

의사결정나무기법을 이용한 건설재해 사전 예측모델 개발

Prediction Model of Construction Safety Accidents using Decision Tree Technique

조 예 림¹

김 연 철²

신 윤 석^{3*}

Cho, Yerim¹

Kim, Yeon-Choel²

Shin, Yoonseok^{3*}

Department of Architectural Engineering, Graduate School, Kyonggi University, Suwon-Si, Gyeonggi-Do, 16227, Korea ¹

Graduate School of Construction Industry, Kyonggi University, Suwon-Si, Gyeonggi-Do, 16227, Korea ²

Department of Plant-Architectural Engineering, Kyonggi University, Suwon-Si, Gyeonggi-Do, 16227, Korea ³

Abstract

Over the past 7 years, the number of victims of construction disasters has been gradually increasing. Compared with projects in other industries, construction projects are highly exposed to safety risks. For this reason, the research methods of predicting and managing the risk of construction disasters are urgently needed that can be applied to a construction site. This study aims to propose a prediction model for a construction disaster using the decision tree technique. The developed the model is reviewed the applicability by evaluating its accuracy based on disaster data. The top three of the prediction values obtained from the proposed model were enumerated, and then the cumulative accuracy were also calculated. The prediction accuracy was 40 percent for the first value, but the cumulative accuracy was 80 percent. Thus, as more disaster data was accumulated, the cumulative accuracy appeared to be higher. If utilized in construction sites, the model proposed in this study would contribute to a reduction in the rate of construction disasters.

Keywords : decision tree, construction safety accidents, prediction model

1. 서 론

1.1 연구의 배경 및 목적

고용노동부 산업재해 현황분석에 따르면 최근 7년간 건설업의 재해자 수가 2009년 20,998명에서 2015년 25,132명으로 꾸준히 증가한 것으로 나타났다[1,2,3,4]. 산업 재해의 예방을 위한 연구와 부단한 노력에도 불구하고 건설업의 재해율은 줄어들지 않고 있다. 건설공사가 고층화·대형화되고 있고 문화와 언어가 다른 외국인 근로자의 수가 증가함에

따라 건설업의 재해율은 앞으로도 쉽게 감소하지는 않을 것으로 판단된다[5].

건설프로젝트의 특성상 옥외생산, 고소작업, 건설 중장비 작업 등으로 타 산업군에 비해 재해가 빈발하고 중대재해가 발생할 가능성이 매우 크다. 실제로 2015년을 기준으로 산업별 업무상사고 사망재해자 수는 전체 사망자수 955명 중 건설업이 437명(45.76%)으로 가장 많았고, 그 다음으로 제조업 251명(26.28%), 기타 산업 149명(15.60%), 운수·창고·통신업 88명(9.21%), 임업 15명(1.57%) 순으로 높은 것으로 나타났다[4]. 이러한 건설 재해는 직접적인 경제적 손실뿐 아니라 국가신인도의 저하 및 사회적 불안감을 고조시킬 수 있어 국가적인 부담 요인으로 작용할 수 있다.

건설업 분야의 재해율 감소를 위한 연구들이 다양하게 진행되고 있다. 하지만 기존 연구에서는 재해사례를 사후에

Received : January 31, 2017

Revision received : February 28, 2017

Accepted : March 27, 2017

* Corresponding author : Shin, Yoonseok

[Tel: 82-31-249-9721, E-mail: shinys@kyonggi.ac.kr]

©2017 The Korea Institute of Building Construction, All rights reserved.

분석하여 원인을 규명하는 데에 그치거나 특정 현장 및 사고에만 대상이 한정되어 있다는 한계가 있다[6, 7, 8]. 그렇기 때문에 다양한 건설 현장에서 적용이 가능하며 적극적으로 사전에 건설재해를 예측 및 관리할 수 있는 연구가 필요하다.

건설재해를 사전에 예측하는데 있어서 데이터마이닝 기법의 활용이 하나의 대안이 될 수 있다. 데이터마이닝 기법 중 하나인 의사결정나무(decision tree)는 분류(classification), 점수화(scoring), 예측(prediction)에서 자주 사용되는 기법 중 하나이다. 의사결정나무의 큰 장점 중 하나는 분석과정이 나무구조 형태의 규칙으로 표현되기 때문에 분류 또는 예측을 목적으로 하는 판별분석, 회귀분석, 인공신경망과 같은 다른 기법들에 비해서 연구자가 분석과정을 쉽게 이해하고 설명할 수 있다는 점이다[9]. 특히, 의사결정나무는 나무구조로부터 목표변수를 설명하는데 있어서 어떤 예측변수가 더 중요한지를 쉽게 파악할 수 있다[10]. 두 개 이상의 변수가 결합하여 목표변수에 어떤 영향을 주는지 그 상호작용(interaction) 효과의 해석이 용이하다.

의사결정나무의 이런 장점은 다른 데이터마이닝 기법보다 건설재해 예측모델 구축에 더 적합하다고 판단된다. 종종 모델의 정확도와 투명도(transparency)간에는 상호교환(trade-off)이 발생한다[11]. 즉, 건설 재해 예측모델을 이용하여 건설현장에서 안전관리를 수행할 때, 안전 관리자에게는 모델의 정확도와 투명도의 2가지 측면에 있어서 발생 가능한 재해에 대한 예측의 정확도보다 해당 재해가 발생하게 되는 원리나 이유를 파악하는 것이 더 중요하다. 따라서 본 연구에서는 모형에 대한 해석이 쉽고 변수의 상호작용 이해가 용이한 의사결정나무 기법을 이용하여 건설현장에서 발생할 수 있는 재해를 사전에 예측할 수 있는 모델을 개발하고자 한다. 의사결정나무기법을 적용하여 발생 가능한 건설재해와 원인을 사전에 파악하여 안전 관리자가 효율적인 안전관리 업무수행을 함으로써 건설현장의 재해 감소에 기여할 수 있기를 기대한다.

1.2 연구의 범위 및 방법

본 연구에서는 데이터마이닝 기법 중 하나인 의사결정나무 기법을 이용하여 건설현장에서 일어날 수 있는 재해를 사전에 예측할 수 있는 모델을 구축하고, 실제 건설재해 사례 데이터를 적용함으로써 모델의 예측 정확도를 평가하여 건설현장에서의 활용 가능성을 검증하는 것까지를 연구의 범위로 제한한다.

연구의 방법은 국내 건설 현장에서 발생한 재해 현황에 대해 조사하고, 의사결정나무 기법 및 건설현장 산업재해 예방에 관련된 기존 문헌들을 통하여 기법의 이론과 적용 방법에 대해 고찰한다. 모델 구축을 위한 도구는 통계 프로그램 ‘IBM SPSS Statistics 19’을 사용하였다. 모델 구축 및 검증을 위하여 2014년 1년간 발생한 건설 현장 재해 사례를 수집하였고 13,232건의 사례를 데이터베이스화한 후 활용하였다. 구축된 모델의 적용성을 검증하기 위해서 동일한 데이터와 프로그램을 이용하여 판별분석 모델을 구축하고 그 결과를 비교해보았다. 연구의 절차를 정리하면 Figure 1과 같다.

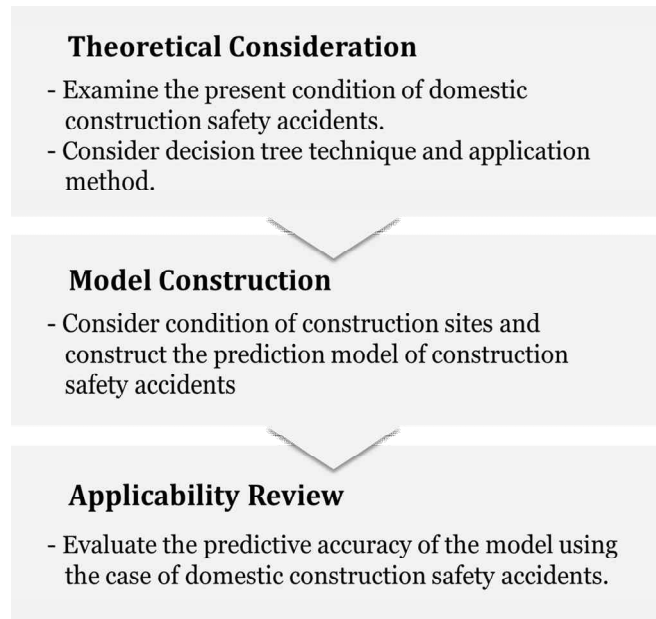


Figure 1. Research methodology

2. 이론적 고찰

2.1 국내 건설 현장 재해 발생 현황

최근 7년 동안 우리나라 건설 현장에서 발생한 산업재해 현황을 살펴보았다. 재해 발생에 따른 재해자 수는 Table 1과 같이 총 162,034명이며, 2009년 20,998명에서 2015년 25,132명으로 매년 꾸준히 증가한 것을 알 수 있다. 연천 인율 역시 2009년 6.55%에서 2015년 7.48%로 증가하였다 [1, 2, 3, 4].

건설 현장 재해 발생 현황을 재해유형별로 살펴본 결과 2015년을 기준으로 총 25,132건의 재해 중 ‘떨어짐’이

8,259건으로 빈도가 가장 높았고 그 다음으로 ‘넘어짐’이 3,594건, ‘맞음’이 3,168건, ‘절단·베임·찢림’이 2,625건, ‘부딪힘’이 2,219건, ‘끼임’이 1,958건, ‘깔림·뒤집힘’이 766건, ‘무리한 동작’이 632건, ‘무너짐’이 327건 순서로 발생 빈도가 높은 것으로 나타났다. 또한, 산업별 업무상사고 사망자 수는 전체 사망자수 955명 중 건설업이 437명으로 이는 45.76%, 거의 절반에 가까운 수치이며 국내 건설 재해 현황의 심각성을 확인할 수 있었다[4].

Table 1. The number of disaster victim in domestic construction industry

Year	Disaster victim (Persons)	The number of accident per year in thousands (%)
Total	162,034	-
2015	25,132	7.48
2014	23,669	7.28
2013	23,600	9.19
2012	23,349	8.38
2011	22,782	7.38
2010	22,504	7.03
2009	20,998	6.55

2.2 의사결정나무(Decision Tree)의 개요

의사결정나무는 데이터마이닝 기법 중 하나로써, 의사결정규칙(decision rule)을 Figure 2와 같이 나무구조로 도표화하여 특정 목표변수에 대해 관심대상이 되는 모집단을 몇 개의 일원적인 소집단들로 분류하고 예측하는 데에 활용되는 기법이다[9,12].

의사결정나무기법은 이해가 쉽고 명확하여 해석이 용이하며 탐색(exploration)과 모형화(modeling)의 두 가지 특성을 모두 지니고 있다. 판별분석과 회귀분석 등과 같은 모수적(parametric) 모형을 분석하기 위해 사전에 이상치(outlier)를 검색하거나 분석에 필요한 변수를 찾아내고 둘 이상의 변수가 결합하여 목표변수에 어떤 영향을 주는지에 대한 해석을 하는 데에 사용될 수 있다[9,11].

의사결정나무는 그 자체가 분류 또는 예측 모형으로 사용될 수 있으며 분석과정의 나무구조로 표현되어 다른 데이터마이닝 기법들에 비하여 간단하게 이해하고 설명할 수 있다는 장점이 있다. 또한, 전문가가 아니어도 입력변수와 출력변수를 쉽게 조정할 수 있고, 모형에 포함되어야 할 교호효

과(interaction)를 찾아내는 것이 용이하다. 이러한 의사결정나무의 장점은 본 연구의 건설 재해 예측 모델 구축에 적합하다고 판단된다.

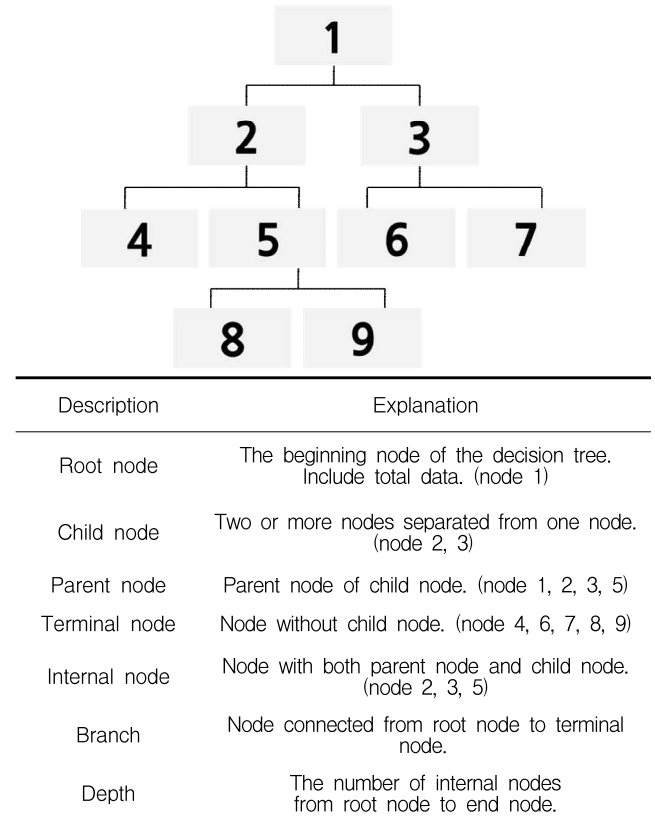


Figure 2. Concept of the decision tree

2.3 기존 연구 분석

건설업 분야에서 데이터마이닝 기법을 접목시켜 건설재해율의 감소를 도모할 수 있도록 연구한 사례를 살펴보았다. Lee and Lee[7]는 데이터마이닝을 이용하여 건설 사고의 연관규칙을 분석하여 건설재해의 위험성을 평가하였다. 하지만 이 연구에서는 재해통계자료를 단순 분석하여 사고들 간 관련성에 대한 규칙을 알아내는 데에 그쳤다. Kim[8]는 데이터마이닝 기법을 활용하여 국내의 소규모 건설현장 재해 중 추락으로 인해 가장 많은 사망사고가 발생한다는 결과를 도출하여 추락 예방 보호구 착용을 향상을 위한 방안을 제시하였다. 그러나 이 연구의 경우 제도적 개선 대상을 소규모 건설현장 및 추락 사고에 한정하여 제시하였다는 한계점이 있다. Shin and Yoo[6]는 사례기반추론 기법을 활용하여 건설현장에서 일어날 수 있는 안전사고의 조기 경보 모델을 구축하기 위한 연구를 하였다. 그러나 사례기반추론

기본 모델은 초보자가 작업하기에 어렵다는 단점이 있다. 따라서 본 연구에서는 누구나 쉽게 이해하고 활용할 수 있는 의사결정나무 기법을 활용하여 건설재해를 사전에 예측할 수 있는 모델을 구축하고자 한다.

3. 건설재해 사전 예측 모델 구축

3.1 데이터 수집 및 변수 선택

의사결정나무 기법을 이용한 건설재해 예측 모델을 구축하기 위해 먼저 변수들을 선정하여야 한다. 건설재해라는 출력변수(종속변수)를 얻기 위해서는 입력변수(독립변수)들을 적절하게 선정하여야 건설재해에 대한 예측의 정확도를 높일 수 있다.

종속변수는 앞서 언급한 빈도 높은 재해유형 중 6개의 주요 유형들을 선택하여 대상으로 정의하였다. 독립변수는 실제로 건설재해 발생에 영향을 끼치는 변수들로 설정해야 하기 때문에 국내에서 발생했던 건설재해 데이터들을 수집한 후 재해 예측 관련 전문가들에게 인터뷰 조사를 하여 총 15개의 독립변수를 선택하였다. 본 연구에서 설정한 변수는 Table 2와 같다.

3.2 의사결정나무 모델 구축

의사결정나무 모델의 구축을 위해서는 매개변수와 옵션을 설정하는 작업이 필요하다. 가장 먼저, 종속변수(output)와 독립변수(input)를 입력한다. 종속변수는 명목형 변수인 재해의 발생 형태(type of accident)를 입력하였다. 재해의 발생 형태는 발생 빈도가 높은 재해유형 중 떨어짐(fall), 넘어짐(wipeout), 맞음(hit), 절단(cut), 끼임(narrowness), 부딪힘(crash) 등 6개의 주요 안전사고 유형을 대상으로 한 것이다. 독립변수는 현장의 조건들을 입력한 것으로 재해자의 성별(gender), 지역(location), 국적(nationality), 재해가 일어난 계절(season), 시간(time) 등을 포함한 명목형 변수 9개와 재해자의 나이(age), 프로젝트 기간(project duration), 프로젝트 규모(project scale), 공정률(rate of process), 노동자 수(workforce), 노동자의 회사 근무 기간(year of service) 등 수치형 변수 6개로 총 15개의 변수로 구성하였다.

다음으로는 나무구조의 확장 방법을 선택해야 한다. 의사결정나무구조의 확장 방법으로는 'ID3, C4.5, C5.0, CART, CHAID' 등의 알고리즘이 있으며, 본 논문에서는 'CHAID' 알고리즘 방식을 선택하였다. CHAID(Chi-

Table 2. Parameter setting

Variable		Type	Feature		
Dependent variable (output)	Type of accident	Nominal	Fall(1), Wipeout(2), Hit(3), Cut(4), Narrowness(5), Crash(6)		
	Gender		Male(1), Female(2)		
	Day of the week		Weekdays(1), Weekend(2)		
	Location		Seoul(1), Capital area(except for Seoul)(2), Etc.(3)		
	Nationality		Korean(1), Foreigner(2)		
	Season		Spring(1), Summer(2), Autumn(3), Winter(4)		
	Time		06-12hour(1), 12-18hour(2), 18-22hour(3), 22-06hour(4)		
	Independent variable (input)		Type of employment	Regular(1), Non-regular(2)	
			Type of occupation	Management(1), Engineering(2), Labor(3), Operation(4)	
			Type of project	Residential and commercial(1), Public and Healthy(2), Etc.(3)	
			Age	Integer(17-83)	
			Project duration	Integer(1-13,752)	
			Project scale	Numerical	Integer(0.1-2,000)
			Rate of process		Real number(5-90)
			Workforce		Integer(5-3,000)
Year of service		Real number(0.8-30)			

squared Automatic Interaction Detection)는 기존의 AID(Automatic Interaction Detection)에 기원을 두어 더 많은 양의 범주형 자료를 분석할 수 있도록 개발한 것이다 [13]. AID는 나무의 가치를 두 개로 쪼개는 이지 분리(binary split)만 가능했으나, CHAID는 카이제곱-검정(이산형 목표변수) 또는 F-검정(연속형 목표변수)을 이용하여 둘 이상의 다지 분리(multiway split)를 수행할 수 있다[14].

다음으로 의사 결정 나무 모델의 기준을 설정해야 한다. '확장 한계'에서 뿌리 노드(node)를 최대 몇 단계까지 확장

을 할 것인지 정하기 위해 최대 트리 깊이를 입력해야 하는데, 본 연구에서는 ‘자동’을 선택하였다. 앞서 나무구조 확장 방법을 ‘CHAID’ 알고리즘 방식을 선택했기 때문에 ‘최대 수준 수’는 3이 된다. 부모 노드와 자식 노드의 최소 크기를 지정하는 ‘최소 케이스 수’는 임의로 값을 변경하며 입력해보았으나 큰 변화가 없었다. 따라서 상위 노드 100, 하위 노드 50, 즉, 기본 값으로 설정하였다. 확장 한계를 설정한 후 ‘CHAID’ 단계로 이동한다. 유의성 수준이 작을수록 노드 분할에서는 단순한 트리가 생성되고 유의성 수준이 클수록 범주 합치기에서는 범주 간의 병합이 억제된다. 카이제곱 통계량은 Pearson을 선택 후 ‘구간’ 단계로 넘어간다. ‘구간’ 단계에서는 척도 독립변수 구간을 지정해야 한다. 본 연구에서는 독립변수들에 대한 척도 독립변수 구간을 고정 숫자로 설정했으며, 구간 값도 ‘최소 케이스 수’의 경우와 같은 이유로 기본 값 10으로 하였다. 모델 구축을 위한 매개 변수 및 옵션 설정 값은 Table 3과 같다.

Table 3. Existing research

Description		Contents	
Extention limit	Maximum tree depth	Automatic	
	The number of minimum cases	High rank node	100 (Default value)
		Low rank node	50 (Default value)
	Level of significance	Node division	0.05
		Category combine	0.05
CHAID	Model estimation	Maximum recursive calculation	100
		Maximum change of cell expected frequency	0.001
Section	Scale independent variable section	Fixed number	10 (Default value)

의사결정나무 모델은 예측 값의 1순위, 2순위, 3순위 및 각각의 발생 확률의 확인이 가능하며 결과도출과정을 직관적으로 파악할 수 있다는 장점이 있다. 구축된 모델인 Figure 3을 예로 들어, 재해가 일어난 계절(season)이 겨울(4)일 경우 노드 1로 이동하며, 봄(1), 여름(2), 가을(3)일 경우 노드 2로 이동한다. 그 다음으로 프로젝트 종류(type of project)의 경우 주거 및 상업용(1), 공공시설(2), 기타(3) 등 변수에 따라 그 다음 노드로 이동하는 것을 한눈에 확인할 수 있다.

4. 의사결정나무 모델의 사례 적용

4.1 예측 정확도 평가

의사결정나무 기법을 이용하여 구축한 모델의 활용 가능성을 검토하기 위하여 통계적 분석방법인 판별분석(Discriminant Analysis) 기법을 이용한 모델과의 예측 정확도를 비교하였다. 판별분석은 집단 간의 차이를 규명해주는 독립변수를 찾아내고 이들의 선형결합(Linear Combination)으로 판별식을 만들어 분류하고자 하는 대상들이 속하는 집단을 찾아내는 방법이다[15]. 판별분석 적용 시 어떤 독립변수를 설정하느냐에 따라 결과 값이 달라질 수 있다. 본 연구에서는, 판별함수의 계수를 결정하는 방법으로 많이 사용되고 있는 ‘Fisher’의 ‘선형판별분석법’을 사용하였다. 판별분석을 실시한 후 도출된 판별함수가 통계적으로 의미가 있는지를 판단하기 위해 ‘Wilks’ Lambda’ 값을 확인하였다. ‘Wilks’ Lambda’의 유의확률 값이 가장 큰 경우가 ‘0.009’로 유의수준(0.05)보다 작으므로 해당 판별함수에 의한 집단 간의 판별 점수의 차이는 통계적으로 의미가 있다고 볼 수 있다[16]. 결과적으로 6개의 종속변수에 대한 판별함수의 계수를 얻을 수 있었다. 모델 예측 정확도 평가를 위해 사용된 데이터는 2014년 1년간 국내 건설현장에서 발생한 재해 사례 데이터 13,232건 중 무작위로 40개를 선택하여 적용하였다. 종속변수 및 독립변수는 의사결정나무 기법을 이용한 모델과 동일하게 적용하였다. 예측 값은 1순위, 2순위, 3순위까지 Table 4와 같이 정리하여 누적 예측 정확도를 구하였다. 의사결정나무 기법과 판별분석 기법을 활용한 모델들의 예측 정확도를 비교한 결과 의사결정나무 기법 기반 모델의 정확도가 더 높은 것으로 나타났다.

4.2 결과 분석

의사결정나무와 판별분석 두 모델의 예측 정확도를 비교한 결과 1순위부터 1+2순위, 1+2+3순위까지의 누적 예측 정확도까지 확인하였을 때 의사결정나무 모델이 더 정확한 것으로 나타났다. 특히, 3순위까지의 누적 예측 정확도는 의사결정나무가 80%, 판별분석이 63%로 정확도가 더 높은 것을 확인할 수 있었다. 예측모델에서 우선순위가 높게 나온 것은 발생 확률이 높은 건설재해 유형을 의미한다. 따라서 본 연구에서 제안된 예측 모델을 통해 위와 같이 1순위 예측 값부터 누적 예측 값을 함께 활용한다면, 각 건설현장의 조건에 따라 발생할 가능성이 높은 건설재해를 사전에 예측해냄으로써 안전관리자의 합리적인 재해 예방 활동이 가능할

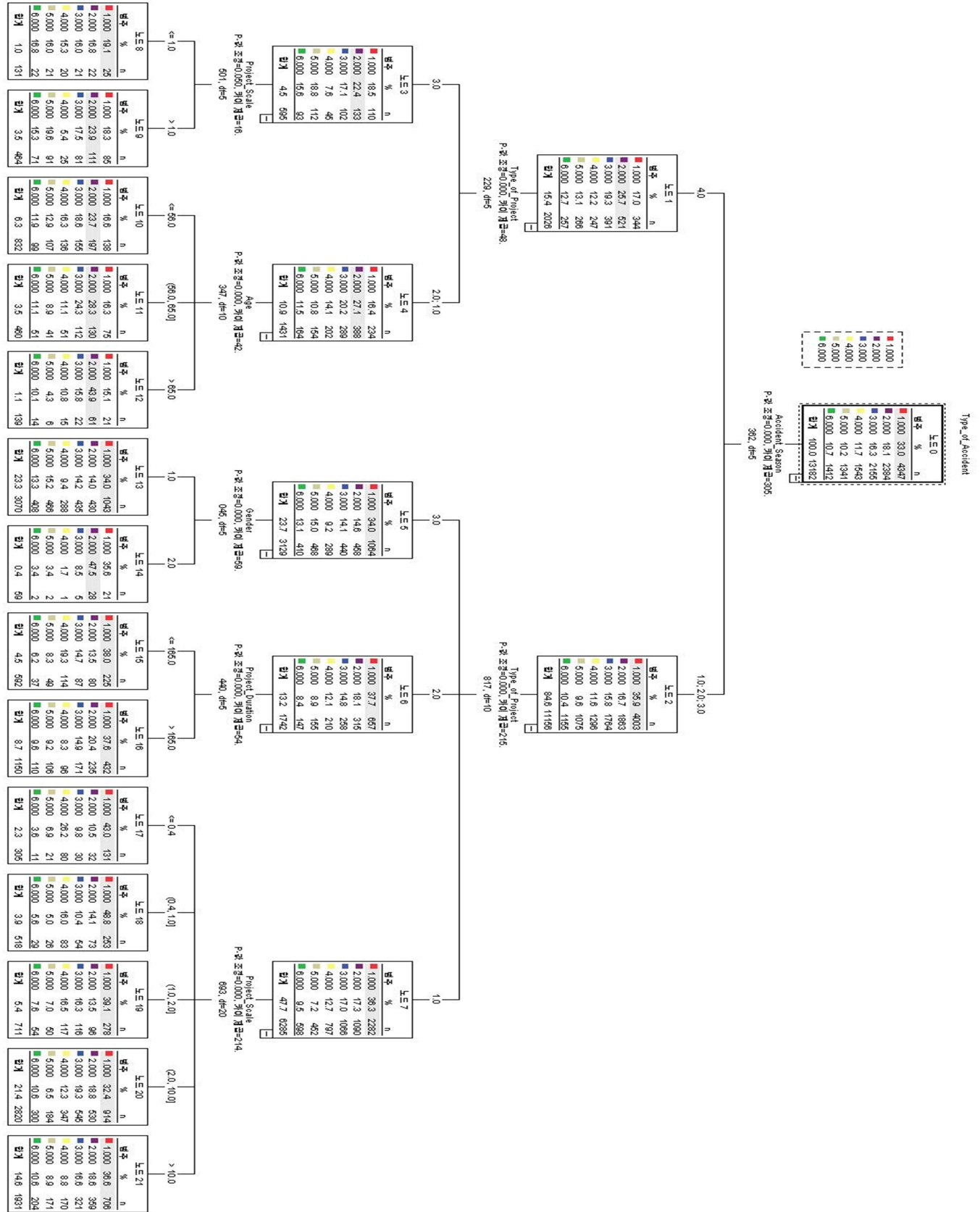


Figure 3. Decision tree model

Table 4. Comparison of prediction accuracy of decision tree and discriminant analysis model

Target	Decision tree model			Discriminant analysis model		
	1st	2nd	3rd	1st	2nd	3rd
Fall	Fall	Wipeout	Hit	Hit	Crash	Cut
Fall	Fall	Narrowness	Wipeout	Narrowness	Crash	Wipeout
Fall	Wipeout	Hit	Fall	Cut	Wipeout	Hit
Fall	Fall	Wipeout	Hit	Wipeout	Fall	Narrowness
Fall	Wipeout	Fall	Hit	Narrowness	Wipeout	Hit
Fall	Wipeout	Hit	Fall	Cut	Narrowness	Crash
Fall	Fall	Wipeout	Hit	Narrowness	Crash	Cut
Fall	Fall	Wipeout	Hit	Crash	Narrowness	Wipeout
Fall	Fall	Wipeout	Hit	Narrowness	Crash	Wipeout
Fall	Fall	Narrowness	Wipeout	Narrowness	Crash	Wipeout
Fall	Fall	Cut	Wipeout	Hit	Cut	Wipeout
Fall	Wipeout	Fall	Hit	Narrowness	Crash	Hit
Fall	Fall	Wipeout	Hit	Wipeout	Cut	Hit
Fall	Fall	Hit	Cut	Wipeout	Hit	Cut
Wipeout	Wipeout	Hit	Fall	Wipeout	Hit	Cut
Wipeout	Wipeout	Hit	Fall	Wipeout	Cut	Hit
Wipeout	Wipeout	Hit	Fall	Wipeout	Crash	Narrowness
Wipeout	Wipeout	Fall	Hit	Narrowness	Wipeout	Crash
Wipeout	Wipeout	Hit	Fall	Wipeout	Hit	Cut
Wipeout	Wipeout	Fall	Hit	Narrowness	Crash	Cut
Wipeout	Fall	Wipeout	Hit	Hit	Fall	Wipeout
Hit	Wipeout	Hit	Fall	Cut	Wipeout	Hit
Hit	Wipeout	Hit	Fall	Wipeout	Hit	Cut
Hit	Wipeout	Hit	Fall	Cut	Hit	Wipeout
Hit	Wipeout	Hit	Fall	Wipeout	Hit	Crash
Hit	Wipeout	Hit	Fall	Wipeout	Hit	Cut
Hit	Fall	Hit	Wipeout	Narrowness	Wipeout	Crash
Cut	Wipeout	Hit	Fall	Wipeout	Cut	Hit
Cut	Wipeout	Hit	Fall	Wipeout	Cut	Hit
Cut	Fall	Cut	Hit	Crash	Cut	Narrowness
Cut	Wipeout	Hit	Fall	Wipeout	Hit	Cut
Cut	Fall	Cut	Wipeout	Wipeout	Hit	Cut
Narrowness	Wipeout	Narrowness	Fall	Narrowness	Crash	Wipeout
Narrowness	Fall	Wipeout	Crash	Narrowness	Crash	Cut
Narrowness	Fall	Narrowness	Wipeout	Narrowness	Crash	Wipeout
Narrowness	Fall	Wipeout	Crash	Wipeout	Narrowness	Crash
Crash	Fall	Wipeout	Crash	Narrowness	Crash	Wipeout
Crash	Wipeout	Hit	Fall	Cut	Crash	Hit
Crash	Fall	Wipeout	Hit	Hit	Crash	Cut
Crash	Wipeout	Fall	Narrowness	Narrowness	Crash	Wipeout
Accumulative accuracy	1st	1st+2nd	1st+2nd+3rd	1st	1st+2nd	1st+2nd+3rd
	40% (16/40)	73% (29/40)	80% (32/40)	17.5% (7/40)	53% (21/40)	63% (25/40)

것이다. 데이터마이닝 기법을 이용한 예측 모델은 데이터의 수가 증가할수록 예측 정확도 또한 높아지기 때문에 시간이 지남에 따라 예측 모델의 활용도 역시 높아질 것으로 예상된다. 또한, 본 연구에서 제안된 의사결정나무기법 기반 건설 재해 예측모델이 건설현장에서 효과적으로 안전 관리가 될 수 있도록 돕는 도구로 사용될 것이라고 사료된다.

5. 결 론

본 연구에서는 건설 현장에서 일어날 수 있는 산업 재해를 사전에 예측하여 근본적인 예방을 도모할 수 있는 예측 모델을 제안하였다. 모형에 대한 해석이 쉽고 변수의 상호작용 효과 해석이 용이한 의사결정나무 기법을 활용하여 건설재해 예측 모델을 제안하였다. 재해 예측 모델 구축을 위해 2014년 1년간 국내 건설 현장에서 발생한 재해 사례를 수집하여 데이터베이스화한 후 13,232건의 데이터를 얻었다. 수집된 재해 데이터에 대하여 재해 예측 관련 전문가들에게 자문을 구해 모델 구축에 사용할 변수를 선택했다. 건설 현장 조건을 고려해 발생 빈도가 높은 유형의 재해를 예측할 수 있는 모델을 구축하고, 실제 건설 재해 사례 데이터를 이용하여 모델의 예측 정확도를 평가하여 실제로 건설현장에서의 활용 가능성을 검증하였다. 의사결정나무 모델의 현장 활용성을 검토하기 위하여 통계적 분석방법인 판별분석 모델의 재해 예측 정확도와 비교하였다. 1, 2, 3순위로 나누고 누적 예측 결과까지 확인하였을 때, 의사결정나무 모델의 예측 정확도(80%)가 판별분석 모델의 예측 정확도(63%)보다 더 높은 것으로 나타났다.

의사결정나무 모델은 전문가가 아니어도 쉽게 이해할 수 있다는 장점을 가지고 있다. 또한, 시간이 지남에 따라 데이터가 축적되면 모델의 예측 정확도가 더욱 높아질 것이며, 건설 현장에서의 활용성 역시 높아질 것으로 판단된다. 본 연구에서 제안한 모델을 건설현장에서 활용함으로써 발생 가능성이 높은 재해에 대하여 안전 관리자가 집중적으로 관리할 수 있도록 도움을 줄 수 있을 것이며, 이는 건설업 재해율의 감소에도 기여할 수 있을 것이다. 또한, 제안된 모델의 정확도 및 활용성을 더욱 향상시키기 위하여 모델의 구성 변수들에 대해 추가 연구를 통한 개선이 필요할 것으로 판단된다.

요 약

건설 산업 재해 예방을 위한 연구와 노력에도 불구하고 최근 7년간 국내 건설업 재해자 수가 꾸준히 증가했다. 건설 현장에서 발생하는 재해는 다른 산업군에 비해 강도 높은 재해가 발생할 가능성이 크기 때문에 근본적으로 예방할 수 있는 방법이 필요하다. 따라서 본 연구에서는 모형에 대한 해석이 쉽고 변수의 상호작용 효과 해석이 용이한 의사결정나무 기법을 활용하여 건설재해 예측 모델을 제안하였다. 제안된 건설 재해 사전 예측 모델의 현장 활용 가능성을 평가하기 위하여 판별분석기법 기반 모델과의 건설 재해 예측 정확도를 비교하였다. 검토 결과 판별분석 모델에 비해 의사결정나무 모델의 누적 예측 정확도가 더 높은 것으로 나타났다. 의사결정나무 기법을 이용한 모델은 시간이 지남에 따라 데이터가 증가하기 때문에 예측 정확도가 더욱 높아지게 된다. 따라서 본 연구에서 제안된 건설 재해 예측 모델이 건설 현장에서 활용된다면 효과적으로 안전 관리를 할 수 있고, 건설업 재해율 감소에도 기여할 수 있을 것으로 기대한다.

키워드 : 의사결정나무, 건설재해, 예측모델

Acknowledgement

This work was supported by Kyonggi University's Graduate Research Assistantship 2016.

References

1. Ministry of Employment and Labor, 2010 Industrial accident analysis of current situation, Sejong-si (Korea): Ministry of Employment and Labor; 2011, 454 p.
2. Ministry of Employment and Labor, 2011 Industrial accident analysis of current situation, Sejong-si (Korea): Ministry of Employment and Labor; 2012, 439 p.
3. Ministry of Employment and Labor, 2013 Industrial accident analysis of current situation, Sejong-si (Korea): Ministry of Employment and Labor; 2014, 529 p.
4. Ministry of Employment and Labor, 2015 Industrial accident analysis of current situation, Sejong-si (Korea): Ministry of Employment and Labor; 2016, 759 p.
5. An SH, Ryu HW, Park MS, Selecting Key Accident Risk Factors

-
- in Building Construction Projects, *Journal of Korea Institute of Building Construction*, 2011 Feb;11(1):45–50.
6. Shin YS, Yoo WS. Early Warning Model using Case-Based Reasoning for Construction Site Safety Accidents, *Journal of Korean society of hazard mitigation*, 2015 Dec;15(6):27–33.
 7. Lee DG, Lee DE. Data mining in construction fatal accident: association rule mining. *Proceeding of the Association of Architectural Institute of Korea*; 2010 Dec 4; Busan, Korea, Busan (Korea): the Association of Architectural Institute of Korea; 2010. p. 426–30.
 8. Kim JH. The improvement of providing protective equipment preventing a fall in small scale construction site [master's thesis]. [Suwon (Korea)]: Kyunghee University; 2013. 76 p.
 9. Kim TH. Advanced system formwork and construction planning model for tall building construction [dissertation]. [Seoul (Korea)]: Korea University; 2013. 140 p.
 10. Lee J. A study of churn prediction model for long-term customers using data mining technique [master's thesis]. [Seoul (Korea)]: Korea University; 2013. 63 p.
 11. Kim JU, Kim ST. Data mining techniques for marketing, sales, and customer relationship management, 3rd ed, Seoul (Korea): Hankyungsa; 2012. 627 p.
 12. Jeong GY. Multivariate control chart procedure using decision tree learning technique [master's thesis]. [Seoul (Korea)]: Chung-Ang University; 2015. 36 p.
 13. Yoo SD. A study on the stability of non-response imputation using CHAID Algorithm, [master's thesis]. [Seoul (Korea)]: Korea University; 2015. 33 p.
 14. Leem YM, Hwang YS. A Feature Analysis of Industrial Accidents Using CHAID Algorithm, *Journal of Korea Safety Management & Science*, 2005 Dec;7(5):59–67.
 15. Park JJ, Seo IK, Kang DY, Lee JH. A Safety Evaluation of Shoulder Rumble Strips on Expressway using Discriminant Analysis, *Journal of the Korean Society of Safety*, 2016 Feb;31(1):95–104.
 16. Kim YC, Yoo WS, Shin YS. Application of artificial neural networks to prediction of construction safety accidents, *Journal of Korean society of hazard mitigation*, 2017 Feb;17(1):7–14.