

PM10 예측 성능 향상을 위한 이진 분류 모델 비교 분석

정용진¹ · 이종성¹ · 오창현^{2*}

Comparative Analysis of the Binary Classification Model for Improving PM10 Prediction Performance

Yong-Jin Jung¹ · Jong-Sung Lee¹ · Chang-Heon Oh^{2*}¹Graduate Student, Department of Electrical, Electronics and Communication Engineering, Korea University of Technology and Education(KOREATECH), Cheonan, 31253 Korea^{2*}Professor, Department of Electrical, Electronics and Communication Engineering, Korea University of Technology and Education(KOREATECH), Cheonan, 31253 Korea

요 약

미세먼지 예보에 대한 높은 정확도가 요구됨에 따라 기계 학습의 알고리즘을 적용하여 예측 정확도를 높이려는 다양한 시도들이 이루어지고 있다. 그러나 미세먼지의 특성과 불균형적인 농도별 발생 비율에 대한 문제로 예측 모델의 학습 및 예측이 잘 이루어지지 않는다. 이러한 문제를 해결하기 위해 특정 농도를 기준으로 미세먼지를 저농도와 고농도로 구분하여 예측을 수행하는 등 다양한 연구가 진행되고 있다. 본 논문에서는 미세먼지 농도의 불균형 특성으로 인한 예측 성능 향상의 문제를 해결하기 위한 미세먼지 농도의 이진 분류 모델을 제안하였다. 분류 알고리즘 중 logistic regression, decision tree, SVM 및 MLP를 이용하여 PM_{10} 에 대한 이진분류 모델들을 설계하였다. 오차 행렬을 통해 성능을 비교한 결과, 4가지 모델 중 MLP 모델이 89.98%의 정확도로 가장 높은 이진 분류 성능을 보였다.

ABSTRACT

High forecast accuracy is required as social issues on particulate matter increase. Therefore, many attempts are being made using machine learning to increase the accuracy of particulate matter prediction. However, due to problems with the distribution of imbalance in the concentration and various characteristics of particulate matter, the learning of prediction models is not well done. In this paper, to solve these problems, a binary classification model was proposed to predict the concentration of particulate matter needed for prediction by dividing it into two classes based on the value of $80\mu g/m^3$. Four classification algorithms were utilized for the binary classification of PM_{10} . Classification algorithms used logistic regression, decision tree, SVM, and MLP. As a result of performance evaluation through confusion matrix, the MLP model showed the highest binary classification performance with 89.98% accuracy among the four models.

키워드 : 기계 학습, 인공신경망, 미세먼지, 분류 모델**Keywords** : Machine learning, Artificial neural network, Particulate matter, Classification

Received 29 September 2020, Revised 7 October 2020, Accepted 14 October 2020

* Corresponding Author Chang-Heon Oh(E-mail:choh@koreatech.ac.kr, Tel:+82-41-560-1187)

Professor, Department of Electrical, Electronics and Communication Engineering, Korea University of Technology and Education(KOREATECH), Cheonan, 31253 Korea

Open Access <http://doi.org/10.6109/jkiice.2021.25.1.56>

print ISSN: 2234-4772 online ISSN: 2288-4165

© This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Copyright © The Korea Institute of Information and Communication Engineering.

I. 서 론

미세먼지는 다양한 크기, 모양, 성분으로 이루어진 물질이다. $10\mu g$, $2.5\mu g$ 이하의 크기에 따라 PM_{10} , $PM_{2.5}$ 로 구분되는 미세먼지는 심혈관, 호흡기, 뇌혈관 질환 등의 건강에 영향을 준다. 때문에 미세먼지는 위험성 물질로 분류되었으며, 사회 구성원들의 활력 저하 원인으로 분석되고 있다[1-4]. 이와 같은 영향을 최대한 피하기 위해 ‘ 좋음’, ‘ 보통’, ‘ 나쁨’, ‘ 매우 나쁨’의 4단계로 구분된 AQI(air quality index)를 기준으로 제공하는 정보를 확인하는 것이 일상이 되었다. 한국의 미세먼지 예보 정확도는 2015년 약 60%의 수준을 나타내었으며, 기상청의 예보 프로세스는 연단위로 약 80%의 예보 정확성을 가지고 있다. 하지만 예보관의 경험에 의한 판단이 적용된 정보이며, 실제 미세먼지 예측 모델의 정확도는 50% 수준을 나타내고 있다[5].

따라서 기존의 통계적 기법과 함께 기계 학습 예측 알고리즘을 적용하여 미세먼지의 예측 정확도를 개선하기 위해 다양한 시도가 이루어지고 있다[6][7]. 하지만 다양한 외부 요인으로부터 발생하는 미세먼지의 특성과 불균형한 농도별 발생 비율의 문제로 인해 예측 모델의 학습이 잘 이루어지지 않는 문제가 있다. 이러한 문제를 해결하기 위해 K. W. Cho의 연구에서는 미세먼지 농도의 예측 정확도를 향상시키기 위해 특정 농도를 기준으로 분리하여 예측하는 모델을 제안하였다. 미세먼지 농도를 저농도와 고농도를 구분하여 deep neural network 기반의 예측 모델을 통해 성능을 비교하였다. 예측 결과 저농도와 고농도에 대한 예측 성능이 향상되었으며, 특히 고농도 예측 성능의 경우 20.62%의 성능 향상을 확인하였다[8]. K. Kaya의 연구에서는 미세먼지 농도 별 발생 비율의 편차로 인한 회귀 모델의 예측 문제를 해결하기 위해 불균형 문제의 해결 방안을 제안하였다. 불균형 데이터를 제한한 upper sampling, down sampling을 통해 클래스의 샘플 수를 동일하게 만드는 방법으로 전체 데이터 세트에서 80%대의 정확도를 확인하였다[9].

본 논문에서는 미세먼지 농도의 불균형 특성으로 인하여 예측 성능 향상의 문제를 해결하기 위한 미세먼지 농도의 이진 분류 모델을 제안한다. 머신러닝 알고리즘 중 logistic regression, decision tree, SVM(support vector machine), MLP(multi-layer perceptron) 모델을 사용한

다. 미세먼지 농도 $80\mu g/m^3$ 을 기준으로 2개의 클래스로 구분하여 이진 분류 모델로 구성한다. 모델들의 최적의 파라미터를 적용하기 위해 grid search cross validation을 통해 파라미터 탐색을 수행하여 모델들을 설계한다. 모델의 평가는 각 모델들의 오차 행렬 기반 정밀도, 재현율, f1-score를 이용하여 진행한다.

II. 데이터 수집 및 구성

2.1. 데이터 수집 및 전처리

미세먼지의 경우 다양한 환경 변수들에 영향을 받는다. 대기오염 물질과 기상 요소들이 대표적이며, 이러한 환경 변수들은 미세먼지 농도 예측을 위한 연구들에 공통적으로 적용되어 연구되고 있다[10-13]. 본 논문에서 사용한 주요 데이터는 표 1과 같다.

Table. 1 Major data

Type	Data
Meteorological factor	Temperature (The average temperature per hour)
	Humidity (The average humidity per hour)
	Wind Speed (The average wind speed per hour)
	Wind Direction (The average wind direction per hour)
Air pollutant	PM_{10} (The average particulate matter per hour)
	O_3 (The average ozone per hour)
	CO (The average carbon monoxide per hour)
	NO_2 (The average nitrogen dioxide per hour)
	SO_2 (The average sulfur dioxide per hour)

해당 데이터는 천안지역의 측정소에서 2009년부터 2018년까지 10년간 1시간 간격으로 측정된 데이터를 사용하였다. 기상 데이터는 온도, 습도, 풍속, 풍향으로 구성되며, 대기오염 물질 데이터는 PM_{10} , O_3 , CO , NO_2 , SO_2 로 구성하였다. 수집 데이터 중 측정 환경에

따른 결측치가 존재하며 이에 따른 오학습을 방지하기 위해 동일 시간의 모든 데이터를 제거하여 구성하였다.

머신러닝 알고리즘을 이용 이용한 분류를 진행하기 위해 데이터들의 전처리가 필요하다. 지도 학습 기반의 알고리즘을 사용하기 때문에 독립 변수에 해당하는 데이터와 종속 변수에 해당하는 데이터를 따로 구분하여 진행하였다. 독립 변수 데이터의 경우 AQI 기반 미세먼지 농도의 범위를 예측하기 위해 사용되는 데이터로서 온도, 습도, 풍속, 풍향, O_3 , CO , NO_2 , SO_2 으로 구성하였다. 풍향의 경우, 16방위로 표현되는 범주형 데이터에 해당하기 때문에 one-hot encoding을 사용하여 16개의 벡터 형으로 변환하였다. 풍향을 제외한 나머지 독립 변수의 경우, 서로 다른 특성을 갖고 있는 수치형 데이터이며, 데이터에 따라 표현되는 수치의 범위를 통일시켜주기 위해 min max scaling을 통해 0과 1사이의 값으로 변환하였다. 종속 변수에 해당하는 데이터는 PM_{10} 에 해당하며 표 2와 같이 환경부에서 예보로 사용하고 있는 AQI를 기준으로 $80\mu g/m^3$ 이하에 해당하는 ‘ 좋음’, ‘보통’에 해당하는 농도를 하나의 클래스로 구분하였으며, $80\mu g/m^3$ 에 초과하는 ‘나쁨’, ‘매우 나쁨’에 해당하는 농도를 또 다른 하나의 클래스로 구분하였다.

Table. 2 The range of particulate matter concentrations based on AQI

Grade	Class 0		Class 1	
	Good	Moderate	Bad	Very bad
$PM_{10}(\mu g/m^3)$	0 ~ 30	31 ~ 80	81 ~ 150	150 ~

2.2. 데이터 구성

지도 학습 모델에 사용되는 데이터는 크게 학습을 위한 training set과 학습된 모델의 평가를 위한 test set으로 구성된다. 모델의 학습 시 사용되는 training set의 경우 학습이 잘 이루어 졌는지 검증이 필요함에 따라 training set은 train set과 검증을 위한 validation set으로 나누어 구성된다. 본 논문에서는 전처리 과정이 끝난 데이터를 75%의 training set과 25%의 test set으로 구성하였으며, training set 중 80%의 데이터를 train set, 나머지 20%의 데이터를 validation set으로 구성하였다. 그림 1은 모델에 사용된 최종 데이터의 구조를 나타내며, 표 3은 데이터 셋의 구성을 나타낸다.

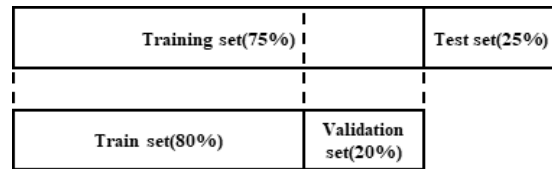


Fig. 1 Structure of data set

Table. 3 Data set configuration

Structure of data set		Samples
Training set	Train set	52,315
	Validation set	13,079
Test set		21,799
Total		87,193

III. 이진 분류 모델 설계

3.1. Logistic regression model

Logistic regression는 선형 회귀와 다르게 종속 변수가 범주형 데이터일 경우 예측 결과가 특정 범주로 분류되기 때문에 regression 기법이 아니라 classification 기법으로 사용할 수 있다.

어떠한 결과를 예측하는 모델의 경우 훈련의 강도에 따라 과대적합 또는 과소적합이 이루어지게 되며, 모델 성능 하락의 주된 원인이 된다. 이러한 문제를 최소화하기 위해 logistic regression는 기본적으로 L2 규제를 적용하기 위한 매개변수 c를 사용한다[14].

따라서 grid search cross validation을 이용하여 최적의 c 값에 대한 탐색을 수행하였다. 탐색할 c 값의 범위는 0.0001, 0.001, 0.01, 0.1, 1, 10, 100, 1000으로 설정하였다. Cross validation의 cv 파라미터는 5로 설정하여 데이터를 분할하였으며, 이에 따라 c값에 순차 접근하여 5회 반복 훈련 및 검증 후 test set을 이용하여 점수를 비교하도록 하였다. 교차 검증 시 validation fold의 전처리를 위해 min max scaler와 모델의 pipeline을 구축하여 탐색하였다. 교차 검증 결과 mean test score가 c값이 10.0일 때 0.808958로 가장 높은 값을 보임에 따라 로지스틱 회귀에 적용할 c의 값은 10.0으로 선정하여 설계하였다.

3.2. Decision tree model

Decision tree는 분류와 회귀에 널리 사용되는 모델이며 기본적으로 특정한 결정에 접근하기 위해 질문을 이어나가며 학습하는 알고리즘이다. 리프 노드가 늘어날수록 훈련 셋의 정확도는 증가하나 과대적합의 문제가 발생할 수 있다[15]. 과대적합의 문제를 막기 위해 사용하는 방법 중 하나는 트리의 깊이를 일정 수준에 도달하면 트리의 성장을 멈추게 하는 방법이다. Decision tree에서 트리의 깊이를 제한하는 파라미터는 max_depth이며, 깊이를 조절함으로써 모델의 성능을 개선시킬 수 있다.

따라서 grid search cross validation을 이용하여 최적의 max_depth 값에 대한 탐색을 수행하였다. 탐색할 max_depth 값의 범위는 1부터 24까지 설정하였으며, k-fold 교차검증의 cv 파라미터를 5로 설정하여 탐색을 진행하였다. 또한 교차 검증 시 validation fold의 전처리를 위해 min max scaler와 모델의 pipeline을 구축하여 탐색하였다. 교차 검증 결과 mean test score가 max_depth값이 4일 때 0.85936013으로 가장 높은 값을 보임에 따라 decision tree에 적용할 max_depth의 값은 4로 선정하여 설계하였다.

3.3. SVM model

SVM은 데이터 분석, 패턴 인식 등 다양한 목적을 위해 사용되는 지도 학습 알고리즘 중 하나이며, 주로 회귀 분석과 분류 분석에 사용된다. SVM은 두 클래스 중 특정 클래스에 속한 데이터들을 기반으로 새로운 데이터가 어느 클래스의 특성을 가지고 있는지 판단하는 비확률적 이진 분류 모델을 만든다. 분류 모델은 두 클래스에 대한 경계를 데이터들이 표현된 공간에 표현되며, 가장 큰 폭의 경계를 도출함으로써 새로운 데이터들을 분류한다[16]. 따라서 SVM은 범주간의 분류를 위한 기준 선을 정의하는 모델이며, 결정 경계라고 표현한다.

SVM은 결정 경계가 어떻게 정의되는지에 따라 성능의 차이가 결정되며, 최적의 결정 경계를 찾는 것이 중요하다. 최적의 결정 경계를 찾기 위해 적용되는 파라미터는 c와 gamma가 있다. C는 결정 경계의 마진을 조절하여 이상치의 허용 범위를 조절하는 파라미터이며, gamma는 결정 경계의 유연성을 조절하여 모델의 과대적합을 방지하는 파라미터이다.

Grid search cross validation을 이용하여 최적의 결정 경계를 구하기 위해 c와 gamma의 최적의 값에 대한 탐

색을 수행하였다. 탐색하기 위한 파라미터 c와 gamma의 범위는 0.0001, 0.001, 0.01, 0.1, 1, 10, 100, 1000으로 설정하였으며, k-fold 교차검증의 cv 파라미터를 5로 설정하여 탐색을 진행하였다. 또한 교차 검증 시 validation fold의 전처리를 위해 min max scaler와 모델의 pipeline을 구축하여 탐색하였다. 교차 검증 결과 c가 1000, gamma가 0.01일 때 mean test score가 0.859238로 가장 높은 값을 보임에 따라 SVM에 적용할 c와 gamma의 값을 1000과 0.01로 선정하여 설계하였다.

3.4. MLP model

MLP는 인공 신경망 구조 중 하나로서 초기 가중치를 임의의 값으로 정의하고 예측 값의 활성 함수 변환 값과 실제 결과 값의 활성 함수 변환 값이 동일하게 나올 때까지 가중치의 값을 계속 수정하는 perceptron을 여러 층으로 구성한 알고리즘이다.

MLP의 구조는 학습 데이터 중 독립 변수를 입력 받기 위한 input layer와 학습을 진행하기 위한 n개의 node를 가지는 hidden layer로 구성하였으며, 활성화 함수는 sigmoid, 최적화 함수는 rmsprop, 손실 함수는 binary crossentropy를 적용하였다. 인공 신경망의 경우 과대적합을 야기하는 요인은 여러 요인이 존재한다. hidden layer를 구성하는 node의 수, 가중치 규제에 대한 L2 규제와 Dropout rate, 학습 방법을 결정하기 위한 batch size가 있으며, 해당 파라미터의 최적 값을 찾기 위해 표 4와 같이 grid를 설정하여 grid search cross validation을 수행하였다.

Table. 4 Parameter grid

Parameter	Value
hidden layer	1, 2, 3, 4
node	20 - 200(interval 20)
L2	0, 0.1, 0.01, 0.001
dropout rate	0, 0.1, 0.2, 0.3, 0.4, 0.5
batch size	20, 40, 60, 80, 100
epoch	100

각 hidden layer 별로 1,200개 조합의 탐색을 진행하였으며, k-fold 교차 검증의 cv 파라미터를 5로 설정하여 수행하였다. 교차 검증 결과 hidden layer는 2, node는 20, L2 규제는 0.001, dropout rate는 0.3, batch size는 80으로 선정하여 설계하였다.

IV. 성능 평가 및 비교

앞 장에서 구성한 데이터 셋과 설계한 4가지의 이진 분류 모델들을 이용하여 분류 성능 평가를 진행하였다. 성능 평가는 오차 행렬을 기반으로 정밀도, 재현율, f1-score를 사용하였다. 그림 2는 훈련된 모델들의 분류 결과를 기반으로 작성한 오차 행렬이다. 표 5는 오차 행렬을 참고하여 계산한 분류 모델들의 성능 비교이다.

Table. 5 Classification performance by models

Class		0 ($80\mu\text{g}/\text{m}^3$ or less)	1 (above $80\mu\text{g}/\text{m}^3$)	f1-score (macro)
Logistic regression	precision	0.9797	0.8167	0.8964
	recall	0.9808	0.8085	
Decision tree	precision	0.9807	0.8176	0.8991
	recall	0.9807	0.8176	
SVM	precision	0.9807	0.8176	0.8991
	recall	0.9807	0.8176	
MLP	precision	0.9845	0.7870	0.8998
	recall	0.9754	0.8551	

Logistic regression의 경우 PM_{10} 의 $80\mu\text{g}/\text{m}^3$ 이하 농도에 대해 정밀도가 0.9797, 재현율은 0.9808로 4가지 모델 중 가장 높은 재현율을 보였으나 $80\mu\text{g}/\text{m}^3$ 초과 농도에 대해 정밀도가 0.8167, 재현율은 0.8085로 4가지 모델 중 가장 낮은 재현율을 보였다. Decision tree와 SVM의 경우 $80\mu\text{g}/\text{m}^3$ 이하 농도에 대해 정밀도가 0.9807, 재현율은 0.9807을 보였으며, $80\mu\text{g}/\text{m}^3$ 초과 농도에 대해 정밀도가 0.8176, 재현율은 0.8176으로 동일한 성능을 보였다. MLP의 경우 $80\mu\text{g}/\text{m}^3$ 이하 농도에 대해 정밀도가 0.9845, 재현율은 0.9754로 4가지 모델 중 가장 높은 정밀도를 보였으나 $80\mu\text{g}/\text{m}^3$ 초과 농도에 대해 정밀도가 0.7870, 재현율은 0.8551로 4가지 모델 중 가장 낮은 정밀도를 보였다. 전체적으로 $80\mu\text{g}/\text{m}^3$ 초과 범위의 농도를 분류함에 있어 낮은 성능을 보였다.

그림 2의 오차 행렬을 통해 분석하였을 경우 분류에 사용된 입력 데이터 중 $80\mu\text{g}/\text{m}^3$ 이하에 해당하는 데이터 비중이 높은 것을 확인할 수 있었으며, 특정 클래스의 비중이 높은 불균형 모델의 정확한 평가를 위해 macro 평균의 f1-score를 사용하였다. f1-score를 기준으

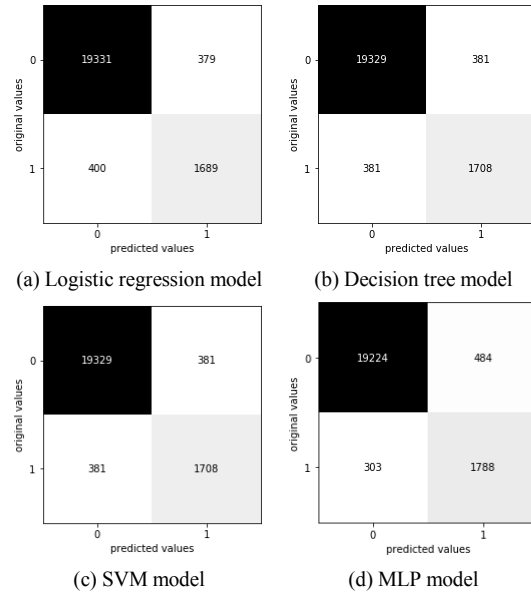


Fig. 2 Confusion matrix

로 logistic regression은 89.64%, decision tree와 SVM은 89.91%, MLP는 89.98%로 MLP가 4가지 모델 중 가장 높은 정확도를 보였다.

V. 결 론

미세먼지 농도를 예측함에 있어 미세먼지가 가지고 있는 불균형적인 특성으로 인해 미세먼지 농도 예측 모델의 학습이 잘 이루어지지 않는 문제가 있다. 이러한 문제를 해결하기 위해 특정 농도를 기준으로 미세먼지를 구분하여 예측을 수행하는 등 다양한 연구가 진행되고 있다. 본 논문에서는 미세먼지 농도의 불균형 특성으로 인하여 예측 성능 향상의 문제를 해결하기 위한 미세먼지 농도의 이진 분류 모델을 제안하였다. 이를 위해 천안 지역에서 10년간 수집한 기상 데이터와 대기오염 물질 데이터를 이용하여 학습 데이터로 구성하였다. 분류를 위한 알고리즘으로 logistic regression, decision tree, SVM, MLP 알고리즘을 사용하였으며, 미세먼지 농도 $80\mu\text{g}/\text{m}^3$ 을 기준으로 2개의 클래스로 구분하여 이진 분류 모델로 구성하였다. 각 모델의 최적화를 위해 grid search cross validation을 통해 파라미터 탐색을 수행하여 모델들을 설계하였다. 평가 결과, $80\mu\text{g}/\text{m}^3$ 이

하의 농도 분류에 있어 약 98%의 성능을 보였으나 $80\mu\text{g}/\text{m}^3$ 초과 농도 분류에 있어 약 80%의 성능을 보였다. f1-score를 통한 분류 모델들의 성능을 비교하였을 경우, MLP 모델이 89.98%로 가장 높은 성능을 보였다.

ACKNOWLEDGEMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(NRF-2019R111A3A01059038) and This paper was supported by the Education and Research Promotion Program of KOREATECH in 2019.

REFERENCES

- [1] M. S. Seo, "The Impact of Particulate Matter on Economic Activity," *The Korean Women Economists Association*, vol. 12, no. 1, pp. 75-100, Jun. 2015.
- [2] A. Valavanidis, K. Fiotakis, and T. Vlachogianni, "Airborne Particulate Matter and Human Health: Toxicological Assessment and Importance of Size and Composition of Particles for Oxidative Damage and Carcinogenic Mechanisms," *Journal of Environmental Science and Health, Part C*, vol. 26, no. 4, pp. 339-362, Nov. 2008.
- [3] K. H. Kim, E. Kabir, and S. Kabir, "A Review on the Human Health Impact of Airborne Particulate Matter," *Environment International*, vol. 74, pp. 136-143, Jan. 2015.
- [4] World Health Organization(WHO), "Health effects of particulate matter. Policy implications for countries in eastern Europe, Caucasus and central Asia," *Regional Office for Europe*, 2013.
- [5] Board of Audit and Inspection(BAI), "Weather Forecast and Earthquake Notification System Operation," *International THE Board of Audit and Inspection of KOREA*, 2017.
- [6] J. W. Cha and J. Y. Kim, "Development of Data Mining Algorithm for Implementation of Fine Dust Numerical Prediction Model," *Journal of the Korea Institute of Information and Communication Engineering*, vol. 22, no. 4, pp. 595-601, Apr. 2018.
- [7] A. Chaloulakou, G. Grivas, and N. Spyrellis, "Neural Network and Multiple Regression Models for PM10 Prediction in Athens: A Comparative Assessment," *Journal of the Air & Waste Management Association*, vol. 53, no. 10, pp. 1183-1190, Oct. 2003.
- [8] K. W. Cho, Y. J. Jung, J. S. Lee, and C. H. Oh, "Separation Prediction Model by Concentration based on Deep Neural Network for Improving PM10 Forecast Accuracy," *Journal of the Korea Institute of Information and Communication Engineering*, vol. 24, no. 1, pp. 8-14, 2020.
- [9] K. Kaya and S. G. Oguducu, "A Binary Classification Model for PM10 Levels," in *2018 3rd International Conference on Computer Science and Engineering (UBMK)*, Sarajevo, pp. 361-366, 2018.
- [10] J. M. Han, J. G. Kim, and K. H. Cho, "Verify a Causal Relationship between Fine Dust and Air Condition-Weather Data in Selected Area by Contamination Factors," *The journal of Bigdata*, vol. 2, no. 1, pp. 17-26, Feb. 2017.
- [11] X. Zhao, R. Zhang, J. L. Wu, and P. C. Chang, "A Deep Recurrent Neural Network for Air Quality Classification," *Journal of Information Hiding and Multimedia Signal Processing*, vol. 9, no. 2, pp. 346-354, Mar. 2018.
- [12] B. T. Ong, S. Komei, and Z. Koji, "Dynamic Pre-training of Deep Recurrent Neural Networks for Predicting Environmental Monitoring Data," in *2014 IEEE International Conference on Big Data (Big Data)*, Washington DC, pp. 760-765, 2014.
- [13] X. Li, L. Peng, X. Yao, S. Cui, Y. Hu, C. You, and T. chi, "Long Short-term Memory Neural Network for Air Pollutant Concentration Predictions: Method Development and Evaluation," *Environmental Pollution*, vol. 231, no. 1, pp. 997-1004, Dec. 2017.
- [14] S. H. Jeon and Y. S. Son, "Prediction of Fine Dust PM10 using a Deep Neural Network Model," *The Korean journal of applied statistics*, vol. 31, no. 2, pp. 265-285, Apr. 2018.
- [15] J. R. Quinlan, "Learning Efficient Classification Procedures and Their Application to Chess End Games," in *Machine Learning*, Berlin, Springer, pp. 463-482, 1983.
- [16] P. H. Huynh, V. H. Nguyen, and T. N. Do, "Enhancing Gene Expression Classification of Support Vector Machines with Generative Adversarial Networks," *Journal of information and communication convergence engineering*, vol. 17, no. 1, pp. 14-20, Mar. 2019.



정용진(Yong-Jin Jung)

2018년 3월 ~ 현재 한국기술교육대학교 전기전자통신공학과 박사과정
2016년 2월 한국기술교육대학교 전기전자통신공학과 공학석사
2014년 2월 공주대학교 전기전자제어공학부 전자공학나노정보공학전공 전자공학트랙 공학사
※관심분야: 미세먼지 예측, 기계 학습, 인공지능망, 심층신경망,



이종성(Jong-Sung Lee)

2017년 3월 ~ 현재 한국기술교육대학교 전기전자통신공학과 박사과정
2014년 8월 한국기술교육대학교 전기전자통신공학과 공학석사
2011년 8월 나사렛대학교 정보통신학부 이학사
※관심분야: 기계 학습, Deep learning, 데이터 분석, 5G network



오창현(Chang-Heon Oh)

1999년 2월 ~ 현재 한국기술교육대학교 전기전자통신공학부 교수
2006년 8월 ~ 2007년 7월 방문교수(University of Wisconsin-Madison)
1993년 10월 ~ 1999년 2월 삼성전자(주) CDMA 개발팀 선임연구원
1990년 2월 ~ 1993년 8월 한진전자(주) 기술연구소 전임연구원
1996년 2월 한국항공대학교 항공전자공학과 공학박사
1990년 2월 한국항공대학교 항공통신정보공학과 공학석사
1988년 2월 한국항공대학교 항공통신공학과 공학사
※관심분야: 무선/이동통신, IoT, 기계학습 기반 통신시스템