

규칙 기반 분류 기법을 활용한 도로교량 안전등급 추정 모델 개발

Developing an Estimation Model for Safety Rating of Road Bridges Using Rule-based Classification Method

정세환¹⁾, 임소람²⁾, 지석호³⁾
Chung, Sehwan¹⁾ · Lim, Soram²⁾ · Chi, Seokho³⁾

Received May 24, 2016; Received June 20, 2016 / Accepted June 20, 2016

ABSTRACT: Road bridges are deteriorating gradually, and it is forecasted that the number of road bridges aging over 30 years will increase by more than 3 times of the current number. To maintain road bridges in a safe condition, current safety conditions of the bridges must be estimated for repair or reinforcement. However, budget and professional manpower required to perform in-depth inspections of road bridges are limited. This study proposes an estimation model for safety rating of road bridges by analyzing the data from Facility Management System (FMS) and Yearbook of Road Bridges and Tunnel. These data include basic specifications, year of completion, traffic, safety rating, and others. The distribution of safety rating was imbalanced, indicating 91% of road bridges have safety ratings of A or B. To improve classification performance, five safety ratings were integrated into two classes of G (good, A and B) and P (poor ratings under C). This rearrangement was set because facilities with ratings under C are required to be repaired or reinforced to recover their original functionality. 70% of the original data were used as training data, while the other 30% were used for validation. Data of class P in the training data were oversampled by 3 times, and Repeated Incremental Pruning to Produce Error Reduction (RIPPER) algorithm was used to develop the estimation model. The results of estimation model showed overall accuracy of 84.8%, true positive rate of 67.3%, and 29 classification rule. Year of completion was identified as the most critical factor on affecting lower safety ratings of bridges.

KEYWORDS: Road bridges, Safety rating, Rule-based classification, Facility Management System

키워드: 도로교량, 안전등급, 규칙 기반 분류, 시설물정보관리종합시스템

1. 서론

1.1 연구의 배경 및 목적

도로교량은 교통 네트워크에서 핵심적인 연결 기능을 수행하고 있으며 다수의 사용자가 시설물을 이용하기 때문에 교량의 사고 발생 시 사회·경제적인 영향이 매우 크다. 따라서 교량을 안전하게 유지·관리하는 것은 매우 중요하다. 하지만 국내 도로교량은 점차 노후화되고 있으며, 30년 이상 경과된 노후교량이 2015년 기준 3094개에서 2025년에는 9664개로 예상되어 유지·관리 대상이 3배 이상 급증할 것으로 전망된다(한국건설기술연구원, 2015).

교량을 안전하게 유지·관리하는 것은 교량의 손상 극복에 필

요한 보수·보강을 수행함으로써 이루어지며, 최적의 보수, 보강 시점 및 물량을 결정하기 위해서는 우선 교량시설물의 현 상태를 정확하게 진단하는 것이 필요하다. 교량의 안전 상태를 진단하기 위해 “시설물의 안전관리에 관한 특별법”(이하 시특법)에서는 관리 대상 시설물의 범위, 점검 방법 등을 제시하는 등 교량을 포함한 각종 시설물의 안전관리 체계를 구축함으로써 시설물 관리 책임을 규정하고 있다.

시특법에서 “시설물”은 “1종시설물”과 “2종시설물”로 구분된다. 이들은 각 시설물의 중요도에 따라 구분되며 도로교량의 경우 1) 상부구조형식이 현수교, 사장교, 아치교 및 트러스교인 교량 2) 최대 경간장 50미터 이상의 교량(한 경간 교량은 제외) 3) 연장 500미터 이상의 교량 또는 4) 폭 12미터 이상이고 연장

¹⁾학생회원, 서울대학교 건설환경공학부 석사과정 (hwani751@snu.ac.kr)

²⁾학생회원, 서울대학교 건설환경공학부 박사과정 (sorami@snu.ac.kr)

³⁾정회원, 서울대학교 건설환경공학부 부교수, 서울대학교 건설환경종합연구소 겸임교수, 공학박사 (shchi@snu.ac.kr) (교신저자)

500미터 이상인 복개구조물이 1종시설물로 분류되고, 1) 경간장 50미터 이상인 한 경간 교량 2) 1종시설물에 해당하지 않는 교량으로서 연장 100미터 이상의 교량 3) 1종시설물에 해당하지 않는 복개구조물로서 폭 6미터 이상이고 연장 100미터 이상인 복개구조물이 2종시설물로 분류된다. 1종·2종 시설물에 대해 시득법에서는 정기적으로 점검을 실시하도록 규정한다.

점검의 세부 절차는 한국시설안전공단에서 발간하는 “안전 점검 및 정밀안전진단 세부지침”을 통해 규정된다. 시설물에 대한 점검은 점검 수준에 따라 정기점검, 정밀점검, 정밀안전진단으로 구분된다. 정기점검은 콘크리트·강재 구조물의 외관 상태를 평가하는 등 간단한 육안검사 수준의 점검이다. 정밀점검은 주요부재에 대한 외관조사망도 작성 및 콘크리트 구조물에 대한 균열, 누수 등의 손상 여부 조사, 강재 구조물의 균열, 도장상태 확인 등 보다 세심한 관찰과 콘크리트의 반발 경도 시험이나 탄산화 깊이 측정 등의 간단한 시험을 포함하며 필요 시 구조해석과 내하력평가를 통한 구조물의 안전성 평가가 수행될 수 있다. 정밀안전진단은 교량 주요부재뿐만 아니라 부속시설에 대한 점검과 같은 더욱 정밀한 외관조사와 콘크리트의 염화물 함유량 시험이나 강재의 비파괴시험 등의 각종 측정·시험을 바탕으로 시설물의 상태 및 안전성을 평가한다.

시설물의 안전등급 산정은 상태평가와 안전성평가로 이루어진다. 상태평가는 시설물의 외관 상태와 결함에 대한 평가이다. 상태등급(condition rating)은 우선 교량의 개별 부재별 등급을 산정한 뒤, 경간·지점별로 존재하는 개별 부재의 등급을 평균하여 교량 전체의 부재별 등급을 산정하고, 마지막으로 부재별 가중치를 적용한 부재 등급의 가중평균을 계산함으로써 산정된다. 안전성평가는 시설물의 구조적 안전성에 대한 평가로, 안전성평가 등급은 구조해석과 내하력평가 결과에 의한 안전율을 기준으로 산정된다. 허용응력설계법에 의해 설계되는 강교는 허용응력과 발생응력의 비율이 안전율로 정해지며, 강도설계법에 의해 설계되는 콘크리트교는 설계강도와 소요강도의 비율이 안전율로 정해진다. 최종적인 시설물의 안전등급은 상태평가 등급과 안전성평가 등급 중 낮은 결과의 등급값으로 산정된다. 등급 구분은 정기점검에는 양호·보통·불량의 3단계 평가 기준이 적용되며, 정밀점검과 정밀안전진단에는 A~E의 5단계 평가 기준이 적용된다(Table 1). A등급은 아무런 문제점이 없는 최상의 상태를 의미하며, E등급은 즉시 사용을 중지하고 필요한 조치가 이루어져야 하는 상태를 의미한다. 특히 C등급 이하는 주요부재의 결함으로 시설물의 기능 저하가 발생하여 이를 해결하기 위한 보수·보강이 필요한 상태를 의미한다.

교량시설물에 대한 다양한 점검 항목에 대해 제한된 시간 안에 검사를 진행하기 위해서는 숙련된 점검인력이 필요하며 정밀 안전진단의 경우에는 관리주체가 외부 점검업체에 용역을 맡겨

수행하기 때문에 교량 안전관리를 위한 전문인력과 예산을 확보하는 것이 필요하다. 그러나 정부 및 지방자치단체 산하의 안전관리 담당 부처에서 활용할 수 있는 예산 및 인력은 부족한 상황이다. 고성석 등(2011)의 연구에 의하면 시설물 관리기관에서 안전관리에 필요한 예산을 완전히 확보하고 있는 경우가 없는 것으로 조사되었으며, 안전관리 담당 부서에서 한 명의 전담 직원이 모든 안전관리 업무를 맡고 있는 등 전문인력 또한 부족한 것으로 조사되었다. 또한 대부분의 지방자치단체에서 시설물 안전점검 및 정밀안전진단에 필요한 예산을 확보하지 못하여 저가에 점검 용역을 주거나 점검을 적시에 실시하지 못하는 실정이다(박세훈, 2004).

진단을 위한 인력과 예산이 부족한 상황에서 비용 효율적인 등급 산정 방법을 제시하기 위하여 본 연구는 교량시설물의 과거 점검 데이터를 분석함으로써 현재의 교량 안전등급을 추정하는 모델을 개발하고자 한다. 본 연구에서는 교량 안전등급을 A~E의 5단계에서 B 이상과 C 이하의 두 등급으로 구분하였다. 교량 안전관리의 관심 대상은 보수·보강이 필요한 C등급 이하인 교량이기 때문에 등급을 5단계로 세분하여 추정하는 것보다 C등급 이하인 등급에 초점을 두는 것이 타당하기 때문이다. 또한 개발한 모델이 단순히 교량 안전등급을 추정할 뿐만 아니라 추정의 근거가 되는 교량의 각종 변수들의 분류 규칙을 제시할 수 있도록 한다. 즉, 교량의 제원, 준공년도, 교통량 등과 같이 교량의 안전등급에 영향을 미칠 것으로 예상되는 다양한 변수들이 실제로 안전등급에 영향을 주는 분류 규칙을 찾아내는 것이다. 여기에서 분류 규칙이란 예를 들어 “준공년도가 1995년 이전인 교량 중 설계하중이 DB-18 이하이고 일일교통량이 50,000

Table 1. Safety ratings of facility

Safety Rating	Description
A (Excellent)	The best condition without damages
B (Good)	The condition that minor damages, not affecting the functionality, occurred on secondary elements and partial repairs are required to enhance the durability.
C (Normal)	The condition that minor damages on primary elements or extensive damages on secondary elements occurred but the overall safety is not affected, and repairs for primary elements or reinforcements for secondary elements are required to maintain the durability and functionality.
D (Insufficient)	The condition that damages occurred on primary elements, so urgent repairs and reinforcements are required, and usage restriction should be decided.
E (Poor)	The condition that severe faults threatening the safety of facility occurred on primary members, the service of facility must be restricted immediately, and reinforcements or a reconstruction is required.

대 이상인 교량은 안전등급이 C이다”와 같이 교량의 안전등급 추정의 근거가 되는 변수 및 변수 값들의 조합이다. 본 연구는 이를 통해 교량 안전등급 악화에 영향을 주는 분류 규칙을 발견함으로써, 발견된 규칙이 향후 교량시설물의 안전관리 의사결정 지원 자료로서 활용될 수 있도록 한다.

1.2 연구의 범위 및 방법

본 연구는 1·2종 도로교량 시설물을 대상으로 한다. 분석에 사용할 데이터는 시설물정보관리종합시스템(Facility Management System, FMS)에서 제공하는 도로교량을 포함한 모든 종류의 1·2종 시설물에 대한 기본적인 정보 및 안전점검 결과와, 국토교통부에서 매년 발간하는 전국 3만여 개 도로교량의 기본제원, 구조형식, 설계하중 등 총 19개 항목에 대한 정보가 기록된 “도로교량 및 터널 현황조사(2015)”에서 수집하였다.

1·2종 시설물 중 최근에 준공되어 정기점검 결과만 존재하는 도로교량은 연구의 범위에서 제외하였다. 정밀점검이나 정밀안전진단은 수행된 뒤 객관적 절차에 의해 A~E등급의 5단계 안전등급이 산정되지만, 정기점검의 경우에는 구체적인 등급이 산정되지 않고 양호·보통·불량의 다소 주관적인 결과가 기록되기 때문이다.

연구의 방법은 다음과 같다(Fig. 1). 먼저 1) 관련 선행연구를 조사하여 연구에서 모델 개발에 활용할 데이터마이닝 기법을

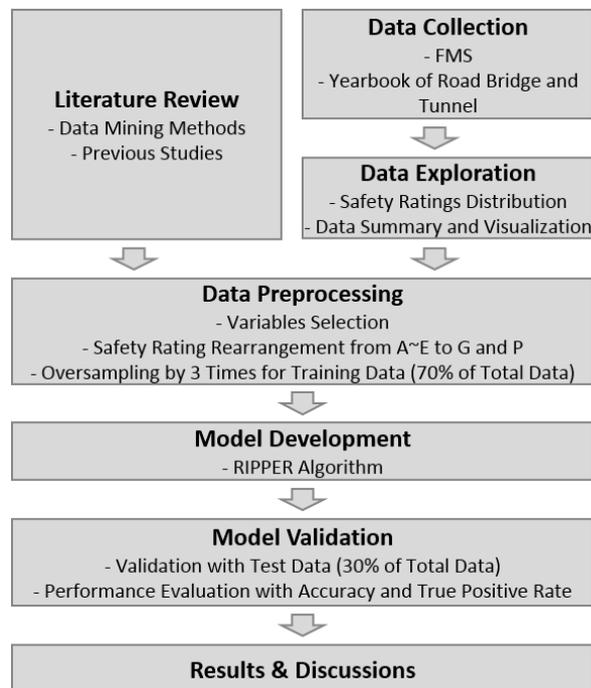


Figure 1. Research process

선정하고 2) FMS와 도로교량 및 터널 현황조사 데이터를 수집, 하나의 데이터베이스를 구축하였으며, 3) 수집한 데이터의 시각화를 통해 데이터를 탐색하고 및 필요한 전처리 과정을 도출하였다. 4) 도출된 데이터 전처리를 수행하고 5) 선정된 데이터마이닝 기법을 적용하여 도로교량 안전등급 추정 모델을 개발하였으며, 6) 각종 성능 지표를 통해 개발된 모델을 검증하고 7) 모델의 안전등급 추정 결과 및 도출된 안전등급 분류 규칙에 대해 논의하였다. 모델 개발에 사용하는 데이터 분석 도구는 오픈 데이터마이닝 소프트웨어인 Weka를 활용하였다.

2. 문헌 고찰

2.1 데이터마이닝 방법론에 대한 고찰

데이터마이닝은 데이터로부터 인간이 의사결정에 활용할 수 있도록 의미 있는 지식정보를 도출해내는 작업으로 정의할 수 있다(Han et al., 2012). 데이터마이닝은 목적에 따라 크게 지도학습(supervised learning)과 비지도학습(unsupervised learning)으로 구분된다. 지도학습은 예를 들어 교량 데이터의 각종 변수(준공연도, 교장, 교통량 등)로부터 교량의 안전등급을 추정하는 것과 같이, 데이터의 여러 변수 중 예측하고자 하는 특정 변수(class)가 존재하며, 이 클래스의 값을 추정하기 위하여 나머지 다른 변수들을 활용하는 방법이다. 비지도학습은 특정 클래스의 값을 맞추고자 하는 것이 목표가 아니라 데이터를 군집화하고 변수들 간의 연관성을 파악하는 것이 목표일 때 활용되는 방법이다.

지도학습을 통해 클래스를 추정하는 모델을 개발하고자 할 때, 개발된 모델의 성능은 주어진 데이터를 얼마나 잘 설명하는 지가 아니라 모델 개발에 사용되지 않은 새로운 데이터의 클래스를 잘 추정하는지에 의해 검증된다. 새로운 데이터의 클래스에 대한 추정 성능을 검증하기 위해, 지도학습 방법에서는 수집한 데이터를 전부 활용해서 모델을 개발하지 않고 데이터의 일부만 모델 개발에 활용하는데 이 데이터를 훈련 데이터(training data)라 한다. 나머지 데이터는 구축된 모델에 투입하여 모델의 정확성을 검증하는 데 활용하며 이를 평가 데이터(test data)라 한다.

지도학습 방법은 추정하고자 하는 변수의 형식에 따라 분류(classification)와 예측(prediction)으로 구분된다. 결과변수가 A·B·C·D·E와 같이 이산화된 값을 갖는 경우를 분류라 하고, 점수와 같이 연속적인 값을 갖는 경우를 예측이라 한다. 본 연구는 추정하고자 하는 특정 변수가 존재하고 이 변수가 A~E등급의 이산화된 값을 가지므로 데이터마이닝 방법 중 지도학습의 분류 방법을 활용하였다.

2.2 선행연구 조사

교량의 안전등급 또는 안전등급에 해당하는 Safety Index 등의 지표를 데이터마이닝의 분류 기법을 이용하여 추정하는 연구는 크게 1) 인공신경망(artificial neural network) 기법을 이용한 연구 2) 의사결정나무(decision tree) 기법을 이용한 연구 3) 규칙 기반 분류(rule-based classification) 기법을 이용한 연구로 구분할 수 있다(Table 2).

인공신경망 기법은 다른 분류 기법에 비해 예측 성능이 뛰어난 것으로 알려져 있다. 그러나 인공신경망 기법을 이용한 모델은 1) 소위 'black-box' 모델로서 어떤 과정에 의해 결과가 분류되는지 사용자가 알 수 없다는 점과 2) 훈련 데이터의 노이즈까지도 예측하는 이른바 과적합(overfitting)된 모델이 만들어질 가능성이 크며 결과적으로 새로운 데이터에 대한 분류 성능이 확보되지 못할 가능성이 크다는 점이 한계점이다.

의사결정나무 기법을 활용한 모델은 데이터가 분류되는 과정이 사용자에게 명확하게 전달된다는 점과 분류에 사용되는 변수

개수가 적을 경우 분류 규칙을 나무 형태로 시각화하기 좋은 점이 있다. 그러나 1) 변수 간의 상호작용이 잘 반영되지 않고 2) 분류에 사용되는 변수가 많아지면 나무가 지나치게 복잡해져서 시각화의 효과가 떨어진다는 단점이 있다.

규칙 기반 분류는 주어진 데이터를 가장 잘 분류할 수 있는 IF-THEN 형태의 규칙을 생성함으로써 데이터의 클래스를 분류하는 기법이다. 이 방법을 활용하여 모델을 개발할 경우 모델의 사용자에게 클래스를 분류하는 규칙을 제공하면서도 의사결정 나무 기법을 활용한 모델과 달리 규칙 개수가 지나치게 많아지지 않는다는 장점이 있다. 교량의 상태와 관련된 규칙을 찾으려 한 선행연구로 Huang and Chen(2012)의 연구에서는 미국의 National Bridge Inventory 데이터를 활용하여 교량 데이터를 유사한 것끼리 군집화하고 이들 군집이 형성되는 규칙을 도출하였다. 그러나 이 연구에서는 규칙과 교량의 상태등급과의 구체적인 연관성이 도출되지 않았다는 한계점이 있다. 이와 달리 본 연구는 국내 도로교량을 대상으로 교량의 안전등급을 결정하는 분류 규칙을 찾아내어 안전등급을 분류하였다. 본 연구에서는 분류 기법 중 Repeated Incremental Pruning to Produce Error Reduction(RIPPER) 알고리즘을 활용하여 분류 모델을 개발하였다. RIPPER 알고리즘은 노이즈가 많은 데이터에 대해서도 분류 성능이 준수하며 특정 클래스에 더 큰 관심을 두는 경우에 관심 대상 클래스의 데이터를 잘 분류해 준다는 장점이 있다(Cohen, 1995).

RIPPER 알고리즘은 크게 규칙 생성 단계와 최적화 단계로 구분할 수 있다. 규칙 생성 단계에서는 우선 주어진 훈련 데이터 중 수가 적은 클래스를 분류할 수 있는 규칙 집합을 생성한다. 구체적으로, 규칙을 생성할 때마다 데이터를 가장 잘 분류할 수 있는 하나의 변수를 찾고 해당 변수가 어떤 값을 가질 때 데이터들의 클래스가 가장 잘 분류되는지 찾는다. 예를 들면 본 연구에서 사용한 훈련 데이터를 가장 잘 분류하는 변수는 준공연도 변수이며 준공연도가 1995년 일 때를 기준으로 1995년 이하인 데이터는 클래스 P이고 나머지 연도는 클래스 G인 데이터가 많다고 할 때, 첫 번째 규칙은 (준공연도 <= 1995) then (클래스 = P) 가 된다. 그러나 준공연도가 1995 이하인 교량이 전부 클래스 P인 것은 아니기 때문에, 이 규칙에 해당하는 데이터들을 다시 가장 잘 분류할 수 있는 변수와 변수값의 조합을 찾는다. 그 결과 설계하중을 기준으로 설계하중이 DB-18인 경우 그렇지 않은 경우에 비해 클래스 P인 데이터가 많다고 하면, 첫 번째 규칙은 (준공연도 <= 1995) and (설계하중 = DB-18) then (클래스 = P)로 바뀐다. 이렇게 생성된 규칙에 대해 규칙에 해당하는 데이터 수와 규칙에 의한 클래스의 분류 정확도가 사용자가 정한 수준 이상일 경우 최종적으로 하나의 규칙이 생성된다. 첫 번째 규칙에 해당하지 않는, 즉 준공연도가 1995년 이후이거나 설계

Table 2. Researches on estimating safety index of bridge

Researcher	Analysis Method	Contents
Cattan and Mohammadi (1997)	Artificial Neural Network	This study proposed an artificial neural network model which used characteristics of bridges and results of structural analysis as input variables to predict inspection ratings by experts.
Lin et al. (2002)	Artificial Neural Network	This study proposed 1) an artificial neural network model using characteristics of earthquake events as input variables to estimate peak ground acceleration (PGA) and 2) an artificial neural network model using PGA and characteristics of bridges as input variables to predict degrees of bridges damages.
Lee et al. (2008)	Artificial Neural Network	This study proposed an ANN-based backward prediction model using recent inspection results and external factors like traffic, climate and population as input variables to predict the previous safety index
Huang (2010)	Artificial Neural Network	This study proposed an artificial neural network model using historical inspection data of bridges as input variables to predict deterioration of bridge decks.
Huang and Chen (2012)	Clustering and Rule-based Classification	This study proposed a 'clustering' methodology for grouping bridges and extracted rules associated with created clusters using National Bridge Inventory data.
Bektas et al. (2013)	Classification Tree	This study proposed a classification tree model using performance scores of bridges elements as input variables to predict safety index of National Bridge Inventory

하중이 DB-18이 아닌 데이터에 대해서 다시 이 과정을 반복함으로써 두 번째 규칙을 생성한다. 이러한 방법으로 데이터 개수가 적은 클래스의 데이터가 모두 포함되는 규칙 집합을 형성하는 작업이 규칙 생성 단계에서 이루어진다. 다음으로 생성된 규칙 중 그 규칙에 해당하는 데이터 개수가 일정 수준 이하인 규칙은 제거하는 “가지치기(pruning)”를 진행한다.

최적화 단계에서는 훈련 데이터 중 무작위로 추출된 일부 데이터로부터 두 개의 규칙을 만든 뒤 이들을 비교한다. 이 때 한 규칙은 규칙 생성 단계에서 생성된 규칙에 변수를 추가함으로써 생성되고, 다른 한 규칙은 아무런 규칙도 참조하지 않고 새로이 규칙을 만들어 나간다. 생성된 두 개의 규칙의 분류 성능을 비교한 뒤 성능이 좋은 규칙을 최종적인 분류 규칙으로 정한다. 이러한 과정을 통해 규칙 집합이 훈련 데이터에 과적합되지 않도록 한다.

3. 분류 모델 구축

3.1 데이터 탐색 및 시각화

한국시설안전공단은 시설물정보관리종합시스템을 통해 시·특법에 의해 1, 2종으로 규정된 모든 종류의 시설물의 안전등급 정보를 공개하고 있다. 도로교량을 포함한 모든 시설물에 대하여 공통적으로 시설물명, 시설물구분, 시설물종류, 종별, 위치, 점검진단일, 차기예정일, 등급 정보를 제공하며, 추가적으로 도로교량의 경우 시설물 주요제원(형식, 연장, 차로수, 경간수, 최

Table 3. Variables of facility management system (FMS)

Variables	Data Types	Examples
Name of Facility	Text	Gayang Bridge, Gangdong Bridge
Facility Classification	Categorical	Bridge, Tunnel, Building
Type of Facility	Categorical	Road Bridge, Railway Bridge
Facility Class	Categorical	First-class, Second-class
Location	Text	Gayang-dong, Gangseo-Gu, Seoul
Last Inspection Date	Date	2014-11-15
Scheduled Inspection Date	Date	2016-11-15
Safety Rating	Categorical	A, B, C, D, E
Structural Type	Categorical	Steel-box, RC-box, PC-box, etc.
Length	Numeric	1,700m
Number of Lanes	Integer	6
Number of Spans	Integer	30
Maximum Span Length	Numeric	180
Data Collection Organization	Categorical	(Organizations that collected data of bridges)
Pictures	Others	(Pictures of bridges)

대경간장), 취합기관, 사진 정보를 제공한다. 각 변수의 데이터 형식 및 데이터 예시는 Table 3과 같다. 그러나 시설물 공통 정보 및 취합기관과 사진 정보는 시설물의 안전등급에 영향을 미치는 요인이기보다 시설물의 관리를 위한 정보로서의 성격이 강하므로 본 연구의 분석 대상에서 제외하였다.

도로교량의 안전등급에 영향을 미칠 것으로 예상되는 시설물 주요제원 정보를 파악하는 과정에서, FMS에서 제공되는 변수의 수가 매우 한정적이기 때문에 도로교량과 관련된 추가적인 정보를 수집하였다. 본 연구는 한국건설기술연구원에서 제공하는 도로교량 및 터널 현황조사서(2015)로부터 도로교량 관련 데이터를 수집하였다. 도로교량 및 터널 현황조사서에는 전국 도로교량의 고속국도, 일반국도, 특별·광역시도, 국가지원지방도·지방도, 시·군·구도상의 도로 교량과 관련된 도로종류, 노선명, 시설명, 시도, 시군구, 읍면동, 리, 총길이, 총폭, 유효폭, 높이, 경간수, 최대경간장, 상부구조, 하부구조, 설계하중, 교통량, 준공년도, 비교의 총 19개 정보가 제공된다(Table 4). 2015년에 발간된 현황조사서에는 2014년 12월 31일 기준으로 전국 29,896개의 도로교량 정보가 제공되었으며, 이 중 본 연구에서는 1, 2종 시설물에 해당하는 6,679개 도로교량의 정보와 FMS에서 제공하는 안전

Table 4. Variables of yearbook of road bridge and tunnel statistics (2015)

Variables	Data Types	Examples
Road Type	Categorical	Expressway, National Way
Road Name	Categorical	Expressway 12, Expressway 55
Facility Name	Text	Gagok Bridge, Ganam Bridge
Province / (Metropolitan) City	Categorical	Seoul, Busan, Gyeonggi-do, Gangwon-do
City / District	Categorical	Seongnam-si, Gangnam-gu, Gangseo-gu
Town (Eup / Myeon / Dong)	Categorical	Nongong-eup, Gonggeun-myeon
Town (Ri)	Categorical	Wicheon-ri, Gagok-ri
Total Length	Numeric	812.0
Total Width	Numeric	11.7
Effective Width	Numeric	10.7
Height	Numeric	25.0
Number of Spans	Integer	22
Maximum Span Length	Numeric	60.0
Superstructure Type	Categorical	PSC I, Steel Box
Substructure Type	Categorical	T Pier, V Pier, Rahmen, Others
Design Load	Categorical	DB-24, DB-18
Traffic	Integer	15094
Year of Completion	Integer	1984
Note	Text	(Explanatory note about bridges)

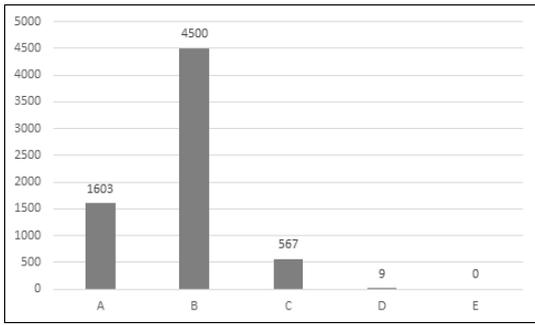


Figure 2. Distribution of safety ratings

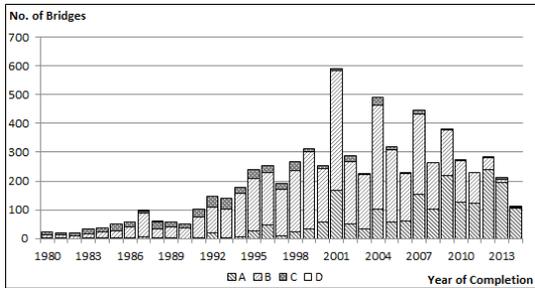


Figure 3. Distribution of safety ratings by year of completion

등급 정보를 수집하였다.

수집된 데이터를 탐색한 결과, 총 6,679개의 교량 중 안전등급이 A인 교량이 1,603개, B인 교량 4,500개로 전체의 91.4%를 차지하는 것으로 나타났다. 안전등급이 C인 교량은 567개로 전체의 8.5%였으며, D는 전체 중 9개밖에 존재하지 않는 것으로 나타났다. 특히, 안전등급이 E인 교량은 단 한 개도 존재하지 않는 것으로 나타났다(Fig. 2).

준공연도 별 교량 개수 및 등급 분포를 분석한 결과, 1990년대 이후부터 교량 건설이 급격히 증가하였으며 최근에 준공된 교량일수록 A등급의 비중이 높고 준공 후 오래 된 교량일수록 B등급 및 C등급의 비중이 증가하는 경향성을 확인하였다(Fig. 3).

3.2 모델의 입력 변수 선택

본 연구에서는 분류 모델에 투입할 변수를 선택하기 위해 각 변수의 1) 교량 안전등급과의 관련성 2) 분석의 정밀도 수준 3) 결측치의 비율, 4) 수치형 변수인 경우 다른 수치형 변수와의 상관관계를 고려하였다. 도로교량 및 터널 현황조서의 변수 중 시설명 변수는 교량의 안전등급과 관련이 없으므로 분석에서 제외하였다. 분석의 지리적 정밀도는 시·도 수준으로 설정하여 시도 변수만 활용하고 세부 주소 변수(시군구, 읍면동, 리)는 제외하였다. 노선명 변수와 비교 변수의 경우 결측치의 비율이 각각 20.1%와 98.8%로 높은 편이므로 유의미한 분석에 어려움이 있을 것으로 판단되어 제외하였다. 마지막으로 총길이, 총폭, 유효폭, 높이, 경간수, 최대경간장 변수 간의 상관관계를 분석하여

Table 5. Coefficients of correlation between numerical variables

	Total Length	Total Width	Effective Width	Height	Number of Spans	Maximum Span Length
Total Length	1	-	-	-	-	-
Total Width	.07	1	-	-	-	-
Effective Width	.07	.94	1	-	-	-
Height	.23	-.12	-.09	1	-	-
Number of Spans	.88	.02	.02	.13	1	-
Maximum Span Length	.36	.11	.10	.23	.10	1

수치형 변수들 간의 상호 연관성을 확인하였다(Table 5). 분석 결과, 유효폭과 총폭 그리고 총길이와 경간수 변수 간에 상관관계 수가 큰 것을 확인하였으며 총길이 변수의 경우 높이와 최대경간장 변수와도 일정 부분 상관관계가 있는 것을 확인하였다. 본 연구에서는 유효폭과 총길이의 두 변수를 제외하고 최종적으로 도로종류, 시도, 총폭, 높이, 경간수, 최대경간장, 상부구조, 하부구조, 설계하중, 교통량, 준공년도, 안전등급의 12개의 변수를 분류 모델의 투입변수로 선택하였다.

3.3 훈련 데이터 및 검증 데이터 설정

본 연구에서 개발한 모델이 실제로 새로운 데이터가 입력되어도 안전등급 추정을 올바르게 할 수 있는지 검증하기 위해, 모델 개발 시 수집한 6,679개의 데이터 중 70%인 4,675개의 데이터를 엑셀의 난수생성 함수를 통해 무작위 추출하여 훈련 데이터로 설정하였으며 나머지 30%인 2,004개의 데이터를 검증 데이터로 설정하였다. 무작위 추출 결과 안전등급이 B 이상인 교량이 4295개로 훈련 데이터의 91.9%를 차지하여 원본 데이터와 유사한 안전등급 분포를 보이는 것을 확인하였다.

3.4 안전등급의 클래스 통합 및 오버샘플링

도로교량의 안전등급은 B등급 이상인 교량이 90%를 넘는 매우 편중된 분포를 보였다. 이와 같이 클래스가 한쪽으로 편중된 데이터를 활용하여 분류 모델을 개발할 경우 모델의 분류 성능은 일반적으로 좋지 않다. 극단적인 예를 들면 “모든 교량이 안전등급 B등급 이상임”이라는 결과를 내놓는 모델을 개발할 경우 이 모델의 분류 정확도는 91.4%이며, 여러 변수를 활용한 복잡한 분류 모델을 만들더라도 분류 정확도가 향상되지 않을 가능성이 크다. 도로교량을 안전하게 유지·관리에 있어 중요한 것은 안전성능이 저하된 도로교량을 식별해 내는 것이다. 즉, 안전성능이 불량한 도로교량의 안전등급을 낮게 추정할 가능성을 높이는 것이 도로교량의 안전등급 추정 모델에 있어 중요하다. 따라서 도로교량의 안전등급 분류 모델의 성능은 모델의 전체적인 분류 정확도뿐만 아니라 실제로 낮은 안전등급의 교량을

얼마나 잘 찾아내는지를 통해 평가되어야 한다.

분류 기법을 활용할 경우 전체 클래스를 비슷하게 고려하는 것이 아니라 특정 클래스를 분류하는 것에 더욱 초점을 두어야 할 경우가 존재한다. 이러한 경우 관심 대상인 데이터를 positive하다고 하며, 그렇지 않은 데이터를 negative하다고 한다. 또한 분류 모델을 통해 데이터의 클래스를 분류한 결과 실제로 positive한 데이터 중에서 분류 모델에 의해 positive하다고 분류된 데이터를 true positive라 하며 실제로는 negative인 데이터 중 모델에 의해 positive라고 분류된 데이터를 false positive라 한다. 실제 클래스가 negative한 데이터에 대해서도 분류 결과에 따라 true negative와 false negative로 구분할 수 있다(Table 6).

이러한 구분은 보통 positive 데이터의 수가 적은 경우에 고려되는데, 이는 전체적인 분류 정확도만으로는 분류 모델의 성능을 판단하기에 부족하기 때문이다. True positive 데이터와 true negative 데이터는 분류가 정확하게 된 데이터이며 false positive 데이터와 false negative 데이터는 분류가 잘못된 데이터이다. 따라서 분류 모델의 전체적인 예측 정확도(accuracy)는 식 (1)과 같이 표현된다.

$$Accuracy = \frac{TP + TN}{Total} \quad (1)$$

실제로는 positive인 데이터를 얼마나 잘 분류했는지는 true positive rate(TPR) 지표를 통해 표현된다. 이는 실제로 positive인 데이터 중에서 분류 모델에 의해 정확하게 positive로 분류된 데이터의 비율이다(식 (2)).

$$TPR = \frac{TP}{Actual\ Positive} \quad (2)$$

본 연구에서 개발하고자 하는 안전등급 분류 모델의 성능은 전체적인 분류 정확도뿐만 아니라 TPR을 고려하여 판단되어야 한다.

클래스가 편중된 경우에도 분류 모델의 성능을 향상시키기 위해서 오버샘플링(oversampling), 언더샘플링(undersampling), 역치조정(threshold moving), 앙상블(ensemble) 기법 등을 활용할 수 있다. 그런데 역치 조정 방법의 경우 클래스가 세 개 이상인 경우에 적용이 불가능하며, 다른 기법들은 클래스가 세

Table 6. Confusion matrix

		Predicted Class	
		Positive	Negative
Actual Class	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

개 이상인 경우에도 적용은 가능하지만 모델의 분류 성능을 향상시키는 효과가 매우 떨어진다(Wang et al., 2012).

그러므로 안전등급이 낮은 도로교량에 대해서도 등급 추정 정확도를 확보하기 위해서는 현재 5단계로 구분되어 있는 안전등급을 그대로 활용하는 것보다 이를 두 개의 클래스로 통합하는 것이 필요하다. 안전등급의 정의에 따라 A등급과 B등급은 시설물의 기능 발휘에 지장이 없는 상태를 의미하는 반면, C등급 이하는 시설물이 본래의 기능을 온전히 발휘할 수 없는 상태를 의미한다. 또한 국토교통부의 제2차 시설물의 안전 및 유지관리 기본계획에서는 시설물 유지관리의 목표를 A, B등급의 비율을 90% 이상으로 유지시키는 것으로 정하고 있다. 따라서 본 연구에서는 교량의 기능이 온전히 발휘되지 않는 C 이하의 등급을 분류의 관심 대상으로 설정하고 이를 P(Poor)라는 하나의 클래스로 통합하였다. 나머지 A등급과 B등급은 기능이 온전히 발휘되는 것으로 볼 수 있으며 이를 G(Good)라는 하나의 클래스로 통합하였다.

오버샘플링 기법과 언더샘플링 기법은 positive 데이터의 수가 적을 때 positive 데이터와 negative 데이터의 비율을 비슷하게 만들기 위해 사용되는 기법이다. 오버샘플링은 수가 적은 positive 데이터의 개수를 증가시키는 방법이고 언더샘플링은 수가 많은 negative 데이터를 감소시키는 방법이다. 오버샘플링은 데이터 개수를 증가시키기 때문에 모델 구축에 필요한 계산 시간이 증가하는 단점이 있는 반면 전체 데이터를 활용하여 모델을 개발하기 때문에 모델의 신뢰도가 손상되지 않는다는 장점이 있다. 언더샘플링은 반대로 데이터 개수를 감소시키기 때문에 계산 시간이 단축된다는 장점이 있지만 전체 데이터 중 일부만을 모델 개발에 사용하기 때문에 모델의 신뢰도가 저하된다는 단점이 있다(박성훈, 2014).

본 연구에서 수집한 6,679개의 데이터는 처리 속도의 작은 변화에 민감하게 반응할 만큼 큰 데이터가 아니기 때문에, 계산 속도를 향상시키는 것 보다 데이터를 최대한 많이 활용하여 분류 모델의 신뢰도를 확보하는 것이 중요하다. 따라서 본 연구에서는 오버샘플링 기법을 적용하여 연구의 관심 대상인 클래스 P인 데이터를 증가시켰다. 오버샘플링을 몇 배나 적용할지 정하기 위해 분류 모델의 TPR과 생성되는 규칙 개수가 오버샘플링 정도에 따라 어떻게 변화하는지 분석하였다(Fig. 4). 그 결과, 오버샘플링이 없는 경우에 비해 오버샘플링이 2배, 3배로 적용되면서 TPR이 크게 증가하였으나 3배 이후로는 TPR이 크게 증가하지 않거나 오히려 감소하는 것을 확인하였다. 또한 생성되는 규칙의 개수는 대체로 오버샘플링 비율이 증가할수록 증가하는 경향을 보였으나 9배 이상으로 오버샘플링이 적용될 경우 다소 감소하는 것을 확인하였다. 이러한 결과는 positive 데이터가 증가함에 따라 훈련 데이터에 존재하는 positive 데이터에만

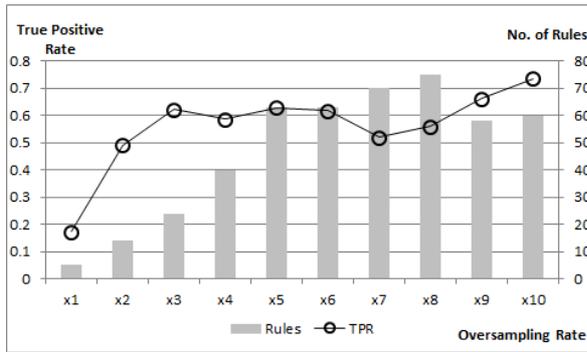


Figure 4. Changes in true positive rate and number of rules by oversampling rate

과적합된 규칙이 생성되면서 결과적으로 새로운 데이터에 대한 분류 성능이 떨어지는 것으로 볼 수 있다. 따라서 본 연구에서는 오버샘플링 비율을 3배로 한 훈련 데이터를 활용하여 분류 모델을 개발하였다.

3.5 모델 매개변수 설정 및 모델 검증

RIPPER 알고리즘의 실행에 필요한 매개변수는 folds의 개수와 최적화 횟수가 있다. Folds의 개수는 규칙 생성 단계에서 규칙 집합을 생성하고 이를 가지치기하는 데 사용되고, 최적화 횟수는 최적화 단계를 몇 번이나 반복할지 지정하는 매개변수이다. 본 연구에서는 매개변수의 조합에 따라 검증 데이터에 대한 모델의 분류 성능이 어떻게 변하는지 분석함으로써 최적의 파라미터 조합을 설정하였다(Table 7). 분석 결과, 규칙의 개수는 folds 값이 작을수록, 최적화 횟수가 클수록 증가하였으나, TPR은 변수 값의 변화에 따른 뚜렷한 경향이 나타나지 않았다. 본 연구에서는 클래스 P에 대한 TPR이 가장 높게 나타난 매개변수 조합인 Folds 10, 최적화 횟수는 500으로 분류 모델을 개발하였다.

데이터의 개수가 적어서 훈련 데이터와 검증 데이터로 나누는 것만으로는 새로운 데이터에 대한 모델의 예측력을 확인하기 어려울 때 k겹 교차검증(k-folds cross validation) 방법을 사용할 수 있다. 이 방법은 우선 데이터를 k개의 부분집합으로 분할한 뒤, 첫 번째 부분집합을 검증 데이터로 활용하고 나머지를 훈련 데이터로 활용하여 모델 성능을 확인하고 이후 두 번째, 세 번째 데이터를 차례로 검증 데이터로 활용하여 모델의 성능을 검증하는 방식이다. 부분집합의 개수를 10개로 하여 10겹 교차검증을 수행한 결과 클래스 P에 대한 TPR은 0.817로 나타나 수집되지 않은 새로운 데이터에 대한 예측력도 높은 수준이다.

4. 결과 및 논의

구축된 분류 모델의 전체적인 분류 정확도는 84.8%, 클래스

Table 7. Classification performance according to number of folds and number of optimization

TPR / Number of Rules	Folds: 3 (default)	10	20	50	100	500
Optimization : 2 (default)	0,622 / 24	0,582 / 14	0,612 / 17	0,515 / 7	0,510 / 7	0,526 / 7
10	0,612 / 31	0,638 / 19	0,571 / 25	0,612 / 17	0,577 / 8	0,536 / 7
20	0,617 / 34	0,648 / 22	0,571 / 26	0,602 / 18	0,607 / 11	0,566 / 9
50	0,607 / 34	0,648 / 22	0,587 / 33	0,622 / 18	0,648 / 12	0,577 / 11
100	0,602 / 35	0,653 / 23	0,582 / 33	0,622 / 23	0,633 / 12	0,577 / 11
500	0,597 / 49	0,673 / 29	0,577 / 35	0,617 / 26	0,628 / 20	0,582 / 14

P에 대한 TPR은 67.3%, 규칙 개수는 29개이다. 생성된 규칙은 Table 8과 같으며, 가장 마지막 규칙을 제외한 28개의 규칙은 모두 데이터를 클래스 P로 분류하는 규칙이다. 표에서 cover는 규칙에 해당하는 데이터 개수를 의미하고 error는 규칙에 해당하지만 분류가 잘못된 데이터의 개수를 의미한다.

규칙 집합의 상위에 있는 규칙일수록 주어진 훈련 데이터를 잘 설명하는 규칙이며, 몇 가지 규칙으로부터 교량의 성능 저하에 관한 일반론을 도출하였다.

- 1) 준공연도가 1995년 이전, 즉 준공 후 대략 20년 이상 된 교량 중 설계하중 DB-18로 설계된 2등급은 설계기준이 높게 요구되는 다른 교량에 비해 안전등급이 낮을 가능성이 크다. 이는 낮은 설계하중이 적용되어 설계된 교량은 내하력이 상대적으로 작기 때문에, 같은 차량하중에도 손상이 발생하기 쉽기 때문이다(오병환 외, 2001).
- 2) 준공연도가 1993년 이전이고 교통량이 23967 이하인 교량이 낮은 안전등급을 가질 가능성이 크다. 이는 차량통행이 많은 주요 도로의 교량은 그 중요도 때문에 유지관리가 잘 되어 안전등급이 B 이상으로 유지되는 반면, 차량통행이 적은 교량은 상대적으로 안전관리에 소홀하여 낮은 안전등급을 갖게 될 가능성이 큰 것으로 생각할 수 있다.
- 3) 1995년 이전에 준공된 RC 슬래브교는 안전등급이 낮을 가능성이 크다. 프리스트레스 콘크리트나 강에 비해 RC는 재료의 내하력이 작으며 손상이 발생할 가능성이 더욱 크다(김영진, 유동우, 2002). 따라서 RC 슬래브교의 안전등급이 낮을 가능성이 크다고 볼 수 있다.
- 4) 준공연도 변수는 29개 규칙 중 24개 규칙에서 첫 번째 변수로 등장하였으며 총 34번 등장하였다. 이는 교량의 안전성 성능 저하에 교량의 다른 제원보다 준공 후 경과 연수가 가장 큰 영향을 미친다는 것을 의미한다.

그러나 생성된 규칙으로부터 일반적인 논의 외에 심도 있는 도로교량의 노후화 규칙을 찾아내기 어려웠다. 이는 교량의 안전성능에 영향을 미칠 수 있는 많은 변수 중 일부의 변수만을 모델 개발에 활용하였기 때문이며, 또한 FMS에서 제공하는

Table 8. Set of rules

Rules	Cover	Error
(Year of Completion (\leq 1995) and (Design Load = DB-18)	487	151
(Year of Completion (\leq 1998) and (Year of Completion (\leq 1993) and (Traffic (\leq 23967)	310	112
(Year of Completion (\leq 1998) and (Total Width (\geq 18) and (Province / City = Jeollabuk-do)	43	7
(Year of Completion (\leq 1998) and (Number of Spans (\geq 7) and (Total Width (\leq 10.3) and (Year of Completion (\leq 1995) and (Number of Spans (\leq 12)	39	9
(Year of Completion (\leq 1998) and (Superstructure Type = RC Slab) and (Year of Completion (\leq 1995)	76	28
(Year of Completion (\leq 1998) and (Traffic (\leq 23231) and (Height (\geq 16) and (Maximum Span Length (\geq 50) and (Maximum Span Length (\leq 102)	40	10
(Year of Completion (\leq 1998) and (Total Width (\geq 18) and (Number of Spans (\leq 4) and (Year of Completion (\leq 1994) and (Height (\geq 5.8)	44	14
(Year of Completion (\leq 1999) and (Traffic (\leq 4797) and (Maximum Span Length (\leq 28) and (Total Width (\geq 9.5) and (Traffic (\geq 1130)	43	13
(Year of Completion (\leq 1998) and (Total Width (\geq 19) and (Number of Spans (\geq 8) and (Province / City = Gyeonggi-do)	40	13
(Year of Completion (\leq 2005) and (Maximum Span Length (\geq 60) and (Year of Completion (\geq 2004) and (Substructure Type = Others) and (Total Width (\leq 12.3)	36	6
(Year of Completion (\leq 1998) and (Number of Spans (\leq 2) and (Number of Spans (\geq 2) and (Maximum Span Length (\leq 50.1)	24	6
(Year of Completion (\leq 1998) and (Road Type = Cityway) and (Height (\geq 5) and (Number of Spans (\geq 6) and (Number of Spans (\leq 9) and (Substructure Type = T Pier)	21	3
(Year of Completion (\leq 2007) and (Traffic (\leq 3243) and (Number of Spans (\geq 7) and (Superstructure Type = PreFlex)	9	0
(Year of Completion (\leq 2007) and (Total Width (\leq 10) and (Maximum Span Length (\leq 20) and (Substructure Type = Gravity Pier)	16	4
(Year of Completion (\leq 2007) and (Total Width (\geq 14.7) and (Year of Completion (\leq 2000) and (Province / City = Jeollanam-do) and (Maximum Span Length (\geq 50) and (Year of Completion (\geq 1996)	18	0
(Year of Completion (\leq 2002) and (Total Width (\leq 10) and (Height (\geq 6) and (Height (\leq 8) and (Maximum Span Length (\geq 31) and (Total Width (\geq 9.2)	23	5
(Year of Completion (\leq 2007) and (Province/City = Gyeongsangbuk-do) and (Year of Completion (\geq 2007) and (Traffic (\geq 28018) and (Maximum Span Length (\leq 35) and (Total Width (\leq 12.6)	33	6
(Year of Completion (\leq 2006) and (Province / City = Jeollabuk-do) and (Total Width (\geq 15.8) and (Maximum Span Length (\geq 45) and (Number of Spans (\geq 3)	18	3
(Year of Completion (\leq 1998) and (Year of Completion (\leq 1993) and (Substructure Type = Rahmen Pier) and (Maximum Span Length (\leq 40) and (Maximum Span Length (\geq 28.4)	34	7
(Year of Completion (\leq 1997) and (Province / City = Gyeongsangbuk-do) and (Total Width (\geq 20)	11	2
(Year of Completion (\leq 1997) and (Traffic (\leq 5524) and (Height (\geq 10.6) and (Height (\leq 13.4) and (Traffic (\geq 1726)	13	1
(Year of Completion (\leq 2006) and (Total Width (\geq 19.5) and (Traffic (\leq 33362) and (Province / City = Gyeonggi-do) and (Height (\geq 9.6) and (Number of Spans (\leq 6)	14	2
(Year of Completion (\leq 1997) and (Substructure Type = T Pier) and (Height (\geq 10.6) and (Year of Completion (\leq 1986)	10	1
(Year of Completion (\leq 2004) and (Height (\leq 8.7) and (Height (\geq 8.2) and (Number of Spans (\geq 9) and (Traffic (\geq 16656)	13	1
(Total Width (\leq 10) and (Traffic (\leq 446) and (Height (\geq 6.4)	10	1
(Height (\leq 13.4) and (Height (\geq 13.1) and (Maximum Span Length (\leq 30)	13	4
(Traffic (\leq 7348) and (Traffic (\geq 7000) and (Traffic (\leq 7000)	6	0
(Total Width (\leq 10) and (Superstructure Type = Preflex) and (Province / City = Gyeongsangnam-do)	6	0
else then (class = G)	3985	99

안전등급은 해당 도로교량의 가장 최근 점검 결과이기 때문에 과거의 안전등급 변화와 보수보강으로 인한 등급 상승이 고려되지 않았기 때문이다.

5. 결론

국내 도로교량의 노후화는 더욱 심화될 것으로 예상되며 노후화 대응을 위한 교량의 현재 안전성능 진단에는 많은 인력과 예산이 소요된다. 따라서 본 연구는 규칙 기반 분류 기법을 활용하여 도로교량의 안전등급을 추정할 수 있는 분류 모델을 개발하였다. 그 결과 29개의 분류 규칙으로 이루어진 모델을 개발하였으며, 분류 성능은 전체 예측 정확도 84.8%, True Positive Rate 67.3%인 것으로 나타났다. 본 연구에서 제시하는 모델의 안전등급 분류 결과는 안전등급에 기반을 둔 도로교량의 유지·관리 의사결정을 위한 지원 자료로서 활용될 수 있을 것이다.

향후에는 도로교량의 기본체원뿐만 아니라 더 많은 정보와 외부 지표를 활용하여 분류 모델을 개발함으로써 분류 성능을 향상시키고 도로교량의 유지·관리 의사결정자에게 유의미한 교량 안전등급 분류 규칙을 제공할 수 있을 것으로 기대된다. 또한 등급을 분류하는 모델을 통해 두 가지의 등급을 구분하기 위한 규칙을 추출하는 대신 연관규칙(association rules) 분석 기법을 활용하여 안전등급과 직접적으로 관련된 규칙을 추출하는 방법도 고려해볼 수 있을 것이다.

감사의 글

이 논문은 2015년도 서울대학교 유망선도연구자 지원사업의 후원을 받아 수행되었으며, 2015년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No. 2015R1A5A7037372).

References

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

Cattan, J., Mohammadi, J. (1997). Analysis of Bridge Condition Rating Data Using Neural Networks, *Microcomputers in Civil Engineering*, 12(6), pp. 419-429.

Cohen, W. W. (1995). *Fast Effective Rule Induction*, Proceedings of the Twelfth International Conference on Machine Learning, Tahoe City, California, USA.

Han, J., Kamber, M., Pei, J. (2012). *Data Mining : Concepts*

- Journal 38 of KIBIM Vol.6, No.2 (2016). and Techniques, 3rd ed., Morgan Kaufmann.
- Huang, R. Y., Chen, P. F. (2012). Analysis of Influential Factors and Association Rules for Bridge Deck Deterioration with Utilization of National Bridge Inventory, *Journal of Marine Science and Technology*, 20(3), pp. 336–344.
- Hua), Artificial Neural Network Model of Bridge Deterioration, *Journal of Performance of Constructed Facilities*, 24(6), pp. 597–602.
- Kim, Y. J., Yoo, D. W. (2002). Lessons and Analysis of Event in Domestic Bridge Failures, *Journal of Korean Society of Civil Engineers*, 50(8), pp. 34–40.
- Ko, S. S., Park, H. K., Lee, H. C., Yeo, S. K., Shin, K. S., Choi, D. Y. (2011). A Study on Survey and Improvement Plan of Facility Safety Management.
- Korea Infrastructure Safety and Technology Corporation. (2015). Facility Management System, <http://www.fms.or.kr> (Nov. 18, 2015).
- Korea Infrastructure Safety and Technology Corporation (2010). Detailed Guideline of Safety Inspection and Precise Safety Diagnosis.
- Korea Institute of Civil Engineering and Building Technology (2015). Yearbook of Road Bridges and Tunnels (2015).
- Lin, T. K., Lin, C. C. J., Chang, K. C. (2002). A Neural Network Based Methodology for Estimating Bridge Damage after Major Earthquakes, *Journal of the Chinese Institute of Engineers*, 25(4), pp. 415–424.
- Lee, J., Sanmugarasa, K., Blumenstein, M., Loo, Y. C. (2008). Improving the Reliability of a Bridge Management System (BMS) Using an ANN-based Backward Prediction Model (BPM), *Automation in Construction*, 17(6), pp. 758–772.
- Ministry of Land, Infrastructure and Transport of Korea (2015.8.11.). Special Act on the Safety Control of Public Structures.
- Ministry of Land, Infrastructure, and Transport of Korea (2007). The Second Basic Plan for Safety Management and Maintenance of Public Structures.
- Oh, B. H., Lee, M. K., Jeong, B. S., Lee, W. P., Kim, W. S., Choi, Y. C., Jang, S. Y., Hong, K. O., Kim, S. I., Kim, M. S., Lee, H. H., Seok, J. S., Lee, J. H., Lim, S. N., Kim, K. H., Lim, H. T., Shin, D. H. (2002). A Study on Developing an Expert System for Predicting Remaining Life of Concrete Decks of Road Bridges.
- Park, S. H. (2014). Large Unbalanced Data Classification Based on Hadoop for Prediction of Traffic Accidents, Masters Thesis, Konkuk University.
- Park, S. H. (2004). Survey on Road Bridges Safety Management and Improvement Plan, Korea Infrastructure Safety and Technology Corporation, *Journal of Facility Safety*, 13, pp. 124–130.
- Wang, S., Yao, X. (2012). Multiclass Imbalance Problems: Analysis and Potential Solutions, *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics*, 42(4), pp. 1119–1130.