# 소방방재

# 인공신경망을 활용한 기상 및 화재 데이터기반의 화재발생 확률 예측

Fire Prediction based on Weather and Fire Data using Artificial Neural Network

홍석영\* · 조성현\*\* · 김민수\*\*\* · 문잌\*\*\*\*

Hong, Seokyoung\*, Cho, Sunghyun\*\*, Kim, Minsu\*\*\*, and Moon, Il\*\*\*\*

#### **Abstract**

Today, more than 40,000 fire accidents occur every year, resulting in considerable casualties and property damage. This study develops an artificial neural network model to predict the probability of fire in Seoul City using fire and weather data collected from January 1, 2008, to December 31, 2017. The integrated dataset combines weather data collected every hour with fire data to determine the probability of a fire; 70,484 data are used for artificial neural network learning while 30,459 are used for evaluation. The analysis calculates the probability of a fire between 0 and 100% based on the information yielded by an artificial neural network comprising three hidden layers, with the month, day, hour, temperature, humidity, and wind speed used as variables. Relative Absolute Error (RAE) is used to evaluate the performance of the model, which reveals the neural network model's superiority over the decision tree method.

Key words: Fire Prediction, Weather and Fire Data, Artificial Neural Network, Relative Absolute Error

#### 요 지

최근 국내의 화재 발생 현황으로 매 년 40,000건 이상의 화재사고가 꾸준히 발생하고 있으며 이로 인해 상당한 인명피해와 재산피해가 발생하고 있다. 본 논문은 이에 대한 예측 및 대응시스템 구축을 위해 2008년 1월 1일부터 2017년 12월 31일까지 수집된 서울시 화재데이터와 기상데이터를 활용하여 서울시의 화재발생 확률을 예측하는 인공신경망 모델을 개발하였다. 매 시간 단위로 수집된 기상데이터와 화재발생 여부를 알 수 있는 화재데이터를 통합한 통합데이터를 생성하였고 70,484건의데이터를 인공신경망 학습에, 30,459건의 데이터를 평가에 사용하였다. 본 논문의 결과물은 월, 일, 시, 온도, 습도, 풍속 변수를 사용하여 3개의 은닉층으로 구성된 인공신경망을 거쳐 0~100% 사이의 화재발생 확률을 계산한다. 모델 성능 평가에는 상대 절차 오차(Relative Absolute Error, RAE)가 사용되었으며 인공신경망 모델이 의사결정트리에 비하여 성능이 우수한 것으로 나타났다.

핵심용어: 화재 예측, 기상 및 화재 데이터, 인공신경망, 상대절대오차

<sup>\*</sup>정회원, 연세대학교 화공생명공학과 박사과정(E-mail: hsyoung224@yonsei.ac.kr)

Member, Ph.D. Cadidate, Department of Chemical and Biological Engineering, Yonsei University

<sup>\*\*</sup>연세대학교 화공생명공학과 박사과정(E-mail: muckre@yonsei.ac.kr)

Ph.D. Cadidate, Department of Chemical and Biological Engineering, Yonsei University

<sup>\*\*\*</sup>연세대학교 화공생명공학과 석사과정(E-mail: minsu77@yonsei.ac.kr)

Master Cadidate, Department of Chemical and Biological Engineering, Yonsei University

# 1. 서 론

최근 10년 간 매년 40,000건 이상의 화재사고가 발생하고 있으며(Fig. 1), 이 사고로 인하여 매일 6명의 인명피해 및 127만원의 재산피해가 발생하고 있다. 계속해서 발전하는 소방시스템에도 불구하고 화재사고 발생은 줄어들지 않고 있으며 화재사고 특성상 사고발생을 줄이기 위한 명확한 대책수립이 불가능하다. 합리적인 소방정책대안을 수립하기 위해 국가화재정보시스템(NFDS)은 2007년도부터 각종화재정보 데이터를 축적하기 시작하였으며, 데이터를 활용한 화재정보 및 화재통계 분석에 관한 연구들이 활발하게 진행되고 있다.

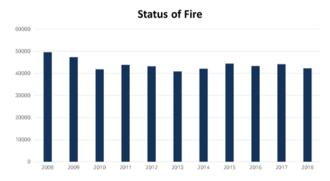


Fig. 1. Annual Frequency of Fire Accident (2008~2018)

NFDS는 1년 6개월 동안 수집한 데이터를 기반으로 화재 위험 예측모델을 개발하였고 화재의 위험정도를 표현하기 위해 화재위험지수(Chang et al., 2008)를 개발하였다. 화재 위험지수를 회귀분석을 기반으로 수학적 모델을 생성하여 나타내었는데 회귀분석 특성상 화재 위험을 화재 요인 간의 선형관계로 표현하여 화재의 복잡한 현상을 제대로 표현하지 못하는 한계가 있다.

Ryu and Kim (2012)은 화재 발생에 영향을 미치는 요인

분석을 알아보기 위하여 의사결정트리로 화재 예측 모델을 생성하여 습도가 화재와 연관성이 높다는 것을 연구하였으며, Kim et al. (2013) 은 의사결정트리를 활용하여 지역별 화재확률을 예측하는 모델을 생성하여 Brier 점수를 통해 예측 모델의 성능을 검증하였다. 하지만 의사결정트리를 활용한 예측 모델은 단계별로 한 개의 화재 요인변수만을 고려하여 분류한다는 점에서 화재 요인변수 간의 관계를 고려할 수 없다는 한계가 있다.

따라서 본 연구에서는 요인변수 간 복잡한 관계를 고려할 수 있는 인공신경망 모델을 활용하여 화재확률을 예측할수 있는 모델을 구축하였다. 2008년부터 2017년까지의화재데이터와 기상데이터를 통합하여 매 시간 화재발생여부를 나타내는 데이터를 생성하고, 인공신경망 모델을 학습 시킨 후 화재 발생확률 예측 결과와 실제 화재발생여부를 비교하여 모델의 성능을 검증하였다. 또한 본 연구의결과를 선행연구와의 정확도 비교를 통해 본 연구의 우수성을 입증하였다.

# 2. 인공신경망을 활용한 화재확률 예측

본 연구에서 제안하는 화재확률 예측모델은 매 시간 측정 되는 날씨 정보를 바탕으로 인공신경망 예측모델을 활용하 여 화재발생 확률을 계산한다. 본 연구는 데이터 수집, 데이터 통합, 예측모델 생성 및 확률계산 단계로 구성되며(Fig. 2) 2008년부터 2017년까지 10년 간 축적된 화재데이터 및 기상 데이터가 사용되었다.

생성된 예측모델은 월, 일, 시, 온도, 습도, 풍속 변수를 사용하며, 매 시간 해당 지역의 화재발생 확률(0%~100%)을 알 수 있어 실시간 화재 대비가 가능하다.

#### 2.1 데이터 수집

예측모델 생성을 위해 서울시 화재 데이터와 서울시 기상

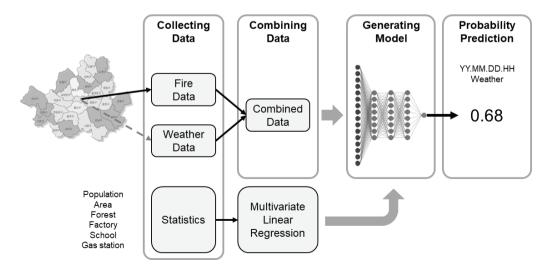


Fig. 2. Prediction Procedure of Fire Accident

데이터, 그리고 서울시의 인구, 면적, 토지이용 등을 나타내는 통계데이터를 수집하였다. 데이터 수집기간은 2008년 1월 1일부터 2017년 12월 31일까지 총 10년이며 화재 데이터는 50,864건, 기상 데이터는 87,686건, 통계데이터는 10건(년단위)을 수집하였다.

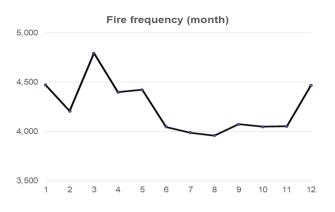
화재데이터는 국가화재정보센터(NFDS)를 통해 수집하였으며 화재사고 발생시간(년, 월, 일, 시) 및 화재유형, 화재 발생 지역, 발화요인, 발화장소 등의 정보를 포함하고 있다. 기상데이터는 기상청을 통해 수집하였으며 매 시간 측정한 온도, 습도, 풍속, 풍향, 강우량 등의 정보를 포함하고 있다. 마지막으로 통계데이터는 서울시의 인구, 면적, 토지 이용(임야, 공장, 학교, 주유소) 현황을 포함하고 있다.

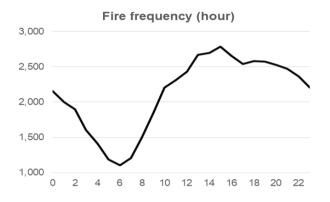
수집한 화재 및 기상 데이터의 다양한 정보 중에서 예측모델 생성에 사용할 주요 변수들을 선정하여 두 종류의 데이터를 통합하기 위한 전처리 작업을 진행하였다. 변수는 데이터분석을 기반으로 선정하였으며 분석결과는 Fig. 3에 나타내었다. 시간에 따른 분석결과를 보면 화재는 1월부터 12월사이 중 3월에 가장 많이 발생하며 7~8월 사이에 적게 발생한다. 또한 0시부터 23시 사이 중 점심시간 13~15시에 많이발생하여 새벽시간 5~7시에 적게 발생한다. 다음으로 날씨에 따른 분석결과로 화재는 기온이 19 ℃~28 ℃ 사이일때 많이 발생하고, 습도가 50~65일때 많이 발생한다. 따라서화재 데이터에서는 년, 월, 일, 시를 화재 요인변수로 선정했으며 기상 데이터에서는 년, 월, 일, 시, 온도, 습도, 풍속정보를 변수로 선정하였다. 두 종류의 데이터는 Tables 1 and 2에 나타내었다.

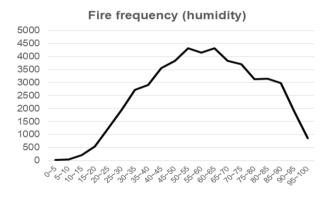
## 2.2 데이터 통합

예측모델을 생성하기 위한 인공신경망을 학습시키기 위해서는 주어진 상황(월, 일, 시, 온도, 습도, 풍속)에서 화재발생 여부를 알 수 있는 데이터가 필요하다. 화재 데이터는 발생한 화재에 대한 데이터만 포함하고 있어 발생하지 않은 상황에 대한 정보를 학습할 수 없다. 따라서 수집된 두 종류의데이터를 바탕으로 연속적인 시간축에서 화재 발생여부를 나타내는 통합데이터 생성이 필요하다.

화재 데이터와 기상 데이터에서 동일한 시간에 수집된데이터를 병합하여 연속된 시간축에서 발생한 화재를 'Yes'로 표시한 후, 화재가 발생하지 않은 나머지 시간대에 'No'를 표시한다. 화재 발생시간과 기상 데이터 측정시간이 동일한 데이터는 Tables 2 and 3에 음영으로 나타내었다. 또한, 같은 시간대에 발생한 2회 이상의 화재건수(ex. 2008년 1월 1일 0시, 2017년 12월 31일 17시)에 대하여 발생한화재 횟수만큼 통합데이터에 반영하였다. 통합데이터는총 100,943건의 데이터를 포함하며 발생화재 50,875건과미발생화재 50,068건으로 구성된다. 발생과 미발생사이의비율이 50:50을 이루고 있으므로 균형 있는 데이터구성이다.







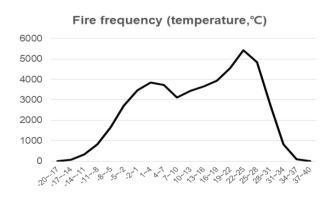


Fig. 3. Frequency of Fire in Seoul According to Month, Hour, Humidity, and Temperature

Table 1. Fire Data in Seoul

No.	Year	Month	Day	Hour	District
1	2008	1	1	0	Mapo
2	2008	1	1	0	Yongsan
3	2008	1	1	1	Jongno
4	2008	1	1	3	Seongdong
50,861	2017	12	31	17	Seongdong
50,862	2017	12	31	17	Nowon
50,863	2017	12	31	22	Songpa
50,864	2017	12	31	23	Seodaemun

Table 2. Weather Data in Seoul

No.	Year	Month	Day	Hour	Temperature $(^{\circ}C)$	Humidity (%)	Wind Speed
1	2008	1	1	0	-7	52	3.5
2	2008	1	1	1	-7.3	52	4.9
3	2008	1	1	2	-7.5	52	4.1
4	2008	1	1	3	-7.8	52	3.9
•••							
87,680	2017	12	31	17	1.7	30	2.9
87,681	2017	12	31	18	0.7	31	2.3
87,682	2017	12	31	19	0	31	2.2
87,683	2017	12	31	20	-1	32	0.9
87,684	2017	12	31	21	-1.6	35	1
87,685	2017	12	31	22	-2.1	36	1.7
87,686	2017	12	31	23	-2.6	40	0.6

Table 3. Combined Data in Seoul

					Weather Data			Fire Data
No.	Year	Month	Day	Hour	Temperature	Humidity	Wind Speed	Fire
					(℃)	(%)	(m/s)	Accident
1	2008	1	1	0	-7	52	3.5	Yes
2	2008	1	1	0	-7	52	3.5	Yes
3	2008	1	1	1	-7.3	52	4.9	Yes
4	2008	1	1	2	-7.5	52	4.1	No
5	2008	1	1	3	-7.8	52	3.9	Yes
100,936	2017	12	31	17	1.7	30	2.9	Yes
100,937	2017	12	31	17	1.7	30	2.9	Yes
100,938	2017	12	31	18	0.7	31	2.3	No
100,939	2017	12	31	19	0	31	2.2	No
100,940	2017	12	31	20	-1	32	0.9	No
100,941	2017	12	31	21	-1.6	35	1	No
100,942	2017	12	31	22	-2.1	36	1.7	Yes
100,943	2017	12	31	23	-2.6	40	0.6	Yes

#### 2.3 선형회귀 모델 생성

2008년부터 2017년까지 수집한 통계데이터(Table 4)는 1년 동안 서울시에 발생한 화재빈도와 서울시 통계의 관계 분석을 위해 사용되었다. 서울시의 통계 값과 화재빈도사이의 관계는 다중선형회귀(Multivariate Linear Regression)모델링을 통해 나타내었으며 식은 Eq. (1)과 같이 나타났다.

$$fire frequency (annual) = 179880.6196 - 0.00361 \times Population -0.00017 \times Forest -0.00787 \times Factory -0.00336 \times School -0.01757 \times Gas station$$

$$(1)$$

선형회귀모델의 평균절대오차는 37.85이며 이 모델을 2015~2017년도의 연간 화재빈도를 예측하는 데 활용하였다. 예측된 연간 화재빈도는 1년, 8760시간으로 나누어 시간당 평균 화재발생 확률로써 인공신경망의 출력층 편향값으로 사용되었다.

#### 2.4 인공신경망 모델 생성

앞에서 구축된 통합데이터는 인공신경망 모델을 학습하 는 데 사용된다. 인공신경망은 기계학습의 통계학적 학습 알고리즘 중의 하나로 여러 층의 뉴런이 연결되어 있는 구조를 나타내며 연결 간의 가중치를 갱신하며 학습이 이루어지는 특징이 있다(Neal, 1996). 신경망은 크게 입력층 (Input Layer), 은닉층(Hidden Layer), 출력층(Output Layer) 으로 구성되며 Fig. 4는 인공신경망 구조의 간단한 예를 나타낸다. 16개의 변수를 입력층에서 입력받고 8개의 노드 로 이루어진 은닉층 안의 활성함수 $h_k(x)$ 를 거친 후 출력층 의 출력 값이 최종 확률 값으로 계산되는 형태이다. 본 연구에서는 인공신경망을 35개의 입력노드와 3개의 은닉층 그리고 확률계산을 위한 1개의 출력노드로 구성하였다. 은닉층은 경험에 의한 최적화를 통해 3개의 층을 사용하였 으며 각 층은 64, 32, 16개의 노드로 구성하였다. 은닉층의 활성함수(Activation function)는 ReLU 함수를 사용하였고, 확률계산을 위한 출력층의 함수는 Sigmoid 함수를 사용하 였다.



Vacr	Year Frequency	Den 1.dien	Area Forest	Land use			
Y ear		Population		Factory	School	Gas station	
2008	5579	10200827	605252502	148806102	2914056	23471545	574659
2009	5420	10208302	1210521516	148528585	2896177	23581217	575887
2016	5000	9930616	605250743	140529457	2970959	24219268	583037
2017	5000	9857426	605196264	139957858	2939243	24426922	571253

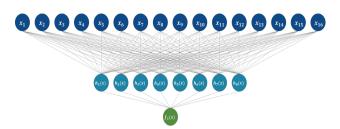


Fig. 4. Structure of Artificial Neural Network

인공신경망은 오차역전파(Back-propagation) 알고리즘에 의해 학습되며(Hagan and Menhaj, 1994) 인공신경망의효율적인 학습을 위해 통합데이터의 변수들을 Table 5에따라 범주화(Categorization)하여 범주형 변수(Categorical variable)로 학습을 진행하였다. '월' 변수는 계절적 특성을반영하여 봄, 여름, 가을, 겨울 4가지 범주로 나누었고, '일' 변수는 4일마다 하나로 묶고 '시' 변수는 3시간마다 하나로묶어 각각 8가지 범주로 나누었다. '온도', '습도', '풍속' 변수는 데이터 분포도에 따라 매우 낮음, 낮음, 보통, 높음, 매우 높음 5가지 범주로 나누었다. 본 연구에서는 2008년부터 2014년까지 통합데이터 70,484건을 학습데이터로, 2015년부터 2017년까지 통합데이터 30,459건을 평가데이터로사용하였다.

학습이 완료된 신경망모델은 월, 일, 시, 온도, 습도, 풍속데이터를 입력받아 서울시의 화재발생확률을 0~100% 값으로 나타낸다. 예측된 확률을 실제 화재발생 여부와 비교를 통해 모델의 정확성을 평가한다.

# 3. 결 과

본 연구에서는 앞서 선행된 의사결정트리 모델과의 결과 비교를 위해 상대 절대 오차(RAE)를 모델의 성능 검증 지표로 선정하였다. RAE는 예측모델의 성능을 검증할 때 많이 쓰이는 지표이며 모델의 의존성, 작은 변화에 대한 민감도 등의 측면에 있어서 우수한 평가지표로써 사용된다 (Armstrong and Collopy, 1992).

Table 5. Classification of Variables

Variable	Number of Class	Class	
Month	4	Spring, Summer, Fall, Winter	
Day	8	1~4, 5~8, 9~12, 13~16, 17~20, 21~24, 25~28, 29~31	
Hour	8	0~2, 3~5, 6~8, 9~11, 12~14, 15~17, 18~20, 21~23	
Temperature	5	Very Low, Low, Normal, High, Very High	
Humidity	5	Very Low, Low, Normal, High, Very High	
Wind Speed	5	Very Low, Low, Normal, High, Very High	

상대절대오차는 Eq. (2)과 같다.

$$RAE = \frac{\sum_{i=1}^{n} |p_i - a_i|}{\sum_{i=1}^{n} |\bar{a}_i - a_i|}$$
 (2)

n은 평가데이터의 개수,  $p_i$ 는 예측모델에서 계산된 화재 발생 확률값,  $a_i$ 는 실제 화재발생 여부(발생:1, 미발생:0),  $a_i$ 는 단순통계에 의해 계산된 화재발생 확률값을 나타낸다. 상대절대오차의 값이 1보다 크면 예측모델의 정확도가 낮은 것을 의미하고, 1보다 작으면 예측모델의 정확도가 상대적으로 높다는 것을 의미한다.

Table 6에서는 Ryu et al. (2014) 연구의 의사결정트리모델과의 상대절대오차를 비교한 결과를 보여준다. 의사결정트리모델리 모델의 서울시 상대절대오차는 0.9594를 나타내며, 본연구의 예측모델 상대절대오차는 좀 더 우수한 정확도를 보여주었다. 서울시 통계데이터(인구, 면적, 토지이용)를 사용하지 않고 월, 일, 시, 온도, 습도, 풍속으로만 예측을 진행한 인공신경망 모델은 상대절대오차 값으로 0.9492 값을 나타내었으며, 통계데이터를 사용하였을 시 보다 개선된 0.9488 값을 나타내었다.

Table 6. RAE Comparison between Decision Tree and ANN

Region	Prediction Method	RAE
	Decision Tree	0.9594
Seoul	ANN model (without statistics)	0.9492
	ANN model (with statistics)	0.9488

# 4. 결 론

본 논문에서는 서울시 기상데이터와 화재데이터를 결합 한 화재발생 통합데이터를 생성하였으며, 7년 데이터를 사 용해 인공신경망을 학습시킨 후 3년 데이터로 성능평가를 진행하였다. 화재는 여러 가지 다양한 변인들의 상호작용을 통해 발생하는 사고이므로 단순하게 날짜와 기상데이터만으로 예측하기에는 한계가 존재한다. 따라서 본 연구에서는지역별 통계데이터(인구, 면적, 토지이용)를 고려하여 정확도를 높인 정교한 예측모델을 개발하였다. 서울시 통계데이터를 활용하여 선형회귀모델을 개발하였으며, 모델의 결과를 인공신경망 학습에 사용하였다. 월, 시를 예측변수로사용하는 기존의 의사결정트리와 달리 월, 일, 시, 통계데이터를 모두 예측 변수에 반영할 수 있는 장점을 가지고 있어보다 정확도 높은 예측을 할 수 있다. 그 결과, 서울시 화재발생 확률 예측 상대절대오차가 0.9594에서 0.9488로 개선된정확도를 보여주며 더욱 우수한 모델임을 증명하였다.

하지만 통계데이터의 단위가 연 단위인 점을 고려하면 시 단위 예측 모델에 정밀하게 적용될 수 없는 한계가 있다. 따라서 이후 연구에서는 보다 정확한 예측모델 개발을 위해 서, 시 단위로 수집할 수 있는 통계 및 다양한 화재요인들을 추가적으로 반영하여 정확도 향상을 위한 연구를 진행 할 계획이다.

## 감사의 글

본 논문은 2017년도 정부(과학기술정보통신부)의 재원으로 한국연구재단 – 재난안전플랫폼기술개발사업의 지원을받아 수행된 연구임(No. NRF-2017M3D7A1085361).

## References

Armstrong, S., and Collopy, F. (1992). Error measures for generalizing about forecasting methods: Empirical comparison. *International Journal of Forecasting*, Vol. 8, No. 1, pp. 69-80.

Hagan, M.T., and Menhaj, M.B. (1994). Training feedforward networks with the Marquardt algorithm. *IEEE Transactions on Neural Networks*, Vol. 5, No. 6, pp. 989-993.

Kim, Y.J., Ryu, J.W., Song, W.M., and Kim, M.W. (2013). Fire probability prediction based on weather

- information using decision tree. *Journal of KIISE: Software and Applications*, Vol. 40, No. 11, pp. 705-715.
- Chang, J.Y., Kim, H.J., and Lee, S.K. (2008). Research on the development fire risk assessment model. National Emergency Management Agency.
- Neal, R. (1996). *Bayesian learning for neural networks*. Springer-Verlag New York, pp. 10-19.
- Ryu, J.W., and Kim, E.J. (2012). Analysis of regional weather factors affecting fire using decision trees. *Proceedings of 2012 Annual Fall Conference*, Korean

Institute of Fire Science & Engineering, pp. 263-266. Ryu, J.W., Kim, Y.J., Kim, E.J., and Kim, M.W. (2014). A generation method of a fire probability prediction model based on weather forecast. *Journal of KIISE:*Computing Practices and Letters, Vol. 20, No. 2, pp. 68-79.

