



화재시나리오 선정을 위한 머신러닝 기반 인명피해 예측모델에 관한 연구

A Prediction Model of Casualties Based on Machine Learning for Selection of Fire Scenario

서동구* · 박병훈** · 이영현*** · 이원희**** · 김정재***** · 송호준*****

Seo, Donggoo*, Park, Byunghun**, Lee, Younghyun***, Lee, Wonhee****, Kim, Jungjae*****, and Song, Hojun*****

Abstract

This study has developed a model that predicts casualties (dead and injured people) using the Classification And Regression Tree (CART). Based on the fire statistics collected over a decade, this model aims to select the appropriate risk-assessment scenarios and fire prevention and safety methods applicable on individual buildings. Our evaluation indicates that this CART model can accurately predict 48 scenarios based on 5 variables related to the types of fire, fire growth rates, and evacuation situations, and calculate the corresponding probabilities for each occurrence. This model is expected to improve future quantitative fire risk assessments.

Key words : Fire Scenario, Machine Learning, Casualties, Prediction Model, Fire Risk Assessment

요 지

본 연구는 건축물의 화재위험평가에서 중요하게 다루고 있는 화재시나리오 선정방법에 대하여 인명안전에 기인한 선정방법을 제안하기 위해 인명피해(사망자 및 사상자) 예측모델을 10년간 화재통계자료를 기반으로 Classification And Regression Tree (CART)방법을 이용하여 구축하였다. 이 예측모델은 정오행렬에 의해 검증한 결과 정확도, 민감도 및 특이도에서 높은 예측성능을 가지는 것으로 검증되었다. 또한 예측모델을 이용하여 설계화원, 화재성장률 및 피난안전에 관계되는 5개 변수를 선정하여 48개 화재시나리오를 검토한 결과, 변수에 따른 발생확률의 차이를 확인하였고 향후 정량적 화재위험평가에서 활용될 수 있다면 설계의 수준을 높일 수 있을 것으로 기대된다.

핵심용어 : 화재시나리오, 머신러닝, 인명피해, 예측모델, 화재위험평가

*교신저자, 정회원, 한국건설기술연구원 건축연구본부 수석연구원 · 공학박사(Tel: +82-31-910-0622, Fax: +82-31-910-0392, E-mail: donggooseo@kict.re.kr)

Corresponding Author, Member, Ph.D., Senior Researcher, Department of Building Research, KICT

**성균관대학교 일반대학원 산업공학과 스마트품질연구실 박사수료

Ph.D. Candidate, Department of Industrial Engineering, Smart Quality Lab., Sungkyunkwan University

***㈜이레테크 데이터랩스사업부 책임연구원

Senior Researcher, Department of DataLabs, Erectec Inc.

****㈜이레테크 데이터랩스사업부 책임연구원

Senior Researcher, Department of DataLabs, Erectec Inc.

*****㈜이레테크 데이터랩스사업부 책임연구원

Senior Researcher, Department of DataLabs, Erectec Inc.

*****성균관대학교 일반대학원 산업공학과 스마트품질연구실 박사과정

Doctor's Course, Department of Industrial Engineering, Smart Quality Lab., Sungkyunkwan University

1. 서론

1.1 연구배경 및 목적

건축물에 있어서 성능위주설계나 화재위험평가 등에서 주요하게 다루고 있는 것으로써 화재시나리오 선정이 있다. 결정론적 또는 확률론적 방법을 사용하더라도 반드시 화재시나리오를 선정하여야 하고 이를 근거로 화재위험평가를 수행한다. 국내에서 화재시나리오가 법적인 형태로 존재하는 것은 「화재예방, 소방시설 설치·유지 및 안전관리에 관한 법률」 제9조의3 제2항에 근거하여 「소방시설 등의 성능위주설계 방법 및 기준」에 그 시행에 관한 사항으로 7개의 시나리오 중 가장 피해가 클 것으로 예상되는 최소 3개 이상의 시나리오에 대하여 화재예측을 수행하도록 하고 있다.

하지만 가장 화재피해가 클 것으로 예상되는 시나리오를 선정하기 위해서는 근거가 명확해야 되는데 Seo et al. (2017)에 의하면 설계자의 편의에 따라 무난한 화재시나리오 유형으로 편중되는 내용이 지적되는 등 화재시나리오 선정에 명확한 근거가 요구된다고 지적한 바 있다. 이와 같이 화재시나리오의 선정은 화재위험평가를 수행할 때 발화장소나 발화원인 등에 따라 설계화원(design fire), 화재성장률 등이 변화되기 때문에 건축물의 화재상황을 예측하는데 매우 중요한 인자임에는 분명하다(Park et al., 2002; Oh et al., 2003; An et al., 2017).

한편, 화재시나리오를 선정할 경우 통계적으로 가장 발생빈도가 높은 사례를 선정하는 것이 이용되며 이 중에는 소규모 화재나 초기소화 등이 이루어져 인명피해가 발생되지 않은 사례도 포함되게 된다. 대부분의 화재통계를 수집하는 방법은 관할 소방서에서 출동한 사례를 근거로 자동으로 집계되는 것이기 때문에 화재발생빈도가 높다는 것이 화재위험성에 비례하지 않다고 말할 수 있겠다(NFA, 2020). 화재위험평가의 궁극적인 목적은 화재로부터 인간의 생명을 보호하는 것이기 때문에 화재시나리오를 선정할 경우에는 인명피해의 관점에서 시나리오를 결정할 필요가 있다(Hadjisophocleous and Fu, 2004).

이에 본 연구에서는 상기와 같은 배경에 따라 화재통계로부터 얻어진 결과를 인명피해(사망자, 사상자)가 발생한 근거를 기반으로 화재시나리오를 선정할 때 척도가 될 수 있는 예측모델을 구축하여 화재시나리오 선정에 기초적 자료를 제공하는데 그 목적이 있다. 이를 통해 정량적 방법으로 수행되는 화재위험평가가 실제 화재사건에 기초하여 화재시나리오를 선정한다면 설계의 수준을 높일 수 있을 것으로 기대한다.

1.2 연구방법 및 범위

본 연구는 인명피해(사망자, 사상자) 결과를 기반으로 건축물의 화재시나리오 선정에 필요한 예측모델을 구축하

는 것으로 국가화재정보센터(www.nfds.go.kr)에 제공되지 않는 다양한 변수를 파악하기 위해 소방청에 정보공개요청을 통해 10년간 화재사례(2009~2018년)에 대한 세부인자를 분석하였다.

통계자료 분석은 머신러닝 방법을 이용하였으며 Classification And Regression Tree (CART) 방법을 이용하였고, 예측모델의 검증에 대하여 정오행렬(Confusion Matrix)을 사용하여 정확도(Accuracy), 민감도(Sensitivity) 및 특이도(Specificity)를 평가하였다. 또한 예측모델을 이용해 가상의 화재시나리오를 정하여 사망자 및 사상자 발생확률에 대한 결과를 도출함으로써 화재시나리오 선정에 있어서 엔지니어 또는 평가자가 확인할 수 있는 방법에 대해 검토하였다. 본 연구는 건축물에 한정하여 수행한 연구로서 모델을 구축하여 화재시나리오를 선정할 때 새로운 방법을 제시하는 것으로써 개발된 모델은 현장에서 사용이 가능한 수준이 아님을 밝힌다.

2. 화재시나리오 및 머신러닝 관련 선행연구 고찰

2.1 화재시나리오 관련 선행연구

화재시나리오는 전술한 바와 같이 화재위험평가에서 화재성상을 예측하거나 사상자 발생확률 등을 도출할 때 사용되고 있으며, 여기서 발화장소나 설계화원 등을 선정하는 것은 화재성상을 예측하는데 반드시 요구되는 인자로서 대부분 화재발생빈도에 기인하고 있다(Hadjisophocleous and Fu, 2004).

Jang (2016)은 성능위주설계 등 실제 수행된 7개의 사례를 분석하였는데 화원크기, 화재성장속도 및 화원위치에 따라 피난허용시간(available safe escape time, ASET)이 상당한 차이가 있는 것을 검증하였고, 이를 통해 입력데이터의 표준화를 주장하였다. 또한 Jeong et al. (2018)은 성능위주설계 개선방안을 도출하기 위해 전문가 설문조사를 수행하였는데 화재시나리오에 대하여 화원설정과 관련된 상세기준 추가가 필요하다고 기술하였다.

특히, An et al. (2017)은 국내 성능위주설계가 시행된 95개의 사전검토를 대상으로 화재시나리오와 화재시물례 이전에 사용된 입력인자에 대해 분석하였는데 화재발생 장소는 주차장이 24%로 가장 많이 선정되었고, 차 순위로 오피스텔을 포함한 업무시설, 판매시설, 공동주택 등의 순이고, 가연물의 경우에는 발화장소에 따라 다르겠지만 폴리우레탄, 목재 등을 선정하고 있다고 분석되었다.

또한 Kim et al. (2012)은 초고층 주상복합 아파트에서 화재시나리오를 평가할 수 있는 항목에 대하여 전문가 대상 설문조사 결과로 상대지수를 도출하였는데, 화재시나리오를 설정하는데 인자의 중요도에 대하여 화재발생지점, 화재발생빈도가 높은 설계화원, 발생층수, 수직 공간 내·외부 온도차이, 화재발생 시간대, 위험요소, 출입문 개폐여부 등

을 도출하였다. 하지만 전술한 An et al. (2017)의 연구결과에서 볼 수 있듯이 성능위주설계에 활용되는 화재시나리오는 이러한 중요도를 반영하지 못하고 있고 화재발생빈도에 초점을 맞추고 있기 때문에 Seo et al. (2017)가 주장한 무난한 화재시나리오의 선정이 문제가 될 것으로 판단되며, 인명안전 중심의 화재시나리오 선정방법이 필요하다고 할 수 있다.

2.2 머신러닝 방법에 관한 선행연구 및 적용방법

본 연구의 목적은 인명피해가 발생할 수 있는 실제 화재발생 건수를 분석하여 건축물의 용도, 면적, 발화위치 및 발화지점 등 다양한 화재발생 원인 및 결과를 기반으로 화재시나리오를 도출할 수 있는 예측모델을 구축하는 것이다. 데이터의 행과 변수의 수가 많을 경우 회귀분석과 같은 전통적인 통계분석 기법을 적용하는데 이 경우 모형의 성능이 저하될 수 있으며 시나리오를 시각적으로 확인하기 어렵다. 이를 위해 다양한 산업군에서는 예측 및 중요변수 선별을 위해 머신러닝 기법이 많이 사용되고 있다.

의학 분야에서는 폐결핵 의심 환자 예측 및 폐결핵 발생의 주원인을 선별(Aguiar et al., 2012)하였으며, 건설 분야에서는 건설회사의 파산 여부를 예측(Karas and Reznakova, 2017)한 사례가 있다. 본 연구의 주된 목적은 다양한 화재발생 사례를 토대로 화재발생 여부를 정확히 예측하고, 화재발생에 영향을 미치는 주요 변수를 선별하여 인명피해가 발생할 가능성에 대한 시나리오를 구축하는 것이기 때문에 화재발생 후 발화열원, 최초착화물, 화재발생 위치 및 연소확대범위 등 다양한 요인에 의한 결과를 확인하여야 한다. 이에 대하여 전통적인 통계분석 기법을 적용하면 교호작용 효과(interaction effect), 이상값(outlier), 통계적 가정사항 등 고려해야 할 사항이 많아지고 이는 모형의 결과해석 및 성능의 저하로 귀결될 수 있다(Bzdok et al., 2018). 따라서 본 연구는 화재시나리오 구축 시 다양한 요인들이 고려되어 인명피해를 기반으로 한 시나리오가 필요하다는 점에서 머신러닝 방법을 사용하였다.

2.3 머신러닝 및 분석방법 선정근거

머신러닝이란 알고리즘을 학습하여 인간의 지능을 모방하는 컴퓨터 알고리즘의 한 분야로 새롭게 도래한 빅데이터 시대에서 패턴인식, 금융, 의료학, 생물학 및 엔지니어링 등 다양한 분야에서 적용되고 있으며, 텍스트 또는 이미지 등의 비정형 데이터 분석에서도 적용되고 있다(El Naqa and Murphy, 2015). 고전적인 통계분석 기법을 이용한 예측은 확률모형의 생성뿐만 아니라 모형을 이용한 결과의 추론에 중점을 두고 있으나, 머신러닝은 전통적 모형의 가정사항인 정규성(normality), 등분산성(homoscedasticity), 독립성(independence)에 관계없이 우수한 예측 성능을 가지는 모형에 적용할 수 있다.

특히 데이터에 변수 간 복잡한 교호작용이 존재할 경우나

변수의 개수가 관측값의 개수보다 많을 경우 유용하게 사용될 수 있다. 머신러닝은 크게 지도학습과 비지도학습으로 구분되며, 결과를 예측하기 위해서는 지도학습을 사용한다. 지도학습 기법에는 로지스틱 회귀, 나이브 베이즈 분류기, 서포트 벡터 머신, 의사결정나무, 랜덤포레스트, 인공 신경망 모형 등이 있다(Osisanwo et al., 2017). 이중 의사결정나무 모형은 머신러닝의 장점 중 하나인 통계적 분포를 가정하지 않고 결과 예측과정을 나무(tree)형태로 나타낼 수 있기 때문에 본 연구에서는 다양한 변수를 근거로 하는 화재시나리오를 파악하는데 적합할 것으로 판단되며, 의사결정나무 기반 모형인 CART 방법을 분석에 활용하였다(Chan et al., 2001).

2.4 CART 방법의 정의 및 특징

본 연구에서 사용한 CART 모형에 기반이 되는 의사결정나무란 Fig. 1과 같이 각 변수에서 하나의 값을 기준으로 지정한 이후, 그것을 순서대로 정렬하여 그룹을 분류하는 모형이다(Kotsiantis, 2007). 의사결정나무의 각 노드(node)에는 분류결과의 기준을 나타내며 각 가지(branch)는 다음 노드에 어떠한 값이 올 수 있는지 예측할 수 있는 값이 표현된다. 각각의 관측값에 대해 그룹은 뿌리노드(root node)에서 시작해서 정렬된 기준에 의거하여 분류된다. CART 모형을 이용하면 예측을 위한 규칙을 공식화할 수 있고 중요한 변수를 선별하기에도 용이하다(Brezingar-Masten, and Masten, 2012).

CART 모형의 주요장점은 분류규칙에 대한 해석이 쉽고 비모수적 방법으로 변수 간 복잡한 관계를 파악하기 용이하며 매우 강건한 모형으로 데이터에 이상값이 존재해도 예측 결과가 잘 바뀌지 않는다(Di Marco and Nieddu, 2014).

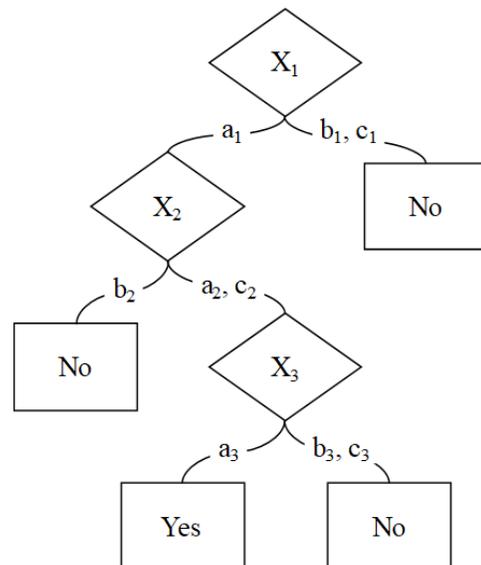


Fig. 1. Decision Tree Diagram

3. 인명피해 기반 예측모델 분석도구 및 전처리

3.1 화재시나리오 구축방법 및 분석도구

머신러닝 방법을 활용하여 화재통계 자료를 분석하기 위해서는 목표변수(Y)와 예측변수(X)를 설정하는 것이 선행되어야 한다. 화재발생빈도를 다양한 예측변수의 조합을 통해 인명피해 정도를 목표변수인 사망자 및 사상자 발생유무(유 = 1, 무 = 0)로 분류하고, 실제 구축된 화재시나리오를 기반으로 발생 여부를 확률로 예측하였다. 본 연구의 분석은 통계분석 도구인 Minitab 20 (Minitab LLC, USA)을 활용하였다.

3.2 변수정보 및 데이터 전처리

본 연구에서 사용된 데이터세트는 화재발생 연도기준 2009년부터 2018년까지의 10년간 데이터로 총 256,056건(행)이며, 분석에 사용된 전체 변수는 58개(열)이다. 여기서 총 데이터의 수는 임야화재, 자동차 화재 등은 제외하고 건축물의 화재발생만을 분석하였으며, 건축물의 상태가 철거나 신축공사 등에 해당되는 데이터는 제외하고 사용 중인 건축물만 집계하였다. 데이터는 유형에 따라 크게 연속형 또는 범주형으로 분류할 수 있는데 Table 1과 같이 연속형 변수는 재산피해, 건물층수 등 13개, 범주형 변수는 발화원인, 발화원인 등 45개로 구분하였다. 또한 범주형 변수 중 목표변수 2개(사망자 및 사상자 여부)를 제외한 43개의 변수는 예측변수로 분석에 활용하였다.

또한 분석의 신뢰도 및 정확도 향상을 목적으로 데이터 전처리(data preprocessing)를 수행하였는데 부정확한 데이터, 상식적인 수준을 벗어나는 이상값 및 데이터 행의 정보가 존재하지 않는 결측값(missing value)을 제거하였다. 더욱이 예측변수의 유형 중 범주형 변수의 경우는 변수 내의 각 수준의 빈도(또는 비율)가 낮을 때 모형 성능이 저하되는 현상이 발생되기 때문에 범주의 수준수를 통합 및 수정하여 재범주화를 실시하였다.

한편, 전체 데이터 중 초기소화활동, 피난유도, 자동소화설비 작동여부 등에 관한 변수가 존재하였으나 대부분 결측값

Table 1. Raw Dataset Information for Fitting the Prediction Model

No	Variables	Name	Type
1	Y	Death	○
2	Y	Casualties	○
3	X	Property damage/Floor area	●
4	X	Category of Property damage	○
5	X	Disappearance area	○
...
58	X	Ignition point	○

○ : Categorical, ● : Continuous

으로 구성(자동소화설비 작동 결측값은 247,147건(96.5%), 초기소화활동 결측값은 213,062건(83.2%)되어 있어 이를 포함할 때 데이터 분석이 어렵기 때문에 제외하고 진행하였다.

3.3 파생변수 생성 및 핵심변수 선정

모델의 성능 향상을 위해 전술한 바와 같이 기존 관측된 예측변수들을 활용한 새로운 파생변수를 추가로 생성하였다. 일반적으로 파생변수를 생성하기 위해서는 예측변수는 사칙연산을 적용하고 범주형 변수는 논리값을 이용한다. 이에 Table 2와 같이 화재발생계절, 화재발생일자 등 13개의 범주형 변수와 연면적 대비 바닥면적 비율, 지상층수 대비 재산피해 등 8개의 연속형 변수에 대하여 21개의 파생변수를 생성하였다.

또한 예측모델 구축에 앞서 파생변수를 포함한 어떤 변수가 목표변수에 영향을 미치는지 확인하기 위해 카이제곱 검증 및 이항로지스틱 회귀분석을 수행하였고, Table 3에 목표변수에 영향을 주는 재산피해, 연면적 및 발화원인 등 34개의 예측변수를 선정하였다.

Table 2. Derived Variables for Fitting the Prediction Model

No	Variables	Name	Type
1	X	Season	○
2	X	Basement	○
3	X	Fire location (ground or basement)	○
4	X	Ratio of floor area to total floor area	●
5	X	Ratio of property damage to gross floor area	●
...
21	X	Ratio of lost area to total floor area	●

○ : Categorical, ● : Continuous

Table 3. Final Variables for Fitting the Prediction Model

No	Variables	Name	Type
1	Y	Death	○
2	Y	Casualties	○
3	X	Property damage/Floor area	●
4	X	Combustion area	●
5	X	Gross floor area	●
...
34	X	Time of fire	○

○ : Categorical, ● : Continuous

3.4 불균형 데이터 처리

불균형 데이터란 다수 범주의 관측값 수와 소수 범주의 관측값 수의 차이가 크게 나타나는 경우의 데이터를 의미한다. 예를 들어 암 발생, 공정에서의 불량 데이터 발생, 금융 거래 시 카드기 등의 경우가 있다. 일반적으로 주요변수의 특성은 소규모 수로 구성된 경우가 많으며 이를 정확하게 분류하는 것은 매우 중요하다.

불균형 데이터로 구성되어 있는 경우에는 모형의 성능을 과대 또는 과소로 추정하는 경우가 발생되는데 이러한 문제에 대해서는 샘플링 기법을 통해 데이터의 수를 조정함으로써 문제를 해결할 수 있다. 샘플링 기법은 다수 범주의 데이터를 축소하는 언더샘플링 방법과 소수 범주의 데이터를 확장하는 오버샘플링 방법이 있다. 언더샘플링 방법에는 Random under sampling (RUS), Tomek links, Condensed nearest neighbor rule (CNN), One-sided selection (OSS) 등의 방법이 있고, 오버샘플링 기법에는 Random over sampling (ROS), SMOTE, Borderline-SMOTE, ADASYN 등이 있다.

본 연구에서의 이상값과 결측값을 제외한 최종 분석 데이터의 수는 사망자 예측모델은 비사망자 221,237건(99.17건), 사망자 1,856건(0.83%)이며, 사상자 예측모델에서는 비사망자 212,653건(95.32%), 사상자 10,440건(4.68%)으로 극심한 데이터 불균형이 존재하였다. 이에 불균형 문제를 해결하기 위해 다수 범주의 데이터를 무작위로 샘플링하여 소수의 데이터 수에 맞게 조정하는 방법으로 RUS방법을 사용하였다(Batista et al., 2004). 그 결과 최종 분석데이터 수는 사망자 예측모델은 비사망자 및 사망자 각각 1,856건, 사상자 예측모델은 비사망자 및 사상자 각각 10,440건으로 도출되었다.

4. 인명피해 예측모델 구축 및 검증

4.1 예측모델 구축방법

본 연구의 원(raw)데이터는 목표변수 사전비율이 심한 불균형 자료구조이므로 모델의 분류 정확도 개선을 위해 RUS 방식을 이용하여 사망자의 데이터 수인 1,856건으로 비사망자의 데이터 수를 5:5의 비율로 조정하여 총 3,152건으로 분석을 수행하였으며, 비사망자의 데이터 수도 동일한 방법으로 20,880건이다.

예측모델의 성능 검증은 K-폴드 교차검증 방법(K-fold cross-validation)을 이용하였다. 이 검증방법은 적은 표본 데이터셋에서 예측과 성능을 향상시킬 수 있는 검증방법이다(Nie et al., 2020). 여기서, K 값은 일반적인 수준인 10으로 설정하였다.

4.2 예측모델 검증지표

목표변수가 0과 1만 존재하는 범주형(이분형) 데이터로 구성된 경우 예측모델이 얼마나 정확하게 분류하는지 성능의 검증이 필요하다. 본 연구에서는 인명피해 예측모델의

성능을 평가하고 검증하기 위해 Table 4와 같이 정오행렬을 사용하여 정확도, 민감도, 특이도를 검증지표로 확인하였다.

정오분류표의 true positive (TP)는 실제 관측값이 1인데 모델에서 1로 정확히 예측하는 건수, false negative (FN)은 실제 관측값이 1인데 모델에서 0으로 예측하는 건수, false positive (FP)는 실제 관측값이 0인데 모델에서 1로 예측하는 건수, true negative (TN)은 실제 관측값이 0인데 모델에서 0으로 정확히 예측하는 건수를 의미한다.

정확도는 전체 관측값 중 실제 관측값과 모델에서 예측값이 일치하는 비율, 민감도는 실제 관측값이 참(true)인 경우 참으로 예측하는 비율, 특이도는 실제 관측값이 거짓(false)인 경우 거짓으로 예측하는 비율을 말하며, Eps. (1)-(3)과 같이 표현할 수 있다(Shobha and Rangaswamy, 2018).

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

$$Sensitivity = \frac{TP}{TP+FN} \quad (2)$$

$$Specificity = \frac{TN}{FP+TN} \quad (3)$$

4.3 인명피해 예측모델 구축결과 및 검증

4.3.1 사망자 예측모델

사망자 예측모델을 구축하기 위해 CART 방법의 다중나무구조 분류 결과는 Fig. 2와 같다. 총 32개 터미널 노드로 구성된 나무모형이 생성되었으며, 나무의 최상단 뿌리노드는 3,712건의 관측값으로 사망자 및 비사망자 건수의 비율은 각각 50%로 나타났다. 또한 Table 5와 같이 사망자 예측모델에서 변수 중요도를 확인한 결과 화재로 인해 사망자에 영향을 주는 중요한 변수로는 재산피해, 연면적, 바닥면적 순으로 나타났으며, Table 6의 정오분류표와 같이 정확도는

Table 4. Expression for Term of Confusion Matrix

Actual	Predicted	
	Positive (1)	Negative (0)
Positive (1)	TP	FN
Negative (0)	FP	TN

Table 5. Relatively Important top 20% Variables in Death Prediction Model

No	Variables
1	Property damage / Building basement
2	Property damage
3	Property damage / Floor area
4	Category of property damage
5	Disappearance area
...	...
20	Ignition point

84.4%, 민감도는 86.7%, 특이도는 82.0%로 높은 예측성능을 가지는 것으로 나타났다.

Table 6. Confusion Matrix for Death Prediction Model

Actual	Predicted	
	Positive (1)	Negative (0)
Positive (1)	1,610	246
Negative (0)	334	1,522
Accuracy	84.4%	
Sensitivity	86.7%	
Specificity	82.0%	

4.3.2 사상자 예측모델

사상자 예측모델을 구축하기 위해 CART 방법의 다중나무구조 분류결과는 Fig. 3과 같다. 총 30개의 터미널노드로 구성된 나무모형이 생성되었으며, 나무의 최상단 뿌리노드는 20,880건의 관측값으로 사상자 및 비사상자 건수의 비율은 각각 50%로 나타났다. 또한 Table 7과 같이 변수 중요도를 확인한 결과 화재로 인해 사상자에 영향을 주는 중요한 변수는 재산피해, 소실면적, 연면적, 바닥면적 순으로 나타났으며, Table 8의 정오분류표와 같이 정확도는 72.6%, 민감도는 70.5%, 특이도는 74.7%로 높은 예측성능을 가지는 것으로 나타났다.

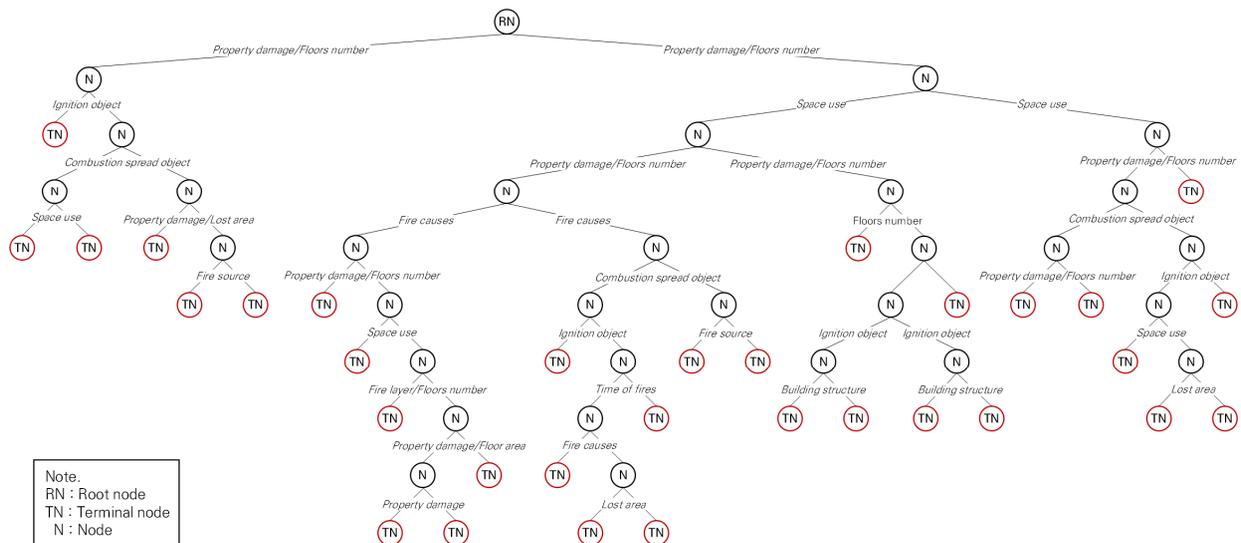


Fig. 2. Classification Result of Tree Structure for Death Model

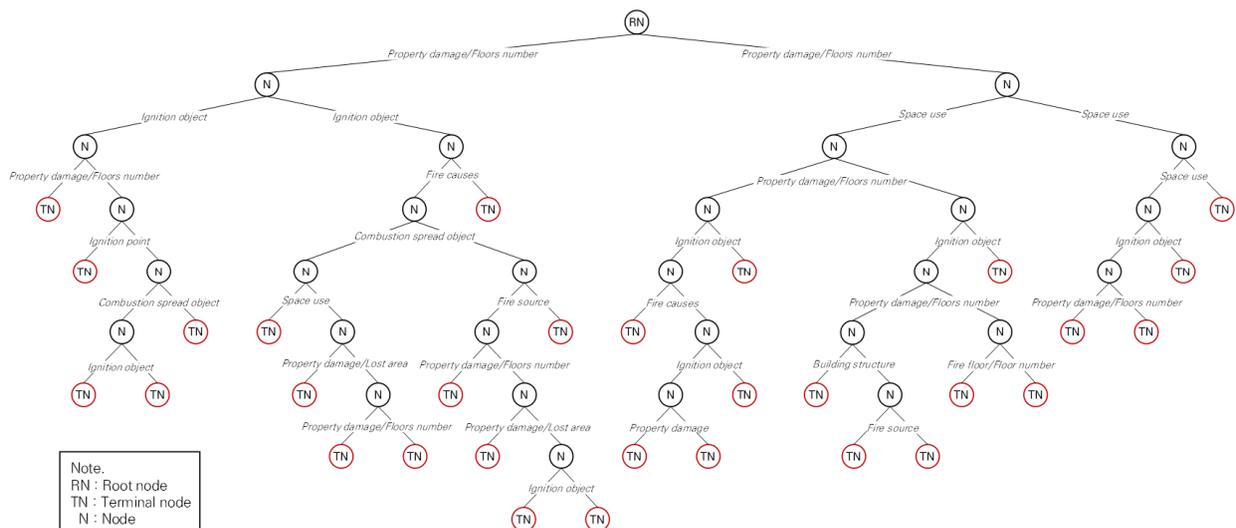


Fig. 3. Classification Result of Tree Structure for Casualties Model

4.4 인명피해 예측모델을 이용한 화재시나리오 검토

4.4.1 화재시나리오 조건

구축된 인명피해(사망자 및 사상자) 예측모델을 이용하여 가상의 건축물을 가정하여 화재시나리오를 검토하였다. 본 연구에서 구축한 예측모델은 다양한 변수가 존재하지만 화재발생 이전의 과정에서 입력할 수 있는 변수로 선정하고, 사망자 및 사상자 확률을 확인하였다. 대상 건축물은 주거용도로서 연면적 4,500 m², 바닥면적은 300 m², 지상 15층 건축물이며, 지하층은 없는 것으로 하였다.

변수로 사용된 부분은 설계화원과 화재성장률과 관계되어지는 발화지점, 최초착화물, 발화원인과 피난위험에 영향

Table 7. Relatively Important top 20% Variables in Casualties Prediction Model

No	Variables
1	Property damage / Building basement
2	Property damage
3	Property damage / Floor area
4	Category of property damage
5	Disappearance area
...	...
20	Ignition point

Table 8. Confusion Matrix for Casualties Prediction Model

Actual	Predicted	
	Positive (1)	Negative (0)
Positive (1)	7,362	3,078
Negative (0)	2,643	7,797
Accuracy	72.6%	
Sensitivity	70.5%	
Specificity	74.7%	

을 미칠 수 있는 발화층, 화재발생시간의 5개의 변수를 이용해 Table 9에 나타난 바와 같이 총 48개의 화재시나리오를 도출하였다.

화재시나리오 관점에서는 다양한 시나리오를 적용하여 검토할 필요가 있으나 4.3절에서 모델의 검증을 하였고 이에 가상의 시나리오에 대한 것의 가능성을 검토하는데 주안점을 두었다. 특히 전술한 바와 같이 동일한 주거용도이지만 고려되는 변수에 따라 사망자 및 사상자 발생확률이 차이가 분명히 나타나는 것을 확인하였다.

이에 발화지점은 주거공간에서 발생될 수 있는 생활공간과 설비공간으로 구분하였고, 화재발생시간은 오후와 심야시간으로 구분하였다. 최초착화물은 쓰레기류, 침구류 및 전기제품 3개로 구분하고, 발화층은 2층과 10층, 발화원인은 전기적 요인과 방화(arson)로 구분하여 시나리오를 구성하였다.

4.4.2 인명피해 예측모델 결과

총 48개의 화재시나리오에 근거하여 예측한 결과는 Fig. 4와 같다. 발화장소의 경우 생활공간과 설비공간으로 구분할 수 있는데 주거용도 건축물의 경우 생활공간에서의 화재가 대부분을 차지하고 있어 다른 인자에 영향을 받지 않는 것으로 분석되었다. 특히 사망자 예측모델의 경우 사망자 발생확률이 88.07%로 일률적으로 도출되는 것에 기인했을 때 사망자 발생확률은 발화장소에 영향을 미치는 것으로

Table 9. Conditions of Fire Scenario

Condition	Contents (symbol)
Cases	48 Fire scenarios
Ignition space	Living space (L), equipment space (E)
Time of fires	Afternoon (A), Midnight (M)
First Combustibles	Waste (W), bedclothes (B), electrical product (EP)
Ignition floors	2nd floor (2), 10th floor (10)
Ignition causes	Electrical (EI), arson (Ar)

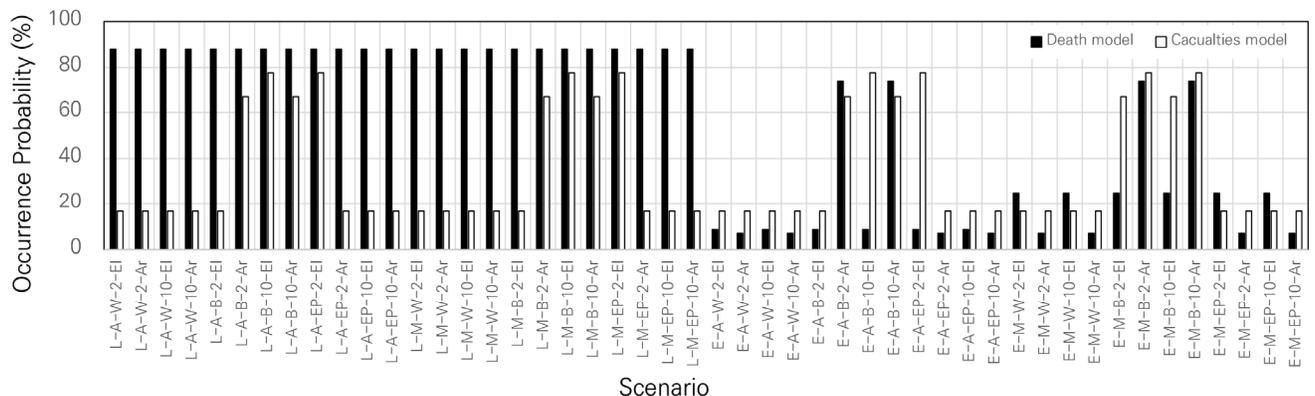


Fig. 4. Classification Result of Tree Structure for Casualties Model

볼 수 있다. 사상자 예측모델도 사망자 예측모델과 마찬가지로 발화장소에 많은 영향을 미치고 있었으나 최초착화물이 침구일 경우에는 발화원인이 전기적(67.35%) 또는 방화(67.35%)일 때 차이가 나타났다. 생활공간인 경우에는 대부분의 발생확률 결과가 상기 서술한 바와 같이 유사하게 도출되지만 최초착화물이 전기물품인 경우 사상자 예측모델에서는 77.68%로 도출됨으로써 심야시간의 2층에서 발화된 화재(L-M-EP-2-EI)의 경우는 화재시나리오에서 검토할 필요가 있는 사례가 될 수 있다.

한편, 발화장소가 설비공간(E)인 경우에는 발화원인에 기인한 결과를 보이고 있는데 여기서도 최초착화물이 침구류(B)와 발화원인이 방화(Ar)일 경우에 높은 사망자(74.07%) 및 사상자(67.35%) 발생확률이 나타났다.

이에 대한 내용을 정리하면 생활공간의 경우에는 사망자 발생확률은 88.07%로 유사하게 나타났으며, 사상자 발생확률은 오후시간대의 저층부(2층) 화재의 경우는 전기제품 화재, 고층부(10층) 화재의 경우는 침구류 화재에서 높은 사상자 발생확률을 확인할 수 있다.

또한 설비공간의 경우 사망자 발생확률은 침구류 화재에서 방화로 인한 화재가 발생할 경우 74.07%의 사망자 발생확률을 확인할 수 있었으며, 그 외의 결과에서는 낮은 사망자 발생확률로 나타났다. 사상자 발생확률의 경우 침구류 화재인 경우는 저층부(2층)의 경우를 제외하고 발화원인이 방화(67.35%), 전기적(77.68%)로 높게 나타났으며 이는 침구류 화재에 영향을 보이는 것으로 보이며 심야시간대는 동일한 화재조건이라도 발화원인이 방화일 경우 전기화재에 비해 높은 경향을 보였다.

분석된 사망자 및 사상자의 발생확률의 의미하는 것은 해당 시나리오에서 화재가 발생했을 때 반드시 발생할 수 있는 것에 기인하는 것이 아니라 개별 모델의 정확도, 민감도 및 특이도 수준에 따른 발생확률이라는 것을 인지할 필요가 있다. 예를 들면 1개의 화재시나리오에서 도출된 사망자 발생확률이 90%라는 것은 이 발생확률이 70%의 예측력을 지니고 있다고 이해하는 것이 중요하다.

본 연구에서는 화재시나리오를 인명피해에 기인하여 다양한 변수의 적용성을 확인하는 것이기 때문에 발생빈도에 기초하지 않고 사망자 및 사상자 발생확률을 도출할 수 있었다. 하지만 사망자 및 사상자 수는 전체 화재데이터에 비례하여 상당히 적은 수이며 모델의 예측력(정확도, 민감도, 특이도)이 높게 나타났으나 실제 모델을 적용하기 위해서는 영향력이 적은 변수량 제외, 원데이터의 신뢰성 확보, 미상 또는 기타와 같은 부정확한 데이터의 분류방법 등을 고려한 추가 연구가 필요하다. 다만, 화재 발생빈도뿐만 확률을 정하고 있는 현 시점에서 다양한 변수에 기인하여 인명피해 발생확률을 도출하여 화재시나리오의 적절성을 확보하는 차원에서 본 연구결과는 유의미하다고 사료된다.

5. 결론 및 향후 연구방향

본 연구에서는 화재통계를 기반으로 화재위험평가에 활용될 수 있는 화재시나리오 선정에 위한 인명피해 예측모델을 구축하였다. 화재발생빈도를 중심으로 이루어지고 있는 화재시나리오 선정방법에 다양한 변수의 영향을 엔지니어 등이 확인할 수 있고 인명안전에 기인한 선정방법을 제시한다는 점에서 보다 안전한 평가가 이루어질 수 있을 것으로 판단되며 다음과 같은 결론을 얻을 수 있었다.

- 1) 화재시나리오 구축을 위해 데이터 전처리, 파생변수 생성, 불균형 데이터 처리 및 RUS방법을 이용해 데이터셋을 구축하고, CART 방법의 다중나무구조 분류를 통해 정오분류표에 의한 검증한 결과, 사망자 예측모델은 정확도, 민감도 및 특이도는 각각 84.4%, 86.7%, 82.0%로 도출되었으며, 사상자 예측모델은 각각 72.6%, 70.5%, 74.7%로 나타나 높은 예측성능을 가지는 것으로 검증하였다.
- 2) 또한 인명피해 예측모델에 대하여 설계화원, 화재성장을 및 피난에 영향을 미칠 수 있는 5개의 변수에 대해 48개 화재시나리오를 도출하고 이에 대한 검토를 가상의 건축물을 대상으로 검토를 수행한 결과, 발화장소가 발생확률에 가장 영향을 미치는 것으로 나타났으며 최초착화물 중 침구류에 있어서는 발화층, 발화원인, 화재발생시간에 따라 사망자 및 사상자 발생확률의 차이를 확인하였다.
- 3) 본 연구에서 화재발생빈도 외의 다른 변수에 대하여 머신러닝 방법을 이용해 영향성을 확인하였으나 영향력이 적은 변수가 많거나 원데이터의 신뢰성 확보 문제 등 현재 화재통계자료로 제공되는 데이터의 한계점을 확인할 수 있었다. 향후 영향력이 높은 변수에 대해서 보다 세부적인 확인이 필요하며, 모델의 신뢰성을 확보하기 위해서는 소방청에서 제공되는 명확한 데이터로 재구성해 화재시나리오 선정에 있어서 보다 체계적인 접근이 필요할 것으로 판단된다.

감사의 글

본 연구는 한국건설기술연구원 ‘건축물의 화재취약 특성을 반영한 화재위험 평가기술 개발(3차년도)’의 연구비 지원에 의해 수행되었습니다.

References

Aguiar, F.S., Almeida, L.L., Ruffino-Netto, A., Kritski, A.L., Mello, F.C., and Wernick, G.L. (2012). Classification and regression tree (CART) model to predict pulmonary tuberculosis in hospitalized patients. *BMC pulmonary*

- medicine*, Vol. 12, No. 1, pp. 1-8.
- An, S.H., Mun, S.Y., Ryu, I.H., Choi, J.H., and Hwang, C.H. (2017). Analysis on the implementation status of domestic PBD (Performance Based Design): Focusing on the fire scenario and simulation. *Journal of the Korean Society of Safety*, Vol. 32, No. 5, pp. 32-40.
- Batista, E.G., Prati, C.R., and Monard, C.M. (2004). A study on the behavior of several methods for balancing machine learning training data. *ACM SIGKDD Explorations Newsletter*, Vol. 6, No. 1, pp. 20-29.
- Brezingar-Masten, A., and Masten, I. (2012). CART-based selection of bankruptcy predictors for the logit model. *Expert Systems with Applications*, Vol. 39, No. 11, pp. 10153-10159.
- Bzdok, D., Altman, N., and Krzywinski, M. (2018). Points of significance: Statistics versus machine learning. *Journal of Nature methods*, Vol. 15, No. 4, pp. 233-234.
- Chan, J.C.W., Chan, K.P., and Yeh, A.G.O. (2001). Detecting the nature of change in an urban environment: A comparison of machine learning algorithms. *Photogrammetric Engineering and Remote Sensing*, Vol. 67, No. 2, pp. 213-226.
- Di Marco, L., and Nieddu, L. (2014). Trigger factors that influence bankruptcy: A comparative and exploratory study. *Rivista Italiana di Economia Demografia e Statistica*, Vol. 68, No. 3/4, pp. 191-198.
- El Naqa, I., and Murphy, M.J. (2015). What is Machine Learning?, *Machine Learning in Radiation Oncology*. Springer, Cham, pp. 3-11.
- Hadjisophocleous, G.V., and Fu, Z. (2004). Literature review of fire risk assessment methodologies. *International Journal on Engineering Performance-Based Fire Codes*, Vol. 6, No. 1, pp. 28-45.
- Jang, K.H. (2016). Proposals on the input data standardization needs of fire and evacuation simulation in performance based design. *Journal of Fire Science and Engineering*, Vol. 30, No. 5, pp. 18-25.
- Jeong, H.M., Hwa, H.W., Son, J.Y., and Jeon, G.Y. (2018). A study on establishment of performance- based design direction through analysis of expert survey. *Journal of The Architectural Institute of Korea Planning and Design*, Vol. 34, No. 2, pp. 23-31.
- Karas, M., and Reznakova, M. (2017). Predicting the bankruptcy of construction companies: A CART-based model. *Engineering Economics*, Vol. 28, No. 2, pp. 145-154.
- Kim, H.J., Lee, S.H., Kim, I.H., and Hong, W.H. (2012). A study on the evaluation index through weight analysis of fire and egress scenario factors: Focused on high-rise mixed-use apartment. *Journal of the Architectural Institute of Korea Planning and Design*, Vol. 28, No. 12, pp. 371-378.
- Kotsiantis, S.B. (2007). Supervised machine learning: A review of classification techniques. *Informatica*, Vol. 31, No. 2007, pp. 249-268.
- National Fire Agency (NFA). (2020). *Fire Statistical Yearbook, 2020 year Edition*, p. 21.
- Nie, Y., Santis, L.D., Carratu, M., O’Niles, M., Sommella, P., and Lundgren, J. (2020). Deep melanoma classification with K-fold cross-validation for process optimization. *2020 IEEE International Symposium on Medical Measurement and Applications (MeMeA)*, 10 July 2020, Bari, Italy.
- Oh, H.J., Baek, S.T., Kim, W.S., and Lee, S.K. (2003). A study on the evacuation performance review for the office buildings. *Journal of Fire Science and Engineering*, Vol. 17, No. 3, pp. 1-6.
- Osisanwo, F.Y., Akinsola, J.E.T., Awodele, O., Hinmikaiye, J.O., Olakanmi, O., and Akinjobi, J. (2017). Supervised machine learning algorithms: Classification and comparison. *International Journal of Computer Trends and Technology (IJCTT)*, Vol. 48, No. 3, pp. 128-138.
- Park, J.K., Roh, S.K., and Seo, Y.M. (2002). A study on the risk assessment system of the underground space. *Journal of Fire Science and Engineering*, Vol. 16, No. 2, pp. 70-74.
- Seo, M.J., Lee, Y.J., An, S.H., Hwang, C.H., and Choi, J.H. (2017). A study on improvement way of fire simulation modelling field through analysis of performance-based design reports of high-rise residential complex building in B metropolitan city. *Journal of Fire Science and Engineering*, Vol. 31, No. 4, pp. 80-85.
- Shobha, G., and Rangaswamy, S. (2018). Machine learning. *Handbook of Statistics, Chapter 8*, Vol. 38, pp. 197-228.

Received ■ September 17, 2021

Revised ■ September 23, 2021

Accepted ■ September 30, 2021