

# 건설재해 사전 예측을 위한 부스팅 알고리즘 적용

## Application of Boosting Algorithm to Construction Accident Prediction

조 예 림\*                      신 윤 석\*\*                      김 광 희\*\*\*  
 Cho, Ye-Rim                      Shin, Yoon-Seok                      Kim, Gwang-Hee

### Abstract

Although various research is being carried out to prevent the construction accidents, the number of victims of construction site is increasing continuously. Therefore, the purpose of this study is construction accidents prediction applying the boosting algorithm to the construction domains. Boosting algorithm was applied to construct construction accident prediction model and application of the model was examined using actual accident cases. It is possible to support safety manager to manage and prevent accidents in priority using the model.

키 워 드 : 건설재해, 부스팅 알고리즘  
 Keywords : construction accident, boosting algorithm

## 1. 서 론

### 1.1 연구의 배경 및 목적

한국노동연구원 연구 자료에 따르면 건설업의 재해자 수는 2005년 15,918명에서 2011년 22,782명으로 지속적으로 증가하였다<sup>1)</sup>. 건설업의 재해 예방을 위한 연구가 다양하게 진행되고 있으나 안전관리체계 개선이나 재해사례분석을 통한 안전관리방안 개선과 같은 사후분석적인 연구가 주를 이루고 있다. 이러한 연구와 노력에도 불구하고 건설현장의 재해율이 감소하지 않는 것은 현재보다 더 적극적인 재해 예방 방법이 필요하다는 것을 의미한다. 본 연구에서는 최근 인공지능 분야의 기계학습 및 예측모델에서 뛰어난 일반화 능력으로 많은 주목을 받고 있는 부스팅 알고리즘을 건설현장에서 발생 가능한 재해를 사전에 예측하는데 적용해보고자 한다. 본 연구의 결과는 건설현장에서 발생할 가능성이 높은 재해 유형을 미리 예측하여 안전관리자가 해당 재해를 중점적으로 관리 및 예방할 수 있도록 지원이 가능할 것으로 사료된다.

### 1.2 연구의 범위 및 방법

본 연구에서는 부스팅 알고리즘을 이용하여 사전재해 예측 모델을 구축하고, 건설재해 사례 데이터를 이용하여 적용성을 검토하는 것까지를 연구의 범위로 제한하였다. 부스팅 알고리즘 모델구축을 위하여 데이터마이닝 상용패키지인 'Statistica Release 7' 을 이용하였다. 건설재해 사례 데이터는 2014년 1년간 국내에서 발생한 건설재해 정보를 수집하였다.

## 2. 부스팅 알고리즘의 개념

일반적으로 어떤 문제의 결과를 예측할 수 있는 컴퓨터 프로그램을 개발하고자 하는 경우 프로그램 개발자는 경험이 많은 전문가로부터 얻을 수 있는 규칙(rule)에 관한 정보를 우선적으로 수집하게 된다. 만약 수집된 규칙이 주먹구구식(rule-of-thumb) 방법이라 개략적이고 부정확하더라도 무작위적인 선택보다는 나은 것이다. 부스팅 알고리즘은 이렇게 개략적이고 부정확할 수 있는 규칙들을 하나로 조합함으로써 보다 정확한 예측을 만들어낼 수 있다고 입증된 방법이다<sup>2)</sup>. 부스팅 알고리즘은 영향력 있는 약 분류기(weak classifier)의 집합을 이용하여 하나의 강 분류기(strong classifier)를 구축하는 단순 학습 알고리즘이다. 이 알고리즘에서는 결합되었을 때 분류 성능이 크게 증가할 수 있는 약 분류기들을 선택하고 이를 이용하여 하나의 강 분류기를 구축함으로써 새로운 데이터가 주어졌을 때 그 분류결과를 예측할 수 있도록 하는 것이다<sup>2)</sup>.

## 3. 건설재해 사전 예측을 위한 부스팅 알고리즘 적용

### 3.1 데이터 수집 및 변수 선택

건설재해 예측 모델 구축을 위하여 2014년 1년 동안 국내에서 발생한 건설안전사고 사례 2만여 건을 수집하였다. 본 연구에서 수집한 사례

\* 경기대학교 건축공학과 석사과정

\*\* 경기대학교 플랜트·건축공학과 부교수, 공학박사, 교신저자(shinys@kyonggi.ac.kr)

\*\*\* 경기대학교 플랜트·건축공학과 교수, 공학박사

중 전체 16,248건의 안전사고 중 추락(32.9%), 전도(14.7%), 맞음(13.3%), 절단(9.5%), 부딪힘(8.7%), 끼임(8.3%), 깔림(5.2%), 불균형 및 무리한 동작(2.7%), 무너짐(2.5%) 순으로 발생형태가 높은 것으로 나타났다. 특히, 추락부터 끼임까지 6가지 안전사고 유형이 전체 안전사고 발생건수의 87.4%를 차지하고 있다. 따라서 본 연구에서는 출력변수인 건설안전사고 발생형태는 6개의 주요 안전사고 유형을 대상으로 하였다.

### 3.2 부스팅 알고리즘 적용 방법

부스팅 알고리즘을 이용한 건설재해 사전 예측 모델의 구축을 위한 세부 적용 절차는 다음과 같다. 첫 번째, 학습용 데이터를 이용하여 재해유형별로 강 분류기를 학습시킨다. 각각의 학습과정마다 학습용 데이터에는 각 사례별 현장 조건들이  $x_i$ 에 할당되고 재해유형은 '0' 또는 '1'의 값을 갖는  $y_i$ 가 할당된다. 두 번째, 강 분류기의 학습이 끝나고, 새로운 사례에 대해 각각의 강 분류기 별로 분류 점수(정 값을 갖는  $h(x)$ 의  $\alpha$ 값의 합 / 전체  $\alpha$ 값의 합)를 산정한다. 산정값이 크다는 것은 이 사례가 다른 강 분류기에서 보다 더 잘 분류된다는 것을 의미한다. 세 번째, 새로운 사례를 가장 높은 분류 점수를 갖는 재해유형 범주로 구분해준다. 이 과정을 도식화한 것이 그림 1이다.

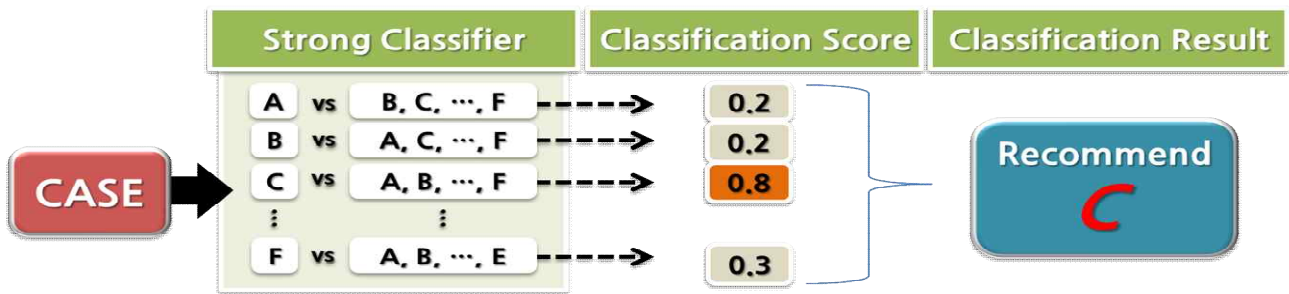


그림 1. 알고리즘 도식화

건설안전사고 사례 데이터를 적용한 결과, 예측한 값의 누적예측정확도가 1순위 40%, 1-2순위 70%, 1-3순위 80%로 높은 정확도를 보였다. 건설재해 사전 예측 모델 적용 결과는 표 1과 같다.

표 1. 건설재해 사전 예측 적용 결과

목표값	예측값(1순위)	예측값(2순위)	예측값(3순위)
찢림	찢림	추락	전도
추락	절단	추락	부딪힘
추락	절단	추락	찢림
찢림	찢림	부딪힘	맞음
전도	찢림	부딪힘	절단
맞음	맞음	부딪힘	전도
맞음	찢림	추락	부딪힘
추락	절단	추락	찢림
추락	절단	찢림	추락
전도	전도	추락	맞음
<b>누적예측정확도</b>	1순위 : 40% (4/10)	1-2순위 : 70% (7/10)	1-3순위 : 80% (8/10)

## 4. 결 론

건설재해예방을 위해 다양한 연구가 진행되고 있으나 건설현장의 재해율은 줄어들지 않고 있다. 본 연구에서는 최근 많은 주목을 받고 있는 부스팅 알고리즘을 현장에서 발생 가능한 재해 유형을 예측하는데 적용하여 더욱 적극적으로 재해를 관리 및 예방할 수 있는 모델을 제안하고 적용성을 검토해보았다. 건설안전사고 사례 데이터의 적용 결과 목표 값에 대해 80%라는 높은 예측정확도를 보였다. 향후 연구에서는 모델의 활용도 향상을 위하여 변수선택에 대한 상세한 연구가 필요할 것이다.

### 참 고 문 헌

1. 배규식, 윤조덕, 안홍섭, 심규범, 건설업 산업재해 현황분석 및 정책방향: 중소기업실업장을 중심으로, 한국노동연구원, 노동복지 및 노동보험 연구, 2013.11
2. Freund Y, Schapire RE, A short introduction to boosting, Journal of Japanese Society for Artificial Intelligence, Vol.14, Issue 5, pp.771~80, 1999.9