에 혼란을 주기 때문에, 중복변수 중 하나의 변수를 제거하는 것이 필요하다(Huang et al., 2012). 통합된 데이터셋에서 중복변수를 찾기 위해 본 연구에서는 수치형 변수 간 피어슨 상관계수와 범주형 변수 간 크래머 V 계수를 계산했으며, 상관계수가 0.6 이상인 경우 두 변수를 중복변수로 간주했다(Han et al., 2011; Reddy et al., 2013; Wu et al., 2014). 그리고나서 중복변수 가운데 목적변수인 부재 상태등급과 상관성이 낮은 하나의 변수는 제거되었다. 수치형 변수와 부재 상태등급 간 상관성 분석을 위해 상태등급은 5개의 결합점수(0.1, 0.2, 0.4, 0.7, 1.0)로 변환되었다(MOLIT, 2021b).

3.3 부재 상태등급 예측모델 개발

3.3.1 모델 구성

본 연구의 목적변수인 부재 상태등급은 5개의 범주(a~e 등급)로 구분되므로, 동시에 여러 개의 범주를 추정하는 다중클래스 분류분석(Multi-class classification)을 수행해야 한다. 그러나, 범주별 데이터 불균형 문제로 인해 데이터 수가 적은 범주(c~e등급)를 학습하여 분류분석의 성능을 높이기에는 한계가 있다. 따라서 본 연구는 부재 상태등급을 예측하기 위해 여러 개의 독립적인 이중 클래스 분류분석(Binary-class classification)을 수행하는 방법을 제안한다(Fig. 1). 즉, 부재 상태등급 예측모델은 AB그룹(a등급, b등급)

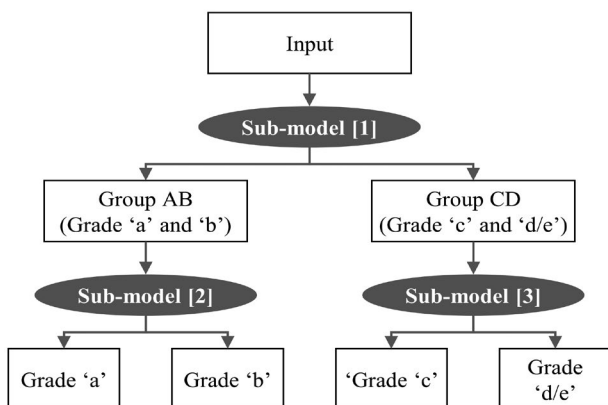


Fig. 1. Model design for element condition estimation

과 CD그룹(c등급, d/e등급)을 예측하는 세부 모델[1], AB그룹에서 a등급과 b등급을 예측하는 세부 모델[2], CD그룹에서 c등급과 d/e등급을 예측하는 세부 모델[3]으로 구성된다. 다만, 상태등급이 e등급인 부재는 매우 소량이므로 d등급으로 포함하여 세부 모델을 구성했다.

3.3.2 데이터 샘플링

전처리 된 데이터는 세부 모델을 학습 및 평가하기 위한 학습 데이터와 평가 데이터로 분리된다. 본 연구의 목적변수인 부재 상태등급은 정상범주(b등급)의 비율이 매우 높기 때문에, 데이터 개수가 많은 다수 범주에 학습 데이터가 편향되는 데이터 불균형 문제가 발생한다. 이를 해결하기 위해 본 연구는 오버샘플링, 언더샘플링과 같은 데이터 샘플링 기법을 활용하여 다수 범주만을 학습하여 나타날 수 있는 예측모델의 성능 저하를 방지하고자 한다.

오버샘플링 기법은 데이터 개수가 적은 소수 범주들에 속한 데이터를 복제하는 기술로서, 학습 데이터 개수가 많지 않을 때 주로 사용된다. 그 중에서도 SMOTE (Synthetic minority over-sampling technique)는 소수 범주에 속한 데이터의 일부를 추출하여 새로운 유사 데이터를 생성하는 기법으로 데이터 복제로 인한 오버피팅을 방지하는 데에 도움을 준다(Chawla et al., 2002). 본 연구는 오분류가 발생할 수 있는 다수 범주와 소수 범주 간 경계에 위치한 데이터를 유사하게 복제하는 borderline-SMOTE를 활용했다. 추가적으로 오버샘플링으로 데이터 불균형 문제가 해소되지 않는 경우 다수 범주의 데이터 일부를 무작위로 제거하는 무작위 언더샘플링 기법도 활용했다. 이러한 과정은 Python 라이브러리인 scikit-learn 0.24.2와 imbalanced-learn 0.8.1을 활용하여 수행했다.

3.3.3 XGBoost를 활용한 모델 학습

샘플링 된 학습 데이터로부터 XGBoost를 활용하여 각 세부 모델을 학습시키기 위해 최적 하이퍼파라미터 조합을 찾는 그리드서치를 수행했다. 본 연구는 <Table 1>과 같은 하이퍼파라미터 유형과 그리드서치 값에 따라 파라미터를 조정했으며, 5겹 교차 검증을 수행하여 평균 정확도가 가장 높은 조합을 최적값으로 선정했다. 정확도(Accuracy)는 검증 데이터 중 예측값과 실제값이 일치하는 데이터의 비율을 의미한다. 최종적으로 서브 모델은 교차 검증을 통해 도출된 XGBoost 최적 하이퍼파라미터 조합을 활용하여 학습한다. 이러한 모델 학습 과정에서는 Python 라이브러리인 scikit-learn 0.24.2와 xgboost 1.5를 활용했다.

Table 1. Hyperparameters for XGBoost

Hyperparameter	Description	Values for grid search
Max_depth	The maximum of number of branches from root node to leaf node	4-10 (interval: 2)
Min_child_weight	The minimum sum of weights of all data required in a child	1, 5, 10
Subsample	The ratio of data for sampling to construct each tree	0.6, 0.8, 1.0
Colsample_bytree	The ratio of columns (variables) for sampling to construct each tree	0.6, 0.8, 1.0
Learning rate	The amount to change weights	0.05-0.1 (interval: 0.01)

3.3.4 모델 평가

학습된 각 서버 모델은 평가 데이터를 활용하여 평가된다. 정확도는 평가 데이터의 불균형 문제로 인해 모델의 예측 성능을 충분히 반영하는 데 한계가 있다. 이에 본 연구는 목적변수의 각 범주별 정밀도(Precision), 재현율(Recall), 서포트(Support)를 고려한 Weighted average F1 score(이하 F1 score)를 활용하여 모델의 성능을 평가했다. 정밀도는 해당 범주라고 예측된 결과 중 실제값과 일치하는 비율, 재현율은 해당 범주에 속한 데이터 중 해당 범주라고 예측된 비율, 서포트는 각 범주에 속한 데이터의 개수를 의미한다.

$$Weighted\ average\ F1\ score = \frac{1}{N} \times \sum_{i=1}^N (S_i \times \frac{2}{\frac{1}{P_i} + \frac{1}{R_i}}) \quad (1)$$

P_i : i 번째 범주의 정밀도
 R_i : i 번째 범주의 재현율
 S_i : i 번째 범주의 서포트
 N : 범주의 개수

4. 연구 결과 및 고찰

본 연구는 제안한 방법론의 유효성을 검증하기 위해 FMS에 등록된 1·2종 도로교량 시설물 중 대표적인 상부 구조형식을 가진 STB (Steel box)교와 PSCI (Pre-stressed concrete I-type)교를 대상으로 샘플 데이터를 구축하고 실험을 진행했다. 주요 구조부재인 바닥판, 주형, 교대/교각의 차기 정밀안전점검 및 정밀안전진단 시 상태등급을 예측하는 모델을 개발하고 평가했으며, 테스트베드를 구축하여 개발된 모델의 활용 가능성을 실증했다.

4.1 샘플 데이터 구축 및 전처리 결과

무작위로 선정된 STB교와 PSCI교 180개소에 대하여 시설물 관리대장, 설계도서, 2015~2021년에 수행된 640건의 e-보고서로부터 15개의 기본정보, 16개의 경간별 상부구조정

보 및 5개의 지점별 하부구조정보, 점검정보(공용년수, 점검년도, 부재별 상태등급, 직진 점검진단의 부재별 상태등급)를 추출하여 유지관리 샘플 데이터를 구축했다. 데이터 전처리를 통해 노이즈 값과 결측치를 처리했으며, 모델 투입을 위해 STB교 및 PSCI교 바닥판, 주형, 교대/교각에 대해 샘플 데이터를 각각 통합했다.

데이터 축소 단계에서는 우선 각각의 샘플 데이터셋에서 같은 범주를 가지는 범주형 변수인 바닥판 종류와 대표상부 구조형식은 삭제했다. 다음으로 상관성 분석을 통해 중복변수를 조사하고 두 변수 가운데 예측모델의 목적변수인 바닥판, 주형, 교대/교각 상태등급과의 상관성이 낮은 변수를 제거했다. 예를 들어 바닥판 상태등급 예측의 경우 '시간장 - 최대경간장(상관계수: 0.91)', '전년도트럭교통량 - 전년도 교통량(상관계수: 0.88)' 등 8개의 수치형 변수 조합과 '노선구분 - 관리기관(상관계수: 1.00)'과 같은 1개의 범주형 변수 조합이 중복변수로 조사되었으며, 두 변수 중 바닥판 상태등급과 상관성이 낮은 최대경간장, 전년도교통량, 노선구분 등 9개의 변수가 제거되었다. 마찬가지로 주형과 교대/교각 상태등급 예측에 대한 중복변수 제거 결과 각각 9개, 7개의 중복변수를 제거했다.

전처리 결과, 바닥판과 주형의 상태등급 예측을 위한 최종 데이터셋은 23개의 설명변수와 1개의 목적변수로 구성되며(Table 2), 교대/교각의 상태등급 예측을 위한 최종 데이터셋은 16개의 설명변수와 1개의 목적변수로 구성되었다(Table 3). 또한, STB교 및 PSCI교 바닥판, 주형, 교대/교각의 상태등급별 데이터 분포는 <Fig. 2>와 같다. 교대/교각의 d/e등급은 데이터 개수가 매우 적어 분석에서 제외했다.

Table 2. Variables of the final dataset for deck and girder condition estimation

(*): Removed variable, (**): Target variable

Concept		Variable
Numerical variable (15)	General information	Construction year(*), Total length, Number of span(*), Maximum span length(*), Total width, Number of lane(*), Average daily traffic(*), Average daily truck traffic
	Structural information by span	Span length, Deck width(*), Deck pavement thickness, Deck thickness, Deck strength, Deck rebar diameter, Deck rebar strength, Girder strength, Girder quantity, Girder spacing(*), Girder height, Crossbeam quantity, Crossbeam spacing
	Inspection information	Bridge age, Inspection year(*)
Categorical variable (9)	General information	Facility class, Route class(*), Region, Management agency, Design live load, Plane shape, Main superstructure type(*)
	Structural information by span	Pavement type, Deck material(*), Girder type
	Inspection information	Condition grade of element(**), Past condition grade of element

Table 3. Variables of the final dataset for abutment/pier condition estimation

(*): Removed variable, (**): Target variable

Concept		Variable
Numerical variable (7)	General information	Construction year(*), Total length, Number of span(*), Maximum span length, Total width, Number of lane(*), Average daily traffic(*), Average daily truck traffic
	Structural information by support	Abutment/pier height, Abutment/pier sphere strength
	Inspection information	Bridge age, Inspection year(*)
Categorical variable (10)	General information	Facility class, Route class(*), Region, Management agency, Design live load, Plane shape, Main superstructure type(*)
	Structural information by support	Abutment support type, Abutment/pier sphere type, Abutment/pier foundation type
	Inspection information	Condition grade of element(**), Past condition grade of element

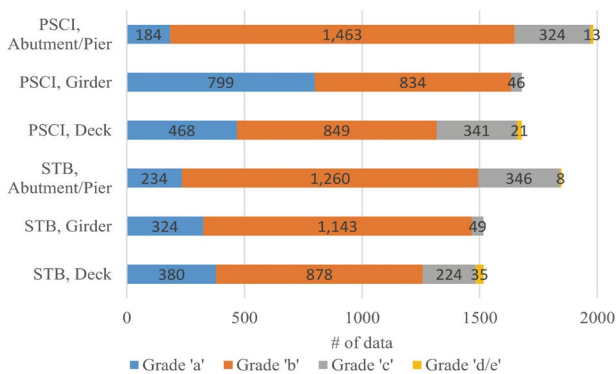


Fig. 2. Data distribution for the condition grades of elements

4.2 부재 상태등급 예측 결과

STB교 및 PSCI교 바닥판, 주형, 교대/교각의 상태등급 예측을 위해 총 14개의 서브 모델을 구성했으며, 각 서브 모델에서는 목적변수의 범주별 데이터 분포에 따라 학습 데이터 (70%)와 평가 데이터(30%)를 구축했다. <Fig. 2>에서 보듯이 부재 상태등급은 각 서브 모델에서 하나의 범주에 매우 치중되어 있기 때문에, 학습 데이터의 불균형 문제를 해소하기 위해 Borderline-SMOTE와 무작위 언더샘플링을 수행했다. 예를 들어 <Table 4>은 STB교 바닥판 상태등급 예측 모델의 서브 모델별 데이터 샘플링 결과를 보여준다.

XGBoost를 활용하여 모델을 학습하기 전, 우선적으로 기존에 활용되었던 트리 기반 기계학습 알고리즘과의 성능 비교를 통해 XGBoost의 우수성을 확인하고자 했다. 대표적으로 <Table 4>의 샘플링 된 데이터를 활용하여 STB교 바닥판 상태등급 예측성능을 비교했다. Decision Tree, Random Forest, XGBoost 각각 최적 하이퍼파라미터를 활용하여 알

Table 4. Data sampling results for deck condition estimation of STB bridge

Sub-model	Class	Total (Original)	# of training data		# of testing data
			Original	Re-sampled	
Sub-model [1]	Group AB	1,423	996	836	427
	Group CD	518	427	418	91
Sub-model [2]	Grade 'a'	410	287	567	123
	Grade 'b'	1,013	709	709	304
Sub-model [3]	Grade 'c'	267	187	144	80
	Grade 'd/e'	35	24	72	11

고리즘을 학습했으며, 검증 데이터를 활용하여 알고리즘의 성능(F1 score)을 비교했다. 그 결과 <Table 5>에서 보듯이 XGBoost이 기존의 트리 기반 기계학습 알고리즘과 비교하여 우수한 성능을 보임을 확인했다.

Table 5. Results of the algorithm performance comparison for deck condition estimation of STB bridge

Sub-model	Algorithm Performance (Weighted average F1 score)		
	Decision Tree	Random Forest	XGBoost
Sub-model [1]	0.842	0.879	0.903
Sub-model [2]	0.787	0.786	0.796
Sub-model [3]	0.944	0.944	0.962
Average	0.858	0.869	0.887

다음으로 14개의 서브 모델별로 샘플링 된 학습 데이터를 활용하여 최적 XGBoost 하이퍼파라미터를 찾기 위한 그리드서치를 수행했으며, 5겹 교차 검증(5-fold cross validation)을 통해 정확도가 가장 높은 최적 하이퍼파라미터 조합을 도출했다. 예를 들어 STB교 바닥판 상태등급 예측을 위한 서브 모델 [1]의 최적 하이퍼파라미터 조합은 10 maximum depth, 1.0 minimum child weight, 0.6 subsample, 0.6 colsample by tree, and 0.07 learning rate와 같다.

마지막으로 각 세부 모델별 평가 데이터를 활용하여 XGBoost 최적 하이퍼파라미터로 학습된 모델의 성능을 검증했다. <Fig. 3>과 <Fig. 4>는 각각 STB교와 PSCI교의 세부 모델별 성능 평가 결과를 보여준다.

14개의 세부 모델은 평균 F1 score 0.869의 성능으로 차기 점검진단 시 주요 부재의 상태등급을 예측했다. 이는 유사연구(Lim & Chi, 2019)에서 확인된 XGBoost를 활용한 교량의 부재 손상발생여부 추정 성능(평균 F1 score 0.852)과 비교했을 때도 준수한 성능임을 알 수 있다. 구체적으로 부재별 성능을 비교하면, 평균적으로 바닥판과 주형에 비해 교대/교각의 상태등급을 예측하는 데에 낮은 성능을 보였다(바닥판 평균 F1 score 0.877, 주형 평균 F1 score 0.883, 교대/교

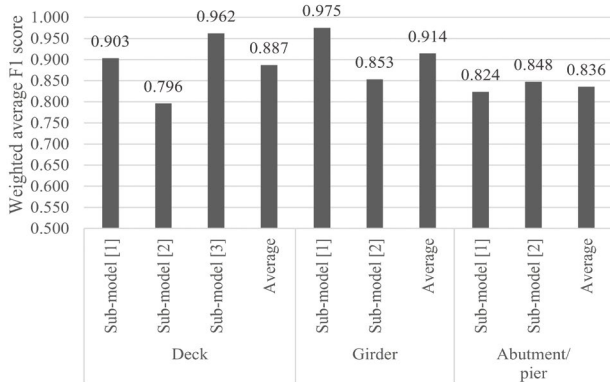


Fig. 3. Model performance by sub-models of STB bridge

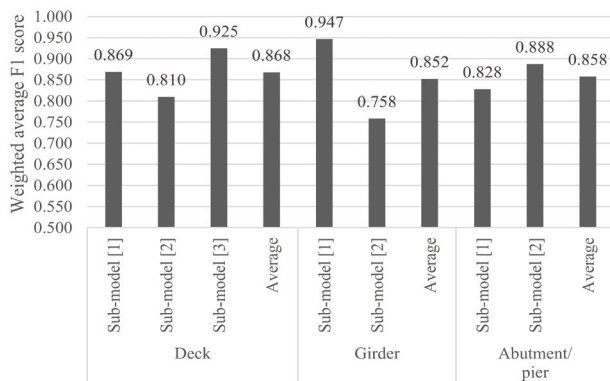


Fig. 4. Model performance by sub-models of PSCI bridge

각 평균 F1 score 0.847). 이는 교대/교각 샘플 데이터의 상태등급 간 데이터 불균형이 상대적으로 컸기 때문이라고 생각된다. 특히 PSCI교 교대/교각의 샘플 데이터는 b등급이 전체 데이터의 73.7%를 차지할 정도로 데이터 불균형 문제가 심각했다. 데이터 불균형 문제가 큰 경우 데이터 샘플링을 통해 소수 범주의 새로운 유사 데이터를 생성하더라도 모델이 소수 범주에 속한 데이터들의 다양한 특성을 학습하기 어려워 예측 성능이 저하될 수 있다.

한편, 서브 모델별로 성능을 비교하면, 평균적으로 서브 모델 [2]의 예측 성능은 서브 모델 [1]과 [3]에 비해 낮았다 (서브 모델 [1] 평균 F1 score 0.891, 서브 모델 [2] 평균 F1 score 0.825, 서브 모델 [3] 평균 F1 score 0.943). 특히 PSCI교 주형의 a등급과 b등급을 예측하는 성능은 F1 score 0.758로 14개의 서브 모델 중에 가장 낮은 성능을 보였다. 실제 점검현장에서는 a등급과 b등급의 외관상 차이는 크지 않아 주로 점검자의 주관에 따라 상태등급이 결정된다고 알려져 있다. 이러한 현상이 반영된 데이터를 학습했기 때문에 서브 모델 [2]의 예측 결과에서도 a등급과 b등급을 쉽게 혼동하는 것으로 판단된다. 반면 바닥판의 서브 모델 [3]에서는 c등급과 d/e등급을 구분하는 설명변수의 차이가 명확했기 때문에 상태등급 예측에 매우 높은 성능을 보였다.

국토안전관리원

시스템소개
데이터관리
모델관리
분석·예측
시스템관리

□ 데이터관리
Home > 데이터관리 > 시설안전데이터

1종
PSCI
노선구분 선택
교량명 입력

검색
엑셀 다운로드

검색결과 : 1항목 (1/1) 페이지 10개씩 보기

번호	시설물번호	시설물명	관리기관	노선	종별	연장(m)	폭(m)	차로수	준공년도
1	BR2010-0000494	강창3교	부산	일반국도	1종	300.00	25.60	5	2010

✓ 기본정보
출처: FMS 시설물관리대장(도로교량)

시설물번호	BR2010-0000494								
시설물명	강창3교	종별	1종	노선	일반국도	시도	대구광역시		
준공년도	2010	관리기관	부산	설계활하중	DB-24	연장(m)	300.00		
경간수	6	최대경간장(m)	50.00	폭(m)	25.60	차로수	5		
상부구조형식	PSCI	평면형상	사각이없는직선교	난간종류	강재(강재파이프)	난간높이(m)	110.00		

✓ 경간정보
경간번호 : 001
출처: FMS 설계도서(도로교량)

경간번호	001	경간장(m)	49.90						
바닥판폭(m)	25.60	바닥판포장유형	LMC	바닥판포장두께(cm)	5.00	바닥판종류	RC		
바닥판두께(cm)	25.00	바닥판강도(kgf/cm ²)	270.00	바닥판주철근직경(mm)	16.00	바닥판주철근강도(kgf/cm ²)	4000.00		
주형구조형식	PSCI	주형강도(kgf/cm ²)	490.00	주형수량	9	주형간격(cm)	290.00		
주형H(cm)	180.00	가로보수량	32	가로보간격(cm)	1200.00				

Fig. 5. Example of data management in testbed



Fig. 6. Example of element condition estimation in testbed

4.3 테스트베드를 활용한 실증

본 연구는 개발된 부재 상태등급 예측모델의 실무적 활용 가능성을 실증하기 위해 테스트베드를 구축했다. Java 기반 웹 어플리케이션 서버를 활용하여 테스트베드를 구축했으며, 샘플 데이터의 탑재를 위한 Maria DB, 예측모델 구동을 위한 Python 등 오픈소스 기반의 소프트웨어를 사용했다. 테스트베드는 FMS 유지관리 데이터 관리 기능과 차기 점검 진단 시 주요부재 상태등급 예측 기능을 제공한다.

FMS 유지관리 데이터 관리 기능을 구현하기 위해 미리 구축된 샘플 데이터를 테스트베드의 데이터베이스에 탑재했다. 사용자가 데이터를 검색하고 확인할 수 있도록 중구분, 대표상부구조형식(STB교, PSCI교) 등에 따른 사용자의 데이터 검색 기능을 구현했다. <Fig. 5>와 같이 사용자 입력에 따라 검색된 시설물 중 하나를 선택하면, 데이터베이스에 저장된 해당 교량시설물의 기본정보, 경간/지점별 구조정보, 점검년도별 점검정보(부재별 상태등급), 교통량정보를 사용자가 직접 확인하고 데이터를 취득할 수 있다.

주요부재 상태등급 예측 기능을 구현하기 위해 JSON 형태로 저장된 14개의 세부 모델이 구동되도록 테스트베드와 연계했다. <Fig. 6>과 같이 사용자가 원하는 교량명과 해당 교량의 전년도트럭교통량, 공용년수, 직전 점검진단 시 경간/지점별 바닥판, 주형, 교대/교각 상태등급을 입력하면, 세부 모델이 구동되어 차기 점검진단 시 경간/지점별 부재 상

태등급 예측 결과를 제공한다. 또한 각 부재에 대한 경간/지점별 예측 결과에 결합도 점수를 부여하여 평균을 계산함으로써(MOLIT, 2021b) 주요부재별 예상등급을 제공한다. 시설물 전체 대표등급은 3개의 주요부재 예상등급의 평균등급으로 제공한다. 종합하면, 본 연구를 통해 구축된 샘플 데이터와 부재 상태등급 예측모델을 활용함으로써 시설물 관리자에게 시설물 유지관리 의사결정에 필요한 시설물 정보 및 시설물 상태 예측정보를 제공할 수 있음을 확인했다.

5. 결론

본 연구는 FMS 데이터를 활용하여 향후 시설물의 상태를 예측하고 유지관리 의사결정을 지원하고자 시설물의 다양한 정보로부터 향후 부재 상태를 추정하는 방법론을 제안했다. 제안한 방법론의 유효성을 검증하기 위해 1·2종 도로 교량시설물을 대상으로 샘플 데이터를 구축하고 XGBoost를 활용하여 부재 상태등급 예측모델을 개발했다. 예측모델의 성능 평가 결과, 차기 주요 부재(바닥판, 주형, 교대/교각) 상태등급을 예측하는 데에 준수한 성능을 보였다(평균 F1 score 0.869). 또한 개발된 예측모델의 실무적 활용 가능성을 실증하기 위해 FMS 유지관리 데이터 관리 기능과 주요부재 상태등급 예측 기능을 제공하는 테스트베드를 구축했다. 이를 통해, 본 연구에서 구축한 샘플 데이터와 예측모델

을 활용하여 시설물 관리자에게 유지관리 의사결정에 필요한 시설물 정보와 시설물 부재별 상태 예측정보를 구체적으로 제공할 수 있음을 확인할 수 있었다. 향후 연구로는, 소수 범주에 속한 c~e등급 데이터를 추가적으로 수집하여 예측모델이 학습하게 함으로써 예측 성능을 높일 수 있으며, 다량의 데이터가 축적된 경우 좋은 성능을 보인다고 알려진 딥러닝 알고리즘을 적용해 볼 수 있다. 또한 과거 보수정보(예, 보수여부, 보수에 따른 성능 향상여부)를 예측모델의 설명변수로 추가함으로써 시설물 보수에 따른 부재의 상태등급 변화를 고려하여 예측모델의 성능과 실무적 설명력을 높일 수 있을 것이다. 아울러 제안된 방법론을 터널, 항만 등 다양한 시설물에 적용하여 상태등급 예측모델을 개발할 수 있다.

감사의 글

본 연구는 국토안전관리원에서 수행하는 기본연구사업(건설시설 안전분야 정보관리체계 개선 및 활용방안에 관한 연구)의 지원으로 수행되었습니다.

References

Alipour, M., Harris D.K., and Barnes L.E. (2017). "Pattern Recognition in the National Bridge Inventory for Automated Screening and the Assessment of Infrastructure." *Proc., Structures Congress 2017*, ASCE, Reston, pp. 279-291.

Bansal, S. (2018). "Data Science Trends on Kaggle." Kaggle, San Francisco, <<https://www.kaggle.com/code/shivamb/data-science-trends-on-kaggle/notebook>> (April 1, 2022).

Bektas, B.A., Carriquiry, A., and Smadi, O. (2013). "Using Classification Trees for Predicting National Bridge Inventory Condition Ratings." *Journal of Infrastructure Systems*, 19(4), pp. 425-433.

Bolukbasi, M., Mohammadi, J., and Arditi, D. (2004). "Estimating the Future Condition of Highway Bridge Components Using National Bridge Inventory Data." *Practice Periodical on Structural Design and Construction*, 9(1), pp. 16-25.

Cesare, B.M.A., Santamarina, C., Members, A., Turkstra, C., and Vanmarcke, E.H. (1993). "Modeling Bridge Deterioration with Markov Chains." *Journal of Transportation Engineering*, 118(6), pp. 820-833.

Chawla, N.V., Bowyer, K.W., Hall, L.O., and Kegelmeyer, W.P. (2002). "SMOTE: Synthetic Minority Over-sampling Technique." *Journal of Artificial Intelligence Research*, 16, pp. 321-357.

Chen, T., and Guestrin, C. (2016). "XGBoost: A Scalable Tree Boosting System." *Proc., the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery, New York, pp. 785-794.

Friedman, J.H. (2001). "Greedy function approximation: a gradient boosting machine." *Annals of statistics*, pp. 1189-1232.

Han, J., Kamber, M., and Pei, J. (2011). *Data mining: concepts and techniques*. Morgan Kaufmann, Waltham.

Huang, J., Huang, N., Zhang, L., and Xu, H. (2012). "A method for feature selection based on the correlation analysis." *Proc., 2012 International Conference on Measurement, Information, and Control*, IEEE, Piscataway, pp. 529-532.

Huang, Y.H. (2010). "Artificial Neural Network Model of Bridge Deterioration." *Journal of Performance of Constructed Facilities*, 24(6), pp. 597-602.

Kim, Y.J., and Yoon, D.K. (2010). "Identifying Critical Sources of Bridge Deterioration in Cold Regions through the Constructed Bridges in North Dakota." *Journal of Bridge Engineering*, 15(5), pp. 542-552.

Korea Authority of Land & Infrastructure Safety (KALIS) (2021). 2021 Statistics on Facilities.

Lim, S., and Chi, S. (2019). "Xgboost application on bridge management systems for proactive damage estimation." *Advanced Engineering Informatics*, 41, 100922.

Martinez, P., Mohamed, E., Mohsen, O., and Mohamed, Y. (2020). "Comparative Study of Data Mining Models for Prediction of Bridge Future Conditions." *Journal of Performance of Constructed Facilities*, 34(1), 04019108.

Ministry of Land, Infrastructure and Transport (MOLIT) (2021a). Facility Management System, <<http://fms.or.kr>> (Oct. 1, 2021).

Ministry of Land, Infrastructure and Transport (MOLIT) (2021b, Dec. 1). Guidelines for Maintenance and Performance Assessments.

Ministry of Land, Infrastructure and Transport (MOLIT) (2022). Traffic Monitoring System, <<http://road.re.kr>> (Feb. 10, 2022).

Morcous, G. (2006). "Performance Prediction of Bridge Deck Systems Using Markov Chains." *Journal of Performance of Constructed Facilities*, 20(2), pp. 146-155.

Reddy, M.C., Balasubramanyam, P., and Subbarayudu, M. (2013). "An Effective Approach to Resolve Multicollinearity in Agriculture Data." *International Journal of Research in Electronics and Computer Engineering*, 1(1), pp. 27-30.

Shmueli, G., Patel, N.R., and Bruce, P.C. (2010). *Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner*,

Wiley, New York.

- Wu, B., Zhang, L., and Zhao, Y. (2014). "Feature selection via Cramer's V-test discretization for remote-sensing image classification." *IEEE Transactions on Geoscience and Remote Sensing*, 52(5), pp. 2593-2606.

요약 : 시설물의 고령화로 인한 유지관리 비용을 줄이고 안전성을 확보하기 위해서는 시설물 유지관리 데이터를 활용하여 향후 시설물의 상태를 예측하고 이를 유지관리 의사결정에 활용하는 것이 중요하다. 이를 위해 본 연구는 XGBoost를 활용하여 다양한 유지관리 정보로부터 향후 시설물의 부재 상태를 추정하는 방법론을 제안함을 목표로 한다. 방법론의 유효성을 검증하기 위해 교량시설물을 대상으로 샘플 데이터를 구축하고, 차기 정밀안전점검 및 정밀안전진단 시 부재 상태등급 예측모델을 개발 및 평가했다. 예측모델의 성능 평가 결과, 주요 부재(바닥판, 주형, 교대/교각) 상태등급을 예측하는 데 준수한 성능을 보였다(평균 F1 score 0.869). 또한 개발된 예측모델의 실무적 활용 가능성을 실증하기 위해 FMS 유지관리 데이터 관리 기능과 주요부재 상태등급 예측 기능을 제공하는 테스트베드를 구축했다. 이를 통해 본 연구에서 구축한 샘플 데이터와 예측모델을 활용하여 시설물 관리자에게 유지관리 의사결정에 필요한 시설물 정보 및 시설물 상태 예측정보를 제공할 수 있음을 확인할 수 있었다. 향후에는 추가적으로 데이터를 수집하고 다량의 데이터가 축적된 경우 좋은 성능을 보인다고 알려진 딥러닝 알고리즘을 활용함으로써 예측 성능을 높일 수 있다. 또한 제안된 방법론을 터널, 항만 등 다양한 시설물에 적용하여 상태등급 예측모델을 개발할 수 있다.

키워드 : 시설물유지관리, FMS, 부재상태예측, XGBoost
