

Outlier 데이터 제거를 통한 미세먼지 예보성능의 향상

전영태^{*}, 유숙현^{**}, 권희용^{***}

Improvement of PM Forecasting Performance by Outlier Data Removing

Young Tae Jeon^{*}, Suk Hyun Yu^{**}, Hee Yong Kwon^{***}

ABSTRACT

In this paper, we deal with outlier data problems that occur when constructing a PM_{2.5} fine dust forecasting system using a neural network. In general, when learning a neural network, some of the data are not helpful for learning, but rather disturbing. Those are called outlier data. When they are included in the training data, various problems such as overfitting occur. In building a PM_{2.5} fine dust concentration forecasting system using neural network, we have found several outlier data in the training data. We, therefore, remove them, and then make learning 3 ways. Over_outlier model removes outlier data that target concentration is low, but the model forecast is high. Under_outlier model removes outliers data that target concentration is high, but the model forecast is low. All_outlier model removes both Over_outlier and Under_outlier data. We compare 3 models with a conventional outlier removal model and non-removal model. Our outlier removal model shows better performance than the others.

Key words: PM_{2.5} Forecasting, Deep Neural Network, Outlier, AI

1. 서 론

국내 미세먼지 농도는 2000년부터 지속적으로 감소하다가 2012년 이후 다시 증가추세로 전환되었다. 특히 중국주요 공업도시지역에서 발생한 고농도 초미세먼지에 의한 스모그가 편서풍을 타고 국내로 유입되는 상황에서 국내 대기정체로 인해 배출된 다양한 오염원이 더해지면서 고농도 미세먼지 현상이 발생하고, 이렇게 발생한 고농도 미세먼지는 일반적으로 3~4일간 지속되는 경향을 보이고 있다[1-4].

이렇게 빈번히 발생하며 장시간 정체되는 고농도 미세먼지는 국민들의 생활과 건강에 직접적인 피해를 입히고 있어서 미세먼지 예보의 정확성에 대한 국민적 요구가 지속적으로 높아지고 있다.

이미 2014년부터 환경부를 중심으로 미세먼지 예보 시스템을 운영 중에 있으며 예보관이 다양한 수치예보 결과 및 관측 현황을 고려하여 최종 확정하는 개념예보의 지수적중률은 약 87%로 양호하나 국민 건강에 직접적인 영향을 주는 나쁨 이상의 고농도 감지확률은 67%로 낮아서 실제 시민이 느끼는 체감

※ Corresponding Author: Hee Yong Kwon, Address: (430-714) 22 Samdeokro 37beon-gil, Manan-gu, Anyang-si, Gyeonggi-do, Korea, TEL: +82-31-467-0878, FAX: +82-31-463-1249, E-mail: hykwon@anyang.ac.kr
Receipt date: Mar. 16, 2020, Revision date: Apr. 29, 2020
Approval date: May. 13, 2020

^{*} Dept. of Computer Eng., Anyang University
(E-mail: qkow110@ayum.anyang.ac.kr)

^{**} Dept. of Information & Communication Eng., Anyang University
(E-mail: shy@anyang.ac.kr)

^{***} Dept. of Computer Eng., Anyang University

오염도와 많은 차이를 보이고 있다[4]. 이를 해결하기 위해서는 고농도 적중률을 향상시키기 위한 연구가 시급하다.

관련 연구로는 대기질 예보에 이미 효과가 입증된 인공신경망을 이용한 연구가 대표적이며, 인공신경망에서 가장 범용적으로 사용되는 MLP(Multi Layer Perceptron)에 날씨, 대기질, 지형, 기상 측정 및 예보 데이터를 입력하여 PM10, PM2.5, O3을 비롯한 대기질을 예보한 연구[5-10], 시계열 자료의 처리에 효과적인 RNN(Recurrent Neural Network) 및 LSTM(Long-Short Term Memory)를 이용하여 대기질을 예측한 연구[11-12], 심층신경망을 이용하여 서울권역의 PM10 및 PM2.5를 예보한 연구[13-14] 등이 있다.

본 연구에서는 미세먼지 예보성능을 향상시키기 위해서 예보 성능을 저하시키는 outlier 데이터를 선별하고, outlier가 예보 성능에 미치는 영향을 분석했다. Outlier의 통계학적 의미는 통계적 자료 분석의 결과를 왜곡시키거나, 자료 분석의 적절성을 위협하는 변수 값 또는 사례를 말한다[15]. 미세먼지 데이터에 있어서 outlier는 저농도인데 고농도로 예보되거나 그 반대인 데이터에 해당되며, 이러한 outlier가 학습 데이터에 포함되면 당연히 예보 성능이 저하된다.

Outlier가 예보 성능에 미치는 영향을 분석하기 위해서 먼저 주어진 데이터에서 outlier를 검출했다. 검출에는 여러 가지 통계학적 기법 중 신뢰구간 방법과 1.5 IQR rule[16]을 적용했고, 통계학적 기법들의 문제를 해결하기 위해서 미세먼지 농도구간을 이용한 outlier 검출방법을 제안하고 적용했다. 예보 모델은 복잡하고, 비선형적이며, 방대한 자료에 대해서 효율적이라고 평가되는 심층 신경망을 사용했고, 입력 데이터는 기상과 대기질 측정 및 예보 데이터와 날씨 데이터를 사용했다.

이어지는 2장에서는 outlier 검출방법들과 검출결과를 설명하고, 3장에서는 검출방법에 따른 예보실험과 결과를 기술하고, 4장에서는 결론 및 향후과제를 제시한다.

2. Outlier

미세먼지 데이터에서 outlier란 정상적인 데이터들과 비교했을 때, 자신이 속한 군집에서 비정상적으로 떨어져 있는 데이터를 의미한다. Outlier를 판단하는 방법에는 주로 신뢰구간에 의한 방법과 1.5 IQR

Rule 등의 통계적 방법을 이용하지만, 미세먼지 데이터에 이와 같은 통계적 기법을 적용 시키면 여러 가지 문제점들이 발생한다. 본 논문에서는 이를 해결하기 위하여 미세먼지 예보 농도 구간을 이용한 outlier 검출 방법을 제안하였다.

2.1 통계적 방법에 의한 outlier 검출

2.1.1 신뢰구간에 의한 검출

신뢰구간 방법은 통계학에서 추정 방법 중 구간추정으로 ‘모수가 어느 범위 안에 있는지를 확률적으로 보여주는 방법’이다[17]. 신뢰구간은 구하고자 하는 모분산을 아느냐 모르냐에 따라서 사용하는 분포도가 달라지는데 이를 Fig. 1에 나타냈다.

본 논문에서 사용된 미세먼지 데이터의 경우 모분산을 모르고 데이터의 개수가 30개 이상인 데이터들의 집합이기 때문에 정규 분포를 사용하여 outlier를 검출하였다. 정규분포를 이용한 outlier는 Fig. 2의 α 부분에 해당하는 부분을 검출하는 것을 의미한다. 먼저 α 를 구하기 위해서는 $1-\alpha$ 에 해당하는 범위를 구해야 한다. 이를 위해 표본 집단의 표본평균 \bar{X} , 표본표준편차 s , 데이터의 크기 n 그리고 신뢰수준에 따른 $Z_{\frac{\alpha}{2}}$ 을 계산한 뒤, 식 (1)을 이용하여 모수의 범위를 추정한다. 신뢰수준은 보통 90%, 95%, 99%를 사용하며 표준정규분포표를 이용하여 구한다. 이때,

- | |
|--|
| <ol style="list-style-type: none"> 1) If the population variance is known,
-> use normal distribution. 2) If the population variance is not known,
-> use T distribution ($n < 30$),
-> use normal distribution ($n \geq 30$)
(n is number of data) |
|--|

Fig. 1. How to determine the distribution.

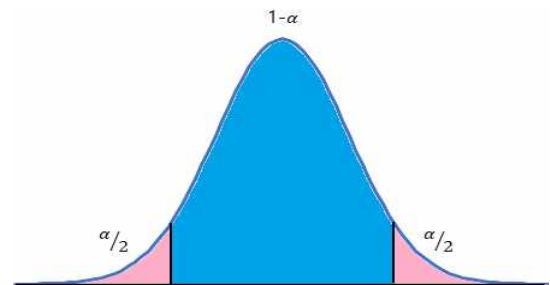


Fig. 2. Normal distribution.

Table 1. Outlier detection result using 95% confidence interval

Time	Total Data	Outliers
T06	999	929
T12		914
T15		913

범위를 벗어나는 데이터들을 outlier로 검출한다.

$$\bar{X} - Z_{\alpha/2} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_{\alpha/2} \frac{s}{\sqrt{n}} \quad (1)$$

Table 1은 95% 신뢰구간 방법을 이용하여 미세먼지 데이터에 대한 outlier를 검출한 결과이다. Table 1의 첫 번째 열에 표기한 T06, T08, T12는 예보시간으로 각각 당일(D+0), 내일(D+1), 모레(D+2)의 첫 구간을 의미하며 자세한 내용은 Table 5에 기술했다. T06시간대의 데이터의 경우 보통구간에 해당하는 데이터들이 총 999개 중 548개로 데이터가 편향된 모습을 보이고 있다. 그로 인해 데이터의 분포가 정규분포를 따르지 않는 것이 확인 되었고, 대부분의 데이터가 Outliers로 검출 되고 있다. 해당 데이터들을 제거하게 되면 학습을 진행하기에는 절대적으로 부족한 데이터들만 남게 되는 문제점이 발견되어 실제 학습에는 미적용 하였다.

2.1.2 1.5 IQR rule

IQR은 사분위수 범위를 의미하고, 데이터의 범위를 표현하기 위한 방법 중 하나로 그 개념을 Fig. 3과 Fig.4[18]에 기술했다. 이는 일반적인 방법인 최대값 - 최소값을 이용한 범위 계산의 경우 이상값 존재

시에 생기는 문제점을 보완한 척도 이다. 이를 구하기 위해서는 먼저 사분위수를 구해야 하는데 사분위수란 자료를 오름차순으로 배열하고, 이를 4등분 했을 때 각각의 4등분 되는 위치의 값을 의미하고 1사분위수(Q1), 2사분위수(Q2), 3사분위수(Q3)으로 구분한다. 이때, 2사분위수의 경우 데이터의 중앙값과 일치한다.

사분위수는 먼저 Q2 즉, 데이터 전체의 중앙값을 구하는 데 이때, 데이터의 개수가 홀수인지 짝수인지에 따라 구해지는 중앙값이 다르다. Fig. 3을 보게 되면 짝수의 경우 중앙이 명확하지 않기 때문에 중앙에서 가장 가까운 두 수의 절반 값을 중앙값으로 한다. 반대로 홀수의 경우 중앙이 명확하기 때문에 중앙에 있는 데이터를 중앙값으로 한다. Q2를 구한 이후 Q1은 아래쪽 절반의 데이터에 대하여 Q2와 동일한 방법으로, Q3는 위쪽 절반에 대하여 Q2와 동일한 방법으로 구한다.

사분위수 범위인 IQR은 식 (2)를 이용하여 구한다.

$$IQR = Q3 - Q1 \quad (2)$$

$$Q1 - 1.5 \times IQR \leq x \leq Q3 + 1.5 \times IQR \quad (3)$$

IQR을 이용하여 outlier를 검출하는 방법은 1.5 IQR rule인 식 (3)을 이용하여 신뢰구간 방법과 같이 범위를 구한 다음 해당 범위에 벗어나는 데이터들에 대해서 outlier로 검출하는 것이다. Table 2에 1.5 IQR rule를 이용하여 outlier 데이터를 검출한 결과를 기술했다.

1.5 IQR rule를 이용하여 Outlier를 검출한 결과 신뢰구간 방법을 이용하여 검출했을 때에 비해 상당

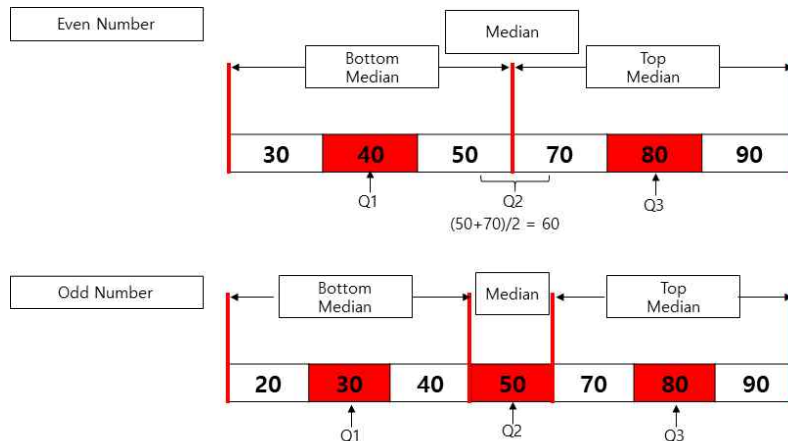


Fig. 3. How to find Quartile.

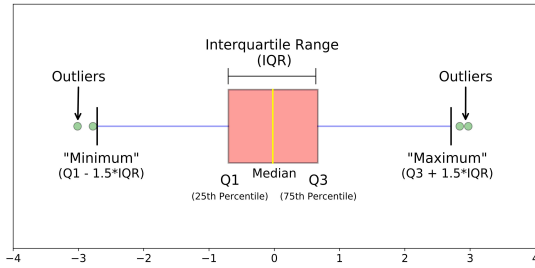


Fig. 4. Different parts of a boxplot.

Table 2. Outlier detection result using 1.5 IQR rule

Time	Total Data	Outliers
T06	999	34
T12		37
T15		36

$$-9 \leq x \leq 55$$

Fig. 5. Result of 1.5 IQR rule (T06).

히 적은 양이 검출되었다. Fig. 5에는 T06시간대에 대하여 1.5 IQR rule를 적용해 구해진 결과를 제시했다. 결과를 보았을 때, 고농도에 해당하는 55를 초과하는 모든 데이터들이 outlier로 검출되어진 모습을 확인 할 수 있다. 이로 인하여 학습을 진행 할 때 고농도에 대한 학습이 더디게 혹은 안 되는 모습을 보이는 문제점을 발견 하였다.

2.2 제안한 outlier 검출 방법

2.1절에서 기술한 바와 같이 신뢰구간 방법의 경우 너무 많은 데이터가 제거되고, 1.5 IQR rule의 경우는 고농도 데이터가 상당수 제거되는 문제점이 발생한다. 이러한 문제로 인하여 미세먼지 데이터에 통계적 방법을 적용하기가 어렵다. 본 논문에서는 이러한 문제점들을 해결하기 위하여 미세먼지의 농도 구간을 이용한 outlier 검출 방법을 제안하였다. 미세먼지 예보지수는 Table 3에 제시한 것과 같이 ‘좋음’,

Table 3. PM2.5 Forecasting concentration section

Low concentration		High concentration	
Good	Normal	Bad	Very bad
0~15	16~35	36~75	76~

1. Over outlier

```
if(target < 36 and output >= 36 ) then
    set this-data as outlier;
```

2. Under outlier

```
if(target >= 36 and output < 36 ) then
    set this-data as outlier;
```

3. All outlier

```
if ( (target < 36 and output >= 36 ) or
    (target >= 36 and output < 36 ) ) then
    set this-data as outlier;
```

Fig. 6. Methods for Outlier data detection.

‘보통’, ‘나쁨’, ‘매우 나쁨’으로 4개의 농도구간(지수)을 가지는데 ‘좋음’과 ‘보통’ 농도를 저농도로, ‘나쁨’과 ‘매우 나쁨’ 농도를 고농도로 구분하여 실험을 진행 하였다.

본 논문에서는 Fig. 6에 기술한 바와 같이 3가지 모델을 생성하여 실험을 진행하였다. Outlier란 소속 군집에서 벗어난 데이터를 의미하기 때문에 학습이 종료 되었을 때 자기 자신의 농도 구간을 벗어난 데이터들을 outlier로 검출하는 방법을 사용하였다. 이때, 구간은 Table 3에 제시한 저농도 구간과 고농도 구간을 기준으로 했다. 해당 방법을 사용하게 되면 어느 정도 학습이 진행되었기 때문에 신뢰구간 방법의 문제점인 대부분의 데이터가 지워지게 되는 문제점을 해결 할 수 있으며, 1.5 IQR rule의 문제점인 범위를 벗어나는 모든 고농도 데이터를 제거 하지 않고 학습이 제대로 진행되지 않은 데이터들만 제거 되기 때문에 고농도 데이터 부족으로 인한 학습부족 현상을 해결할 수 있다. Fig. 6은 outlier 데이터를 검출하는 알고리즘으로, target이 저농도일 때, output이 고농도로 예측하는 Over_outlier, target이 고농도일 때, output이 저농도로 예측하는 Under_outlier, 두 경우를 모두 합한 All_outlier모델이다.

3. 실험 및 결과

3.1 입력 인자와 신경망 예보모델

Outlier가 예보 성능에 미치는 영향을 분석하기 위해서 Fig. 7[19]에 제시한 심층신경망을 기반으로 한 예보 모델을 사용했고, 입력인자로 Table 4에 기술한 기상, 대기질 측정 및 예보 데이터와 날짜(Julian

Table 4. Input Data

Weather	Temperature, Pressure, Relative humidity, Mixed highs, U, V, Geopotential, height_850hpa, 850hpa_U, 850hpa_V, Relative humidity_850hpa, Temperature_850hpa, Geopotential height_925hpa, 925hpa_U, 925hpa_V, Temperature difference_850-925, Dew point temperature, Accumulated Precipitation, Solar radiation
Air quality	CASE4_PM2.5, PM10, O3, NO2, CO, SO2, PM2.5
Date	Julian membership function (January ~ December)

day)[19]을 사용하였다.

본 논문에서 사용한 미세먼지 예보 모델은 매 시간 측정 하는 기상과 대기질 데이터들의 6시간 평균 자료들을 입력으로 하여, 15시에 당일(D+0), 내일(D+1), 모레(D+2)에 대해서 6시간 단위의 예보 값을 생성한다. 예보 구간에 대한 상세 내용은 Table 5에 기술했다. 예보모델은 15시에 예보하기 때문에 T6부터가 예보구간에 해당하고 따라서, T1~T5는 과거, T6~T7이 당일(D+0), T8~T11이 내일(D+1), T12~T15가 모레(D+2) 예보구간이다.

예보구간이 T6~T15로 10개이므로 예보모델은 각 예보구간별로 1개씩 총 10개의 네트워크로 구성된다. 각 구간별 예보모델에는 과거구간인 T1~T5의 측정 자료와 해당 예보구간의 예보자료가 입력된다.

Table 5. Time frame of 15 hour forecast model

Time	Time corresponding to section
T1	D-1 07~12 o'clock, Measure at forecast time
T2	D-1 13~18 o'clock, Measure at forecast time
T3	D-1 19~24 o'clock, Measure at forecast time
T4	D+0 01~06 o'clock, Measure at forecast time
T5	D+0 07~12 o'clock, Measure at forecast time
T6	D+0 13~18 o'clock
T7	D+0 19~24 o'clock
T8	D+1 01~06 o'clock
T9	D+1 07~12 o'clock
T10	D+1 13~18 o'clock
T11	D+1 19~24 o'clock
T12	D+2 01~06 o'clock
T13	D+2 07~12 o'clock
T14	D+2 13~18 o'clock
T15	D+2 19~24 o'clock

Hourly measured data are averaged at every 6 hour interval.

3.2 실험 방법 및 결과

본 논문에서는 인공신경망 학습 진행 시, 먼저 1차 학습을 통해 outlier들을 검출하고, 입력 데이터에서 제거한 뒤, 제거한 입력데이터를 사용하여 2차 학습을 진행하게 된다.

실험에 사용된 데이터는 2015년 01월 01일부터 2018년 12월 31일까지의 서울 지역의 PM2.5를 비롯한 대기질 및 기상 측정값 및 예보 값을 사용하였다. 이 중 학습에 사용된 데이터는 2015년 01월 01일부터 2017년 12월 31일까지 1096일 중 결측 등의 원인으로 데이터가 존재하지 않는 일수를 제외한 999일의 데이터를 사용 하였고, 평가에는 2018년 01월 01일부터 2018년 12월 31일까지 365일 중 데이터가 존재하지 않는 일수를 제외한 344일 데이터를 이용하였다.

Table 6은 1차 학습이 종료되었을 때, 각 모델별로 검출되는 outlier의 개수이다. 총 999개의 데이터 중 적게는 66개에서 많게는 227개의 데이터들이 제거되는 모습을 볼 수 있다.

Table 7은 본 논문에서 제안한 농도구간을 이용한 outlier 검출방법(All, Over, Under)과 통계적 기법인

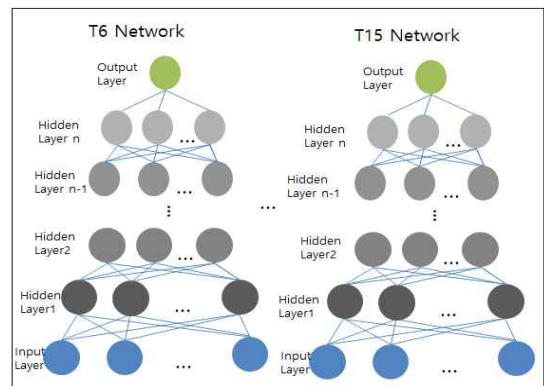


Fig. 7. DNN architecture [3].

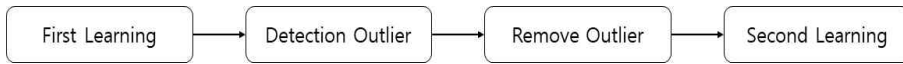


Fig. 8. Outlier model learning algorithm.

Table 6. Number of Removed Outlier Data

Model	T06	T08	T12
Over_outlier	66	98	122
Under_outlier	76	103	105
All_outlier	142	201	227

1.5IQR rule에 의한 outlier 검출방법(1.5IQR)으로 outlier를 제거한 예보 결과와 outlier를 제거하지 않은 기존 모델인 Julian과 수치모델 CASE04의 예보 결과를 기술한 것이다. 예보모델의 성능 평가는 Fig. 9에 제시한 통합대기환경지수(CAI: Comprehensive air-quality index)를 따랐다.

Accuracy는 지수 적중률로 실제 측정값이 좋음(나쁨)일 때, 예보가 좋음(나쁨)으로 정확히 맞춘 비율을 의미하고, POD는 감지확률로, 측정값이 고농도일 때, 예보 역시 고농도로 판단한 비율을 의미한다. FAR은 오경보율로서 실제 측정값은 저 농도이지만 고농도로 예보한 비율이다. (실제 측정값이 고농도이지만 저 농도로 예보한 경우 전자에 비해 중요도가 높지 않고, 실제 현업에서 성능지표로 활용 하지 않아 제외함)

먼저 예보결과를 수치모델(CASE04)과 상호 비교

해보면, 5개의 예보모델(All, Over, Under, 1.5IQR, Julian) 모두 Accuracy는 우수하고, POD는 좋지 않은 성능을 보임을 알 수 있다. 하지만, 수치모델(CASE04)의 경우 80%대의 높은 POD 성능에도 불구하고, FAR이 48%~54%로 높게 형성되어 수치모델의 고농도 감지 성능은 좋다고 볼 수 없다. 반면에 나머지 5개의 예보모델들의 FAR은 17%~26%이다.

둘째로 outlier를 제거한 모델들을 서로 상호 비교해보면 1.5 IQR 모델의 경우 FAR은 제안한 outlier 제거 모델들보다 우수하고, Accuracy와 D+0 구간의 POD는 비슷하나 D+1 이상의 구간에서 POD의 성능이 현저히 떨어짐을 알 수 있다.

다음으로 제안한 outlier 제거 모델들과 기존의 예보모델인 Julian를 비교해보면 Accuracy는 -4% 이내로 비슷한 성능을 보였고, FAR 역시 20%대로 비슷하였으나, POD는 D+2 구간에서 All과 Over 모델이 7%~9%로 크게 향상되었다. 예보에 있어서 일반적으로 현 시점부터 예보시간이 멀어질수록 성능이 떨어진다. 하지만, 제안한 All과 Over outlier 제거 모델은 오히려 예보성능이 떨어지는 D+2 구간에서 성능이 향상되어 장기 예보 모델로서의 활용도가 기대된다.

Level		Prediction value			
		Good	Normal	Bad	Very bad
Observation value	Good	a1	b1	c1	d1
	Normal	a2	b2	c2	d2
	Bad	a3	b3	c3	d3
	Very bad	a4	b4	c4	d4

I: , II: , III: , IV: 

Evaluation item	Equation
Accuracy	$= \frac{(a1 + b2 + c3 + d4)}{N} \times 100(\%)$
POD: Probability of Detection	$= \frac{IV}{(III + IV)} \times 100(\%)$
FAR: False Alarm Rate	$= \frac{(II)}{(II + IV)} \times 100(\%)$

Fig. 9 Items and methods for evaluating the performance of the forecast model.

Table 7. PM2.5 Predict Result

Kind	Point	Model	Day	Accuracy	POD	FAR
Outlier detection	Proposed method	All	D+0	75	67	26
			D+1	74	63	26
			D+2	77	64	22
		Over	D+0	76	67	26
			D+1	75	64	25
			D+2	75	66	23
		Under	D+0	77	67	19
			D+1	75	54	23
			D+2	75	52	22
	Statistical method	1.5IQR	D+0	77	64	18
			D+1	77	43	20
			D+2	74	34	0
Outlier not detection	Original model	Julian	D+0	77	70	24
			D+1	79	63	17
			D+2	76	57	20
	Atmospheric prediction	CASE04	D+0	57	86	48
			D+1	57	89	56
			D+2	58	86	54

4. 결 론

본 논문에서는 PM2.5 미세먼지의 예보성능을 향상시키기 위하여, 예보모델에 적합한 3 가지 outlier를 제시하고, 각각을 제거한 예보모델의 성능을 비교하여 최적의 방안을 제시하였다. Outlier를 제거하지 않은 기존의 Julian모델의 경우 Accuracy와 FAR은 좋은 성능을 보였으나, D+1 이상의 구간에서 POD가 좋지 못한 결과를 보였다. 하지만, 제안한 outlier 제거 모델은 예보성능이 떨어지는 D+1 이상의 구간에서 최대 9%의 POD 성능이 향상되어 장기 예보의 고농도 감지에 outlier 제거가 효과적임을 알 수 있었다. 하지만, outlier 제거 모델의 경우 D+0의 POD와 전반적인 FAR이 각각 약간씩 감소/증가하는 것으로 나타났다. 이와 같은 현상은 제거된 outlier 저농도 데이터에 유효한 데이터가 포함되어 발생한 현상으로 추측된다. 향후 이를 해결하기 위하여 제안한 outlier 검출 방법에 의해 추출된 데이터의 유효성을 검증하는 방법에 대한 연구를 진행할 예정이다.

REFERENCE

- [1] S. Lee, C. Ho, and Y. Choi, "High-PM10 Concentration Episodes in Seoul, Korea: Background Sources and Related Meteorological Conditions," *Atmospheric Environment*, Vol. 45, Issue 39, pp. 7240-7247, 2011.
- [2] S. Lee, C. Ho, Y. Lee, H. Choi, and C. Song, "Influence of Transboundary Air Pollutants for China on the High PM10 Episode in Seoul, Korea for the Period October 16-20, 2008," *Atmospheric Environment*, Vol. 77, pp. 430-439, 2013.
- [3] H. Oh, C. Ho, J. Kim, D. Chen, S. Lee, Y. Choi, et al., "Long-Range Transport of Air Pollutants Originating in China: A possible Major Cause of Multi-Day High-PM10 Episodes During Cold Season in Seoul, Korea," *Atmospheric Environment*, Vol. 109, pp. 23-30, 2015.
- [4] NIER, *A Study of Construction of Air Quality*

- Forecasting System Using Artificial Intelligence(I)*, NIER-SP2017-148, 11-1480523-000 3221-01, 2017.
- [5] A.B. Chelani, D.G. Gajghate, and M.Z. Hasan, "Prediction of Ambient PM₁₀ and Toxic Metals Using Artificial Neural Networks," *Journal of the Air and Waste Management Association*, Vol. 52, Issue 7, pp. 805-810, 2002.
- [6] I.G. McKendry, "Evaluation of Artificial Neural Networks for Fine Particulate Pollution (PM₁₀ and PM_{2.5}) Forecasting," *Journal of the Air and Waste Management Association*, Vol. 52, Issue 9, pp. 1096-1101, 2002.
- [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 and Waste Management Association*, Vol. 53, Issue 10, pp. 1183-1190, 2003.
- [8] G. Corani, "Air Quality Prediction in Milan: Feed-Forward Neural Networks, Pruned Neural Networks and Lazy Learning," *Ecological Modelling*, Vol. 185, Issue 2-4, pp. 513-529, 2005.
- [9] M. Cai, Y. Yin, and M. Xie, "Prediction of Hourly Air Pollutant Concentrations Near Urban Arterials Using Artificial Neural Network Approach," *Transportation Research Part D: Transport and Environment*, Vol. 14, Issue 1, pp. 32-41, 2009.
- [10] S. Thomas and R.B. Jacko, "Model for Forecasting Expressway Fine Particulate Matter and Carbon Monoxide Concentration: Application of Regression and Neural Network Models," *Journal of the Air and Waste Management Association*, Vol. 57, Issue 4, pp. 480-488, 2012.
- [11] J. Fan, Q. Li, J. Hou, X. Feng, H. Karimian, and S. Lin, "A Spatiotemporal Prediction Framework for Air Pollution Based on Deep RNN," *Proceeding of International Society for Photogrammetry and Remote Sensing Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume IV-4/W2, International Symposium on Spatiotemporal Computing*, pp. 15-22, 2017.
- [12] B.S. Freeman, G. Taylor, B. Gharabaghi, and J. The, "Forecasting Air Quality Time Series Using Deep Learning," *Journal of the Air and Waste Management Association*, Vol. 68, Issue 8, pp. 866-886, 2018.
- [13] S. Yu and Y. Jeon, "Improvement of PM₁₀ Forecasting Performance Using DNN and Secondary Data," *Journal of Korea Multimedia Society*, Vol. 22, No. 10, pp. 1187-1198, 2019.
- [14] S. Yu, "Development of PM₁₀ Forecasting Model for Seoul Based on DNN Using East Asian Wide Area Data," *Journal of Korea Multimedia Society*, Vol. 22, No. 11, pp. 1300-1312, 2019.
- [15] Outlier(2000), <https://terms.naver.com/entry.nhn?docId=1943645&cid=41989&categoryId=41989> (accessed March 12, 2020).
- [16] J.W. Tukey, *Exploratory Data Analysis*, Addison-Wesley, 1977.
- [17] Confidence Interval(2018), https://ko.wikipedia.org/wiki/%EC%8B%A0%EB%A2%B0_%EA%B5%AC%EA%B0%84 (accessed April 10, 2020).
- [18] Understanding Boxplots(2018), <https://towardsdatascience.com/understanding-boxplots-5e2df7bcbd51> (accessed March 12, 2020).
- [19] S. Yu, Y. Jeon, and H. Kwon, "Improvement of PM₁₀ Forecasting Performance Using Membership Function and DNN," *Journal of Korea Multimedia Society*, Vol. 22, No. 9, pp. 1069-1079, 2019.



전 영 태

2019년 안양대학교 컴퓨터공학과
학사
2019년~현재 안양대학교 컴퓨터
공학과 석사 과정
관심분야: 패턴인식, 인공지능,
영상처리, 병렬처리응용



권 희 용

1983년 서울대학교 전자계산기공
학과 학사
1985년 서울대학교 전자계산기공
학과 석사
1993년 서울대학교 컴퓨터공학과
박사



유 숙 현

1999년 안양대학교 컴퓨터공학과
학사
2002년 안양대학교 컴퓨터공학과
석사
2011년 안양대학교 컴퓨터공학과
박사

1986년~1995년 한국통신 연구개발단 선임연구원
1995년~현재 안양대학교 컴퓨터공학과 교수
관심분야: 패턴인식, 인공지능, 영상처리, 병렬처리응용

2012년~현재 안양대학교 정보통신공학과 조교수
관심분야: 패턴인식, 영상처리, 병렬처리응용, 딥러닝