



초대형화재사고 예측을 위한 화재사고 분류의 개선 및 발생의 주기성 분석

김창완 · †신동일*

명지대학교 화학공학과 석사과정, *명지대학교 화학공학과 교수
(2020년 1월 6일 접수, 2020년 2월 21일 수정, 2020년 2월 22일 채택)

Improved Classification of Fire Accidents and Analysis of Periodicity for Prediction of Critical Fire Accidents

Chang Won Kim · †Dongil Shin

Department of Chemical Engineering, Myongji University, Yongin, Gyeonggido 17058, Korea
(Received January 6, 2020; Revised February 21, 2020; Accepted February 22, 2020)

요약

일반적으로 화재는 다양한 원인으로 발생하며 무작위로 보이기에 화재의 발생을 예측한다는 것은 매우 도전적인 문제이다. 하지만 모든 화재가 아닌 큰 피해를 주는 초대형 화재사고의 예측이 가능하다면, 선제적 대응을 통한 손실 최소화할 수 있다. 본 연구에서는 국가 전체를 대상으로 초대형 화재사고를 예측하기 위해 기계학습 기법인 k-평균 클러스터링을 이용하여 화재사고를 분류하고, 이를 인위적인 설정이 강한 비전문가 기준, 전문가 기준 분류 결과와 비교하여 예측에 적절한 분류 기준을 제안하였다. 비교 결과 기계학습을 이용한 분류가 일정한 피해 규모와 비율로 분류되어, 예측에 적절한 분류 기준이라 판단하였다. 또한 초대형 화재사고의 주기성을 분석한 결과 일정한 패턴을 보였지만 높은 편차를 보였다. 따라서 단순 예측기법이 아닌 고급 예측기법을 사용하였을 때 초대형 화재사고의 발생 예측이 가능하다고 판단되었다.

Abstract - Forecasting of coming fire accidents is quite a challenging problem cause normally fire accidents occur for a variety of reasons and seem randomness. However, if fire accidents that cause critical losses can be forecasted, it can expect to minimize losses through preemptive action. Classifications using machine learning were determined as appropriate classification criteria for the forecasting cause it classified as a constant damage scale and proportion. In addition, the analysis of the periodicity of a critical fire accident showed a certain pattern, but showed a high deviation. So it seems possible to forecast critical fire accidents using advanced prediction techniques rather than simple prediction techniques.

Key words : Incident classification, Machine learning, Heinrich's law, k-means clustering

1. 서론

우리나라는 경제가 발전함에 따라 여러 위험의 대형화, 고도화, 집적화 및 복합화 되는 현상을 보이고 있다[1]. 이러한 사회 위험이 증가함에 따라 위험관리 기술이나 제도를 증가시켜 사회의 안정화 시키는 것이 중요하다. 특히 우리나라에서 발생하는 대형화재의 경우 대부분 인적 사고로 인한 사

고이며, 자연재해라고 판단된 문제 또한 개발과정에서 생긴 원인으로 인해 발생한 인적 재해 문제로 해석된다. 이와 같은 사회 위험은 전 세계적으로 다양하게 일어나며, 화재사고피해를 줄이기 위해 각국은 화재사고 대응 및 예방에 대한 다양한 노력을 기울이고 있다. 특히 컴퓨터 기술의 발달로 여러 나라에서 국가 화재데이터베이스를 구축하고 분석하고 있다. 미국의 경우, 화재 정보는 국가 화재 정보 보고 시스템(NFIRS)에 의해 체계적으로 저장되고 관리된다[2]. 이와 더불어 캐나다에서는 화재위험지역을 알려주기 위해 화재위험도(AHRA)를

†Corresponding author: dongil@mju.ac.kr

Copyright © 2019 by The Korean Institute of Gas

분석 및 제공한다[3]. 우리나라에서는 매년 발생하는 모든 화재의 조사 데이터를 국가화재정보시스템(NFDS)에 저장한다. NFDS 데이터에는 시간 정보(년, 월, 일), 공간 정보(도시, 산, 해안 등), 환경 정보(온도, 습도, 계절 특성), 소방 대응 정보 등이 포함된다[4]. 우리나라는 매년 평균 4만 건의 화재가 발생하며 NFDS에 많은 자료가 축적되어 있다. 그러나 축적된 NFDS의 빅데이터를 분석하고 앞으로 일어날 화재를 예측하는 일반적인 모델의 연구가 부족한 실정이다.

본 연구에서는 화재사고의 분류를 개선하기 위해 사고관리 기준의 분류와 사고 예측을 위한 분류 측면을 분석 및 비교하였다. 문제를 해결하기 위해 아래와 같은 분류 기법들을 사용하였다.

1. 전통적으로 산업재해 예방에 사용되는 Heinrich 법칙을 국가 화재사고에 적용한 비전문가 수준의 분류
2. 현직 소방관의 경험에 기반을 둔 전문가 기준의 분류
3. 화재데이터에 기반을 둔 기계학습 기법 분류
화재사고의 경우 피해가 작은 사고라도 충분히 피해 확대를 야기할 수 있는 사고이기에 Heinrich에 사용된 3가지 등급(경미한 사고, 소형 사고, 대형사고)이 아닌 소형, 대형, 초대형 화재사고로 분류하였다[5].

II. 이론

2.1. Heinrich의 법칙

Heinrich의 법칙은 산업재해 예방을 위해 사용되는 통계적 방법이다. Heinrich의 법칙은 경미한 사고(near miss), 소형사고(minor incident), 대형사고(major incident)의 비율 분포를 분석하였다. Heinrich는 사망자가 존재하는 경우 대형사고, 사망자는 없지만 부상자가 존재하는 경우 소형사고, 사망자와 부상자가 없지만 재산피해나 아차사고를 경미한 사고로 구분하였다. 이 분류 방법을 이용하여 5,000건의 산업사고를 분석한 결과, 대형사고 1건을 기준하여 소형사고 29건과 경미한 사고 300건의 비율을 가짐을 발견하였다. 이러한 분포는 대형사고가 우연히 발생하는 것이 아니라 소형사고와 경미한 사고가 반복적으로 발생하고, 발생한 소형사고와 경미한 사고의 발생 원인을 무시하고 방치한다면 돌이킬 수 없는 대형사고로 번질 수 있다는 것을 경고하였다[5]. 또한 Heinrich 법칙은 다양한 분야의 사고에도 적용되며, 안전관리체계를 개선하는데

이용되고 있다[6].

2.2. k-평균 클러스터링 알고리즘

레이블이 되어있지 않는 데이터를 분류하는 기계학습 기반 클러스터링 알고리즘은 계층적 클러스터링(hierarchical clustering) 알고리즘, 밀도 기반 클러스터링(density-based spatial clustering of applications with noise) 알고리즘, k-평균 클러스터링(k-means clustering) 알고리즘 등이 있다[7]. NFDS 데이터 특성상 많은 양의 데이터가 있고, 분산되어 있기 때문에 이에 적합한 k-평균 클러스터링 알고리즘을 사용하였다. k-평균 클러스터링 알고리즘은 MacQueen[8]에 의해 처음 제안되었다. k-평균 클러스터링 알고리즘에는 유클리드 거리에 기초한 클러스터링 단계와 클러스터의 중심점 최적화 단계로 구성된다. 클러스터링 단계에서 데이터 포인트 $X = (x_1, x_2, \dots, x_n)$ 을 n개의 포인트 수를 k개의 클러스터 $S = (S_1, S_2, \dots, S_k)$ 로 분할한다. 이때 클러스터 S에서 점 X와 중심점 사이의 유클리드 거리의 제곱을 최소화하기 위한 목적 함수를 사용한다.

$$\operatorname{argmin} \sum_{i=1}^k \sum_{j \in S_i} \|x_j - \mu_i\|^2 \quad (1)$$

(1)의 수식의 μ_i 는 S_i 의 중심점을 나타낸다. 클러스터링 단계에서 분류된 점들을 이용하여 클러스터의 중심점 최적화 단계를 진행한다.

$$\mu_i^{(t+1)} = \frac{1}{|S_i^{(t)}|} \sum_{x_j \in S_i} x_j \quad (2)$$

k-평균 클러스터링 알고리즘은 클러스터 중심점 S_i 가 더 이상 변경되지 않을 때까지 두 단계를 반복한다[9].

2.3. 자기상관계수 함수

자기상관계수 함수(autocorrelation function, ACF)는 무작위의 두 시각에 따른 변수의 상관관계를 나타내는 함수로, 랜덤 프로세스의 거동을 확인하거나 이의 주기를 분석하기 위해 사용된다[10]. 자기상관계수는 두 시점의 공분산을 각각의 표준편차로 나누고, 이를 시차에 대한 함수로 표현된다. 이때 시차 k가 떨어진 변수의 자기상관계수를 구하는 수식은 아래와 같다[11].

$$r_k = \frac{\sum_{t=k+1}^n (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2} \quad (3)$$

III. k-평균 클러스터링 기반의 화재사고 분류

본 연구에서는 k-평균 클러스터링 알고리즘을 이용한 화재사고 등급 분류 단계와 분류된 화재사고의 수를 이용한 초대형 화재사고의 주기성을 계산하는 단계로 구성된다. 화재사고 등급 분류 단계 이전에 NFDS의 정보를 선택하여 데이터 분석을 목적으로 분류한다. 화재사고 등급 분류 단계에서 NFDS 데이터는 초대형, 대형, 소형 화재사고로 분류된다. 이러한 화재사고 등급 분류에는 Heinrich 법칙에 이용된 분류 방법을 이용한 비전문가 기준 분류와 기계학습 기반의 k-평균 클러스터링을 이용한다. 화재사고 등급 분류 후, 초대형 화재사고의 주기성을 탐색하기 위해 초대형 화재사고 1건 대비 대형과 소형 화재사고의 수, 다음 초대형 화재사고까지 시간을 계산한다. 만약 초대형 화재사고가 동일한 주기성을 보인다면 앞으로 일어날 초대형 화재사고의 발생 시기를 예측하는 것이 가능할 것이다.

본 연구에서 사용된 화재사고 등급 분류에는 일반적으로 화재 피해 규모를 분류할 수 있는 정보인 사망자, 부상자, 재산피해를 사용하였다. 3차원의 화재피해 데이터의 분포 특성상 x축과 y축, z축의 중심으로 데이터가 분포되어있어 클러스터의 중심점이 축으로 최적화되는 문제점이 있다. 이를 개선하기 위해 3차원의 점이 위치한 피해 면을 구하였고, 이 면과 원점과의 거리를 계산하여 1차원 데이터로 만들었다. 3가지 정보를 이용한 연간별 멀티 클러스터링 모델과 멀티 클러스터링 모델의 분류 적절성을 판단하기 위해 1가지 정보만 이용한 연간별 단일 클러스터링 3개의 분포 비율과 비교하였다.

기계학습을 이용하여 분류된 초대형 화재사고의 주기성을 탐색하기 위해 초대형 화재사고 이후 발생한 대형 화재사고와 소형 화재사고, 시간을 측정하였고, 이를 도표화하여 주기적 특성 및 추세를 알아보았다.

IV. 결과 및 고찰

본 연구에서는 매년 화재사고를 분류하였을 시 일관된 분류가 이루어지는지 분석하기 위해 6개 연도에 대해 각각 모델 결과를 도출하였다. 만약 매년 일관된 등급별 비율이 보인다면, 주기성을 통한

사고예측뿐만 아닌 사고관리 측면에서도 분류가 유의미하다고 볼 수 있다. 분류에 사용된 NFDS 데이터는 2007-2012년 국내에서 발생한 273,818건의 데이터를 이용하였으며, 세부적으로 2007년에는 47,882건, 2008년에는 49,631건, 2009년에는 47,318건, 2010년에는 41,863건, 2011년에는 43,875건, 2012년에는 43,249건의 화재가 발생하였다. 본 연구에서 사용된 k-평균 클러스터링 모델은 Python 환경을 이용하였다.

4.1. 화재사고 분류

Heinrich 법칙을 원용한 비전문가 기준의 분류는 초대형 화재사고는 사망자가 존재하는 사망사고, 대형 화재사고는 사망자는 존재하지 않지만 부상자가 존재하는 부상사고, 소형 화재사고는 사망자와 부상자가 존재하지 않는 나머지 사고로 구분하였다. 위의 분류에 따라 2007-2012년 NFDS를 분류하면, 초대형 화재사고 1,738건, 대형 화재사고 7,951건, 소형 화재사고 264,129건으로 나타났으며, 이 결과를 Fig. 1과 같이 3-D그래프로 나타내었다.

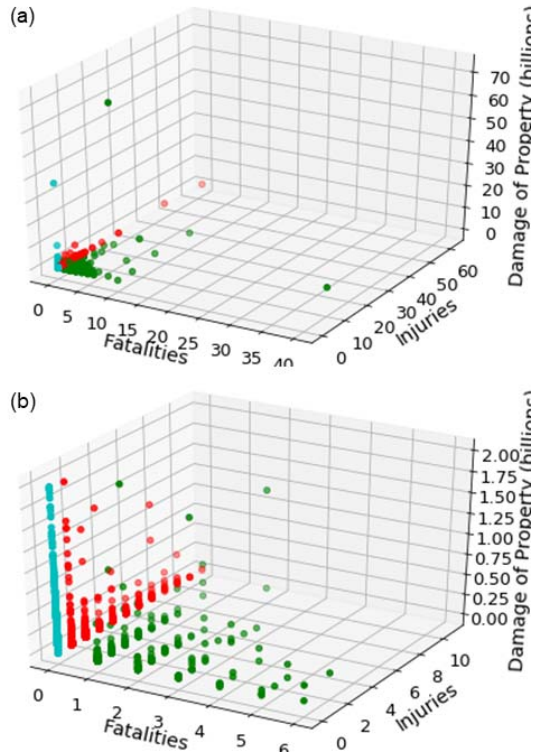


Fig. 1. Results of classification according to non-expert standard.

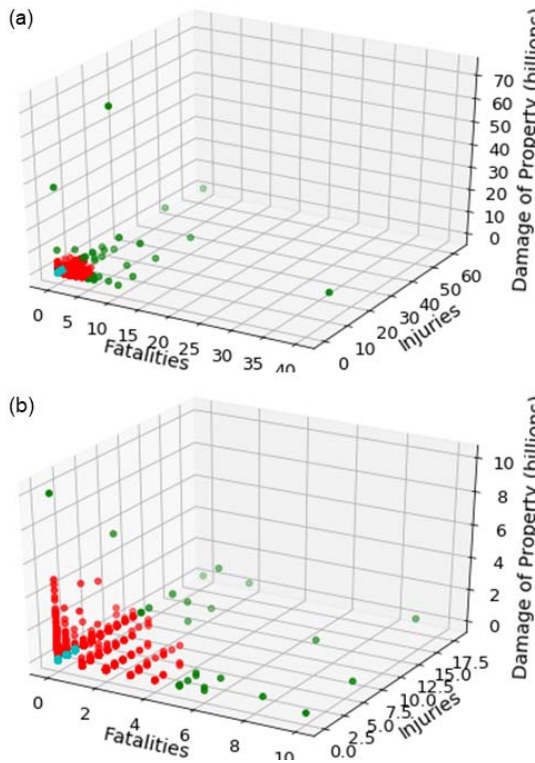


Fig. 2. Results of classification according to expert standard.

그래프에서 x축은 사망자수, y축은 부상자수, z축은 십억원 기준 재산피해를 나타내며, 녹색은 초대형 화재사고를, 적색은 대형 화재사고를, 청록색은 소형 화재사고를 나타낸다. Fig. 1a 그래프는 2007-2012년 정보를 모두 나타내었으며, 이들 중 사망자 6명 이내, 부상자 12명 이내, 재산피해 20억원 이내의 사고만을 Fig. 1b 그래프에 나타내었다. 이 결과에 따르면 단지 사망자와 부상자의 유무에 따라 등급이 나누어졌기에 사망자를 포함한 다수의 사고들이 초대형 화재사고로 분류되었다. 이로 인해 사망자가 발생하지 않았지만, 수많은 부상자나 큰 재산피해를 입은 화재가 대형 또는 소형으로 분류되어 화재피해별 등급 분류라고 보기 어렵다. 전문가 기준의 분류는 전문가의 의견을 반영하여 소형 화재사고를 사망자가 발생하지 않은 사고 중 주변 가구로 확대되지 않은 단일가구 화재로 설정하였고, 이를 2018년 기준 평균 가구원수인 2.4명을 기준하여 3명 이하의 부상자 또는 1억원 이하의 재산피해 사고로 기준하였다[12]. 초대형 화재사고의 경우 기본소방법의 화재조사 및 보고규정에 명시되어있는 사망자 5명 또는 인명피해 10명 또는 50억원 이상의 재산피해를 기준하였다[13]. 대형 화재사고의 경우 소형, 초대형 화재사고를 제외한 나머지 사고로 사망자가 포함된 사고 또는 2가구 이상의 화재피해를 기준으로 설정하였다. 전문가의 의견에 따르면, 기본소방법상으로 구분된 대형화재사고는 단지 관리하기 위한 목적으로 대형화재발생시 신속한 대응을 위해 분류한 것이지

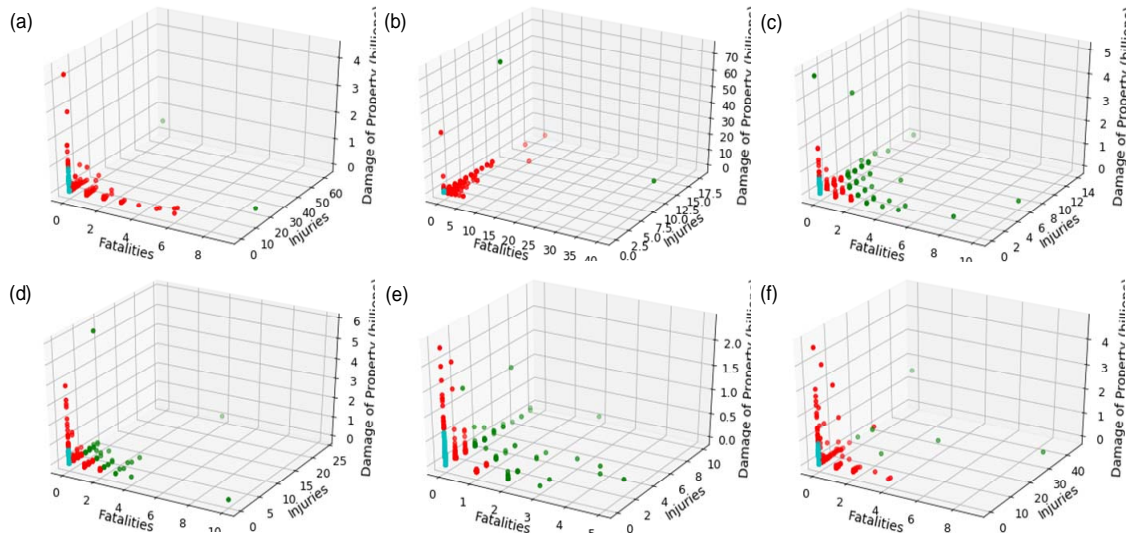


Fig. 3. Results of multiple attribute clustering in 2007-2012.

등급과 예측의 측면은 고려되지 않았다고 한다.

전문가 기준의 분류 결과에 따르면 2007-2012년 사이에 초대형 화재사고 42건, 대형 화재사고 4,048건, 소형 화재사고 269,728건이 나타났으며, 이 결과를 3-D그래프로 나타내었을 때, Fig. 2와 같이 도출된다. Fig. 2a 그래프는 2007-2012년 정보를 모두 나타내었으며, Fig. 2b 그래프는 사망자 10명 이내, 부상자 20명 이내, 재산피해 100억원 이내의 사고만을 나타내었다. 그래프를 살펴보면, 비전문가 기준과는 다르게 어느 측면이어도 큰 피해가 발생하였을 경우 대형 또는 초대형으로 분류될 수 있었다.

Fig. 3은 각 연도별로 진행한 멀티 클러스터링 결과를 3-D 그래프로 나타내었다. 멀티 클러스터의 각 정보의 단위 불균형으로 인해 특정 정보와 클러스터 중심점의 유클리드 거리에 큰 영향을 주는 문제가 발생하였고, 클러스터 중심점이 특정 정보에 맞추어 최적화되는 문제가 있었다. 이를 해결하기

위해 데이터의 최소-최대 정규화 작업을 진행하였고, 이를 위해 2007-2012년 기준 각 정보의 최대 피해인 사망자 40명, 부상자 62명, 재산피해 720억원을 정규화 기준으로 설정하였다. 기계학습을 이용하여 분류를 진행한 결과, 초대형 화재사고는 244건, 대형 화재사고는 9,541건, 소형 화재사고는 264,033건으로 분류되었다. 각 연도별 모델을 사용하였기 때문에 매년 등급별 범위가 다르다는 문제가 있었다. Outlier(이상점)로 인한 문제로 매우 큰 피해를 가진 화재데이터가 있다면 그래프상으로 매우 멀리 위치하게 되고, k-평균 클러스터링의 클러스터링 단계에서 진행되는 유클리드 기준으로 분류되기에 소수의 데이터들만 분류되고, 이들로 인해 중심점이 옮겨지면서 클러스터의 중심점이 원점으로부터 멀어지는 것을 확인할 수 있었다. 하나의 클러스터의 중심점이 원점으로부터 멀어진다면 그로 인해 다른 클러스터들 또한 영향을 받으며,

Table 1. Range of fire losses of critical, major, and minor accidents

		Minor accident			Major accident			Critical accident		
		Fatalities	People injured	Damage of Property	Fatalities	People injured	Damage of Property	Fatalities	People injured	Damage of Property
Non-expert	Min	0	0	0	0	1	0	1	0	0
	Max	0	0	37,600	0	62	5,770	40	33	72,120
Expert	Min	0	0	0	0	0	0	0	0	0
	Max	0	2	1,000	4	9	4,942	40	62	72,120
2007	Min	0	0	0	0	0	0	0	18	5
	Max	0	0	865	6	11	4,293	9	62	493
2008	Min	0	0	0	0	0	0	7	6	7,878
	Max	0	0	854	7	18	37,600	40	10	72,120
2009	Min	0	0	0	0	0	0	0	0	0
	Max	0	0	758	2	3	1,957	10	14	4,941
2010	Min	0	0	0	0	0	0	0	0	0
	Max	0	0	723	2	3	3,897	10	24	5,770
2011	Min	0	0	0	0	0	0	0	0	0
	Max	0	0	646	1	2	2,335	5	10	1,315
2012	Min	0	0	0	0	0	0	0	13	1
	Max	0	0	853	4	14	4,643	9	45	1,531

이로 인해 상대적으로 많은 점들을 포함하게 된다. 하지만 화재데이터 특성상 원점 가까이에 위치한 점들이 90%이상으로 소형 화재사고의 클러스터 중심점은 거의 움직이지 못한다. 따라서 Fig. 3b와 같이 기형적인 분류가 이루어질 수 있다. 하지만 outlier가 없는 경우 Fig. 3c와 같이 피라미드형 피해분류를 이루는 것을 확인할 수 있었다.

비전문가 기준 분류의 결과와 전문가 기준 분류의 결과, k-평균 클러스터링 결과의 범위를 Table 1에 나타내었다. 표는 각기 분류된 소형, 대형, 초대형 화재사고의 사망자수, 부상자수, 백만원 기준 재산피해를 나타내었다. 이 결과를 살펴보면, 비전문가 기준 분류의 결과에서 초대형 화재사고로 분류되어야 할 정도의 피해가 소형 또는 대형으로 분류되거나 대형 화재사고로 분류되어야 할 정도의 피해가 소형 사고로 분류되는 것을 확인하였다. 전문가 기준 분류와 k-평균 클러스터링의 결과는 소형, 대형, 초대형의 피해가 등급별로 역전된 데이터 없이 나옴을 확인할 수 있었다. 하지만 k-평균 클러스터링의 2007, 2008, 2012년 초대형 화재사고를 본다면, outlier로 인해 극소수의 화재사고만이 초대형 화재사고로 분류되었고, 이로 인해 대형 화재사고의 범위가 높게 분류됨을 확인할 수 있었다. 이를 통해 전처리가 필요하다 판단하였지만, 초대형 화재사고의 비율이 전체 화재사고의 1%도 되지 않기 때문에 outlier의 기준이 모호하고, 편차를 통해 outlier를 설정할 시 90% 이상의 화재데이터가 0에 수렴하기 때문에 대형 화재사고마저 outlier로 분

류되는 문제가 있었다. 따라서 outlier를 제외하지 않는다는 가정 하에 분류를 진행하였다. 결과적으로 극소수의 화재사고만 초대형 화재사고로 분류된 3개의 연도를 제외한 2009, 2010, 2011년 결과는 대형 화재사고의 최대 사망자 수가 1-2명, 부상자 수가 2-3명으로 분류되었으며, 이를 통해 연도별 기계학습 기반 분류를 하여도 일정한 피해 범위가 산출됨을 확인하였다. 더불어 k-평균 클러스터링을 통해 산출된 소형 화재사고의 피해 범위가 전문가의 의견에 의해 설정된 대형 화재사고의 기준과 유사한 것을 확인하였다.

NFDS 데이터를 비전문가 기준 분류와 전문가 기준 분류, k-평균 클러스터링 분류를 진행한 결과를 Table 2에 화재사고의 등급별 분포 비율로 나타내었다. 이때 사용된 등급별 분포 비율은 1건의 초대형 화재사고 발생을 기준으로 대형 화재사고와 소형 화재사고의 수로 계산하였다. 비전문가 기준으로 분류한 화재사고는 매년 1건의 초대형 화재사고에 대해 일정한 대형 화재사고 수와 소형 화재사고 수를 보여주며 사고관리 측면 분류에 도움이 될 수 있을 것이라 판단하였다. 전문가 기준 분류의 경우 초대형 화재사고의 수가 적게 분류되어 매년 1건의 초대형 화재사고 대비 3,000건 이상의 소형 화재사고가 존재함을 확인할 수 있었다. 특히 2011년의 경우 단 2건의 초대형 화재사고가 발생하였으며, 이렇게 극소수의 초대형 화재사고를 예측하기에는 큰 무리가 있다고 판단하였다. 따라서 일정한 비율을 보여주지 않는 전문가 기준 분류 방

Table 2. Distribution of ratios of critical accident, major accident, and minor accident

Year	Classification according to non-expert	Classification according to expert	k-means clustering by a single attribute			k-means clustering by 3 attributes
			Fatalities	People Injured	Damage of Property	
2007	1: 4.0: 137.4	1: 81: 5238	1: 47 : 6792	1: 1409: 46472	1: 27: 23914	1: 847: 23093
2008	1: 4.3: 132.8	1: 83: 5431	1: 358: 49272	1: 28 : 857	1: 1 : 49629	1: 971: 23844
2009	1: 4.2: 133.8	1: 118: 7767	1: 25 : 3614	1: 15 : 467	1: 24: 23634	1: 22 : 583
2010	1: 4.5: 159.9	1: 211: 13742	1: 252: 41610	1: 16 : 565	1: 11: 6966	1: 27 : 793
2011	1: 5.1: 188.8	1: 292: 21645	1: 7 : 1559	1: 15 : 577	1: 8 : 1504	1: 13 : 408
2012	1: 5.6: 185.6	1: 49: 3276	1: 44 : 8605	1: 216 : 6991	1: 6 : 3597	1: 214: 5963
Overall	1: 4.5: 152	1: 96: 6422	1: 31 : 4947	1: 26 : 864	1: 9 : 5256	1: 21: 595

법은 예측 변수로 사용하기 힘들다고 판단하였다.

k-평균 클러스터링 분류의 결과로 초대형 화재 사고 수가 2007년 2건, 2008년 2건, 2009년 78건, 2010년 51건, 2011년 104건, 2012년 7건이 도출되었다. 2007년을 살펴보면 부상자수를 이용한 k-평균 클러스터링과 재산피해를 이용한 k-평균 클러스터링에서도 동일하게 소수의 초대형 화재사고로 분류됨을 확인할 수 있다. 그리고 부상자를 이용한 클러스터링은 1건의 초대형 화재사고와 1409건의 대형 화재사고, 46472건의 소형 화재사고로 분류되는 것으로 보아 단 1건의 초대형 화재사고에 많은 부상자가 발생하였음을 알 수 있다. 반면 재산피해의 경우 2건의 초대형 화재사고와 54건의 대형 화재사고, 47828건의 소형 화재사고로 분류되었다. 이는 초대형 화재사고의 클러스터뿐만 아니라 대형 화재사고의 클러스터의 중심점이 높은 값에 위치하는 것을 알 수 있다. 이와 동일하게 2008년과 2012년의 k-평균 클러스터링 결과 outlier에 취약하다는 것을 확인하였고, 사고 관리를 대상으로 k-평균 클러스터링을 사용하기에는 무리가 있을 것으로 판단하였다. 하지만 outlier가 존재하지 않는 3개의 연도에 대해서는 일정한 비율을 보여주며 1년 평균 77건의 초대형 화재사고로 분류하였다.

Fig. 4는 2007-2012년 k-평균 클러스터링 모델에 의해 얻어진 각 연도의 분포 비율을 도표하였다. 이 그래프에 의하면 대형 화재사고의 비율이 증가함에 따라 동일하게 소형 화재사고의 비율이 증가하였다. 이에 대형 화재사고 기준 소형 화재사고 수를 계산한 결과 2007년에는 27.3건, 2008년에는 24.6건, 2009년에는 26.5건, 2010년에는 29.4건, 2011년에는 31.4건, 2012년에는 27.9건으로 매년 일정한 비율을 보였다. 이 결과를 통해 초대형 화

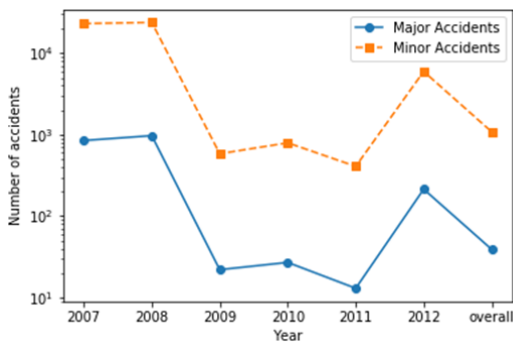


Fig. 4. Number of major and minor accidents per one critical accidents by multi k-means clustering.

재사고가 아닌 대형 화재사고와 소형 화재사고의 분류는 명확히 구분되고 있다고 판단하였다.

4.2. 분류된 화재사고의 주기성 평가

분류의 결과를 종합해 본다면 비전문가 기준 분류의 초대형 화재사고는 1년 평균 289건으로 통계적으로 30시간마다 빈번하게 일어나며, 화재의 크기를 고려하지 않았다. 전문가의 기준의 초대형 화재사고는 1년에 평균 8건이 발생하며, 화재의 크기는 대규모 화재로 분류되었다. 기계학습 기반 분류의 초대형 화재사고는 1년 평균 40건이 발생하였으며, outlier가 없는 기간에 대해서는 1년 평균 77건으로 통계적으로 111시간마다 발생한다. 그리고 기계학습 기반 분류의 경우 데이터를 기반을 둔 분류이기에 시대의 추세를 반영할 수 있다고 판단하였고,

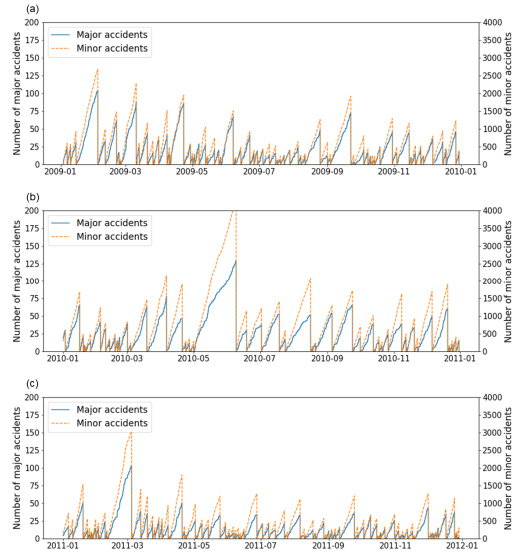


Fig. 5. Plot of accumulated number of major and minor accidents after a critical accident in 2009-2011.

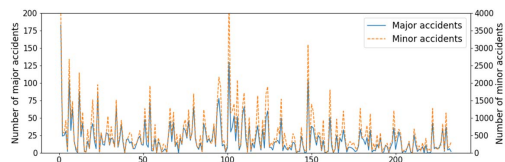


Fig. 6. Periodicity of major and minor accidents per a critical accident in 2009-2011.

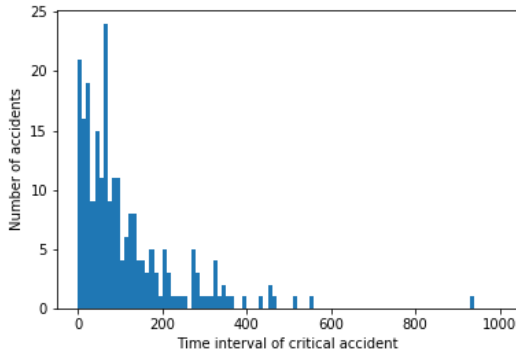


Fig. 7. Histogram of time interval of critical accident in 2009-2011.

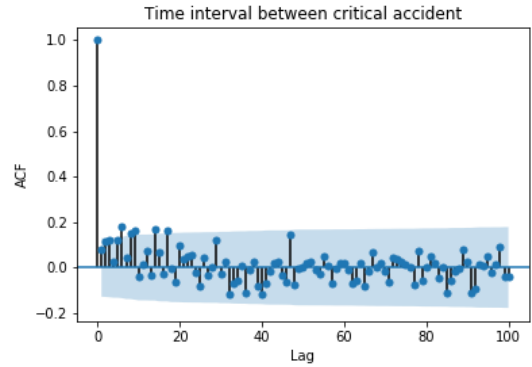


Fig. 9. Autocorrelation of time interval between critical fire accidents.

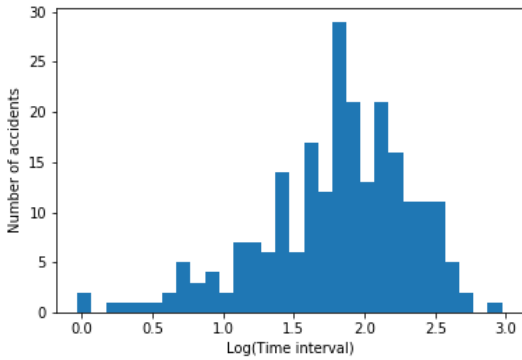


Fig. 8. Histogram on log axis of time interval of critical accident in 2009-2011.

매년 일정한 초대형 화재사고와 대형 화재사고, 소형 화재사고의 비율을 보여주어 화재사고가 반복성을 가지고 있다 판단하였다. 따라서 만약 화재사고의 반복성이 일정한 추세나 주기성을 가지고 있다면 단순한 화재사고 모니터링을 통해 예측을 할 수 있고, 추세나 주기성이 없더라도 일정한 비율을 갖는 대형 화재사고와 소형 화재사고의 발생 수를 이용해 초대형 화재사고를 예측할 수 있을 것이라 판단하였다. 따라서 이를 검증하기 위해 k-평균 클러스터링 모델로 분류된 초대형 화재사고의 주기성 탐색을 시도하였다. 2009-2011년 화재데이터의 초대형 화재사고 1건당 대형 화재사고의 수와 소형 화재사고의 수를 Fig. 5와 같이 도출되었다. 여기서 Fig. 5을 살펴보면 x축은 시간을 나타내며, 왼쪽의 y축은 대형 화재사고의 수로써, 파란색 실선을 이용해 대형 화재사고의 수를 도표 하였고, 오른쪽의 y축은 소형 화재사고의 수로써, 주황색 점선을 이

용해 소형 화재사고의 수를 도표 하였다. 그래프를 분석해 보면, 2009년과 2011년의 그래프에서 어느 정도의 주기성을 갖고 초대형 화재사고가 발생하는 것을 확인할 수 있었다. 하지만 인적사고라는 특성으로 인해 뚜렷한 패턴이나 계절적 요인을 보이지 않았다.

Fig. 5의 결과에서 초대형 화재사고 직전까지 누적된 대형 화재사고와 소형 화재사고 수를 도출한다면 Fig. 6와 같이 나타난다. Fig. 6을 살펴보면, x축은 2009-2011년에 발생한 초대형 화재사고들의 번호를 나타내며 왼쪽의 y축은 대형 화재사고의 수로써, 파란색 실선을 이용해 대형 화재사고의 수를 도표 하였고, 오른쪽의 y축은 소형 화재사고의 수로써, 주황색 점선을 이용해 소형 화재사고의 수를 도표 하였다. 하나의 초대형 화재사고가 발생하기 까지 대형 화재사고의 수는 평균 19.6건으로, 23.17의 편차를 보여주었고, 소형 화재사고의 수는 평균 565.8건으로, 641.24의 편차를 보여주었다.

초대형 화재사고가 발생하기까지의 대형 화재사고의 수와 소형 화재사고의 수는 초대형 화재사고 간의 간격에 매우 큰 영향을 미친다. 대형화재사고와 소형 화재사고가 일정 간격으로 발생한다는 가정 하에 이들 사이에 대형 화재사고와 소형 화재사고가 많다는 것은 그만큼 시간이 많이 지나갔다고 할 수 있다. 따라서 2009-2011년에 발생한 초대형 화재사고들의 시간 간격을 도출한다면 Fig. 7와 같이 나타난다. Fig. 7을 살펴보면 x축은 초대형 화재사고간의 시간 간격으로써, 한 구간에 10시간을 포함하도록 도표 하였다. 왼쪽의 y축은 구간별 빈도수를 도표 하였다. 이 히스토그램에 의하면 초대형 화재사고 시간 간격의 평균은 112시간, 표준편차는 120.52, 분산은 14525.06, 왜도는 2.5, 첨도 10.2를

보여주며 정규 분포에 따르지 않고, 왼쪽으로 기울어진 분포를 보였다. 가장 많이 발생한 구간은 60-70시간으로 총 23건을 보여주었고, 특이치는 936시간을 보여주었다. 이 분포를 정규분포로 만들기 위해 완곡화 방법을 사용하여 시간 간격에 로그 함수를 사용한다면 Fig. 8와 같이 나타난다. 왜도는 2.5에서 -0.8로 줄어들었으며, 첨도는 10.2에서 0.78로 줄어들며 보다 정규분포에 가깝게 조정된 것을 확인할 수 있었다. 이 결과의 평균은 1.80, 표준편차는 0.54, 분산은 0.29를 나타내며, 이를 통해 초대형 화재사고의 시간 간격을 예측할 수 있다고 판단된다.

최종적으로 자기상관계수 함수를 이용하여 초대형 화재사고간의 무작위성을 분석하였다. 초대형 화재사고가 발생 후 다음 초대형 화재사고 발생까지 걸린 시간의 자기상관계수를 도출하면 Fig. 9와 같이 나타난다. 위의 결과로 시차가 증가함에 따라 무작위로 상관함수가 변화함을 알 수 있지만, 20개의 시차 이하에서 신뢰구간을 넘어가는 것으로 보아 낮은 시차에서 뒤의 초대형 화재사고에 영향을 줌으로써 초대형 화재사고 발생 시점이 완벽한 무작위성을 보이지 않는다고 판단된다. 이러한 이유로 초대형 화재사고가 발생한 이후 사람들의 경각심으로 발생하지 않거나, 화재사고가 확대되어 2차, 3차의 피해를 입힐 수 있고, 제도적으로 개선될 경우 바로 개선되기보다 어느 정도 기간이 지나야 적용되기에 낮은 시차에서 보다 높은 자기상관관계를 보이는 것으로 판단된다. 따라서 이러한 무작위성으로 인해 단순한 화재사고 모니터링을 통한 초대형 화재사고 예측보다 고급 예측 기법을 통한 초대형 화재사고 예측이 가능할 것이라고 판단하였다.

V. 결론

본 연구에서는 사고발생의 실시간 모니터링 기반 초대형사고의 예측을 위한 사고 분류 세분화와 주기성 분석을 실시하였다. 2007-2012년 국가화재정보시스템을 이용하여 아래와 같은 화재사고 등급별 분류를 진행하였다.

1. 하인리히의 법칙을 원용한 비전문가 기준의 분류
 2. 소방관의 직관으로 설정된 전문가 기준의 분류
 3. 화재데이터에 기반한 k-평균 클러스터링 분류
- 분류 결과 비전문가 기준의 분류는 매년 유사한 등급별 비율을 보여주었지만, 화재피해별 등급 분류에는 적합하지 않고 일어나는 초대형 화재사고

횟수가 잦아 예측에 적합한 초대형 화재사고의 기준이 아님을 보였다. 전문가 기준의 분류는 명확한 화재피해의 분류가 되었지만, 초대형 화재사고가 1년 평균 8건으로 분류되며 극소수의 사고를 예측하기에는 큰 무리가 있다고 판단하였다. 기계학습 기반 분류는 outlier로 인해 매년 다른 화재피해 범위를 보였지만, outlier가 없는 기간의 경우 매년 유사한 등급별 비율을 보이며 적절한 초대형 화재사고 횟수를 보여주며 예측에 적합한 초대형 화재사고의 기준으로 판단되었다. 따라서 기계학습 기반 분류된 초대형 화재사고의 주기성을 분석한 결과 사고의 무작위성으로 인해 일정한 주기나 추세는 보이지 않았지만, 로그함수를 사용한 초대형 화재사고의 시간 간격이 정규분포에 가깝게 나타났으며, 자기상관계수 함수를 통해 어느 정도 무작위성은 존재하지만 최근 발생한 초대형 화재사고가 다음 초대형 화재사고 발생 시점에 영향을 주지만 어느 정도 시점이 지날 경우 영향이 매우 적어지는 것을 보였다. 이를 통해 초대형 화재사고에 대한 발생 시점을 예측하는 것이 가능하다고 판단되며, 예측된 초대형 화재사고의 발생 시점으로 비상대응을 실시한다면 화재피해를 최소화할 수 있을 것이라 기대된다.

감사의 글

본 연구는 산업통상자원부 스마트디지털엔지니어링전문인력양성 사업의 연구비지원(P0008475-G02P04570001901)에 의해 수행되었습니다. 이에 감사드립니다.

REFERENCES

- [1] D. Park, "Characteristics of Risk in Modern Society and Desirable National Safety and Disaster Management Systems", Human Right and Justice, 452, 6-19 (2015)
- [2] Federal Emergency Management Agency, United States Fire Administration. About the National Fire Incident Reporting System. Available: <https://www.usfa.fema.gov/downloads/pdf/publications/fa-290.pdf>. (2019)
- [3] Public Safety Canada, All-Hazards Risk Assessment (AHRA), [Online]. Available: <https://www.publicsafety.gc.ca/cnt/mrgnc-mngmnt/mrgnc-prprdnss/ll-hzrds-rsk-sssmnt-en.aspx>
- [4] National Emergency Management Agency, National Fire Data System (NFDS), [Online]. Available:

- lable: <http://www.nfds.go.kr/>
- [5] H. W. Heinrich, *Industrial accident prevention: a scientific approach*, McGraw-Hill (1959)
 - [6] T. McSween and D. J. Moran, "Assessing and Preventing Serious Incidents with Behavioral Science: Enhancing Heinrich's Triangle for the 21st Century", *Journal of Organizational Behavior Management*, 37, 283-300 (2017)
 - [7] O. Maimon, and L. Rokach, "Clustering methods." *Data mining and knowledge discovery handbook*. Springer US (2005)
 - [8] J. MacQueen, "Some methods for classification and analysis of multivariate observations", *Proc. 5th Berkeley Symp. on Math. Statist. and Prob.*, 1, 281-297 (1967)
 - [9] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A K-Means Clustering Algorithm", *Journal of the Royal Statistical Society*, 28, 100-108 (1979)
 - [10] A. Papoulis, *Probability, Random variables and Stochastic processes*, McGraw-Hill (1991)
 - [11] Box, G. E. P., and Jenkins, G. *Time Series Analysis: Forecasting and Control*, Holden-Day (1976)
 - [12] 박정훈. (2019년 11월 28일). 전화인터뷰
 - [13] 소방방재청, "화재조사 및 보고규정", In: 소방방재청훈령 제229호, (2010)