

다중 선형 회귀에 의한 광산란 초미세먼지 측정기의 황사 보정 기법

백승훈
중원대학교 컴퓨터공학과 교수

An Asian Dust Compensation Scheme of Light-Scattering Fine Particulate Matter Monitors by Multiple Linear Regression

Sung Hoon Baek
Professor, Department of Computer Engineering, Jungwon University

요약 광산란법을 이용한 초미세먼지 측정기는 초 단위의 측정이 가능하고 휴대할 수 있는 크기로 설계될 수 있다. 또한 하나의 센서로 여러 크기별 ($PM_{1.0}$, $PM_{2.5}$, $PM_{4.0}$, 및 PM_{10}) 농도를 측정할 수 있다. 이 방식은 입자의 개수와 크기를 측정하고 이를 단위 부피당 무게인 농도로 변환하는 과정 때문에 큰 밀도를 가지는 황사에 대해서는 큰 오차를 나타낸다. 본 논문은 광산란 초미세먼지 측정기가 여러 크기별 PM 농도를 이용하여 황사 발생 시 초미세먼지($PM_{2.5}$)의 농도의 오차를 정확히 보정할 수 있고, 황사가 발생하지 않을 때도 영향을 받지 않는, 다중 선형 회귀 기법의 기계학습에 의한 보정 기법을 제시한다. 두 가지 또는 세 가지의 PM 크기 입력만으로도 광산란 미세먼지 측정 장치의 황사 오류를 크게 보정할 수 있음을 보인다. 한 달 동안 중부권대기환경연구소의 베타레이 측정기와 광산란 측정기의 측정값을 비교 분석하였다. 황사가 없는 구간에서 이 두 장비의 상관계수(R^2)는 0.927이었고, 황사를 포함한 전 구간에서 상관계수는 0.763이었지만, 기계학습을 통하여 상관계수가 0.944로 향상되었다.

주제어 : 기계학습, 초미세먼지, 황사, 센서, 융합

Abstract Light-scattering fine particulate matter monitors can measure particulate matter (PM) concentrations in every second and can be designed in a portable size. They can measure the concentrations of various PM sizes ($PM_{1.0}$, $PM_{2.5}$, $PM_{4.0}$ and PM_{10}) with a single sensor. They measure the number and size of particulate matters and convert them to weight per volume (concentration). These devices show a large error for asian dust. This paper proposes a scheme that compensates the $PM_{2.5}$ concentration error for asian dust by multiple linear regression machine learning in light-scattering PM monitors. This scheme can be effective with only two or three types of PM sizes. The experimental results compare a beta-ray PM monitor of national institute of environmental research and a light-scattering PM monitor during a month. The correlation coefficient (R^2) of these two devices was 0.927 without asian dust, but it was 0.763 due to asian dust during the entire experimental period and improved to 0.944 by the proposed machine learning.

Key Words : Machine learning, Fine particulate matter, Yellow dust, Sensor, Convergence

*This work was conducted while a Sabbatical at Jungwon University in 2020.

*Corresponding Author : Sung Hoon Baek(shbaek@jwu.ac.kr)

Received June 29, 2021
Accepted August 20, 2021

Revised July 31, 2021
Published August 28, 2021

1. 서론

초미세먼지는 건강을 악화시키는 사회적 문제이다. 초미세먼지가 호흡기, 심장혈관 사망률, 인체의 여러 질병에 영향을 미친다[1-3]. 또한 미국의 300개 이상의 카운티에서 수집된 결과에서는 초미세먼지 $20\mu\text{g}/\text{m}^3$ 당 중증 COVID-19에 감염될 가능성이 100% 증가한다는 연구결과가 있다[4].

건강과 관련된 초미세먼지의 정확한 실시간 측정은 매우 중요하다. 그러므로 초미세먼지 측정 장치에 관한 연구가 많이 수행되고 있다.

미세먼지 측정 방법으로서 중량농도법, 베타선법, 및 광산란법이 있다. 국가 측정망에서는 베타선법이 많이 사용되고 있다. 베타선법은 중량농도법보다 빠른 1시간 단위로 공기질을 측정할 수 있다. 광산란법은 베타선법에 비해서 크기가 작고 초 단위로 연속적인 측정이 가능하다. 또한, 미국 EPA(environmental protection agency)에서는 중량농도법과 높은 상관관계를 갖는 광산란법 장치(Class III)에 관한 인증제도를 운영하고 있으며, 점차 이 인증에 통과한 장치들이 늘어나고 있다[5,6].

한 시간 지연이 있는 베타선법에 비해서 초 단위로 실시간 측정이 가능한 광산란법 장치에 관해 많은 연구가 진행되고 있다. 이에 따른 연구의 한 부분으로서 광산란법 장치가 황사에 대해 과소평가하는 문제를 해결하는 연구가 필요하다[7].

광산란법 측정 장치는 먼지의 개수와 크기를 측정하는 방식인데 황사는 일반적인 미세먼지(황산염, 질산염)와 크게 다른 밀도를 가지고 있다. 미세먼지 측정 장치의 농도 단위는 단위 부피당 질량인데, 황사의 높은 밀도 때문에 광산란법 측정 장치의 계측값은 베타선법 또는 중량법으로 측정하는 것과는 크게 다른 오차를 발생시킬 수 있다.

본 논문은 황사로 인해 초미세먼지 농도가 저평가되는 광산란법 측정 장치 내에서 다중 선형 회귀 기법에 따른 기계학습(machine learning) 방법으로 해결하고자 한다. 아직 기존 연구에서는 이러한 문제를 해결하는 연구결과는 없었다. 하지만 기계학습과 미세먼지에 관련된 연구는 진행되고 있다.

기계학습으로 직접 측정 없이 다른 기상정보로부터 미세먼지 농도를 예측하는 많은 연구가 이루어졌다. 이런 연구들은 풍향 풍속 상대습도와 같은 기상인자, 하

학인자, 및 위상기반만을 적용하여 다중선형회귀로 시간대별 및 일별 PM_{10} ($10\mu\text{m}$ 이하의 먼지입자) 또는 $\text{PM}_{2.5}$ ($2.5\mu\text{m}$ 이하의 먼지입자) 농도를 예측하는 시도가 활발히 이루어지고 있다[8-15].

다중선형회귀 모델을 이용하여 7가지 기상자료(기온, 풍향, 풍속, 강수량, 습도, 현지기압, 해면기)로부터 서울의 PM_{10} 농도를 예측하는 모델을 제시하는 연구가 있었으며[8]. Harishkumar 등은 인공위성에서 얻은 에어로졸의 광학적 두께로부터 기계학습 회귀 모델을 통하여 $\text{PM}_{2.5}$ 를 예측하는 방법을 제시하였다 [9].

황사와 관련한 연구로서, 황사가 발생하는 동안 공기전파 세균 위험을 기계학습으로 예측하는 연구가 있다 [14].

본 논문에서는 최초로 황사 발생 시 광산란법 측정 장치의 오차를 보정하는 기술을 제시한다. $\text{PM}_{1.0}$, $\text{PM}_{2.5}$, $\text{PM}_{4.0}$, 및 PM_{10} 중 두 가지에서 네 가지의 종류의 측정값으로 황사 발생 시 저평가되는 $\text{PM}_{2.5}$ 의 값을 보정할 수 있으면서 황사가 없을 때는 큰 영향을 받지 않음을 실험 결과로 보여준다.

2. 미세먼지 측정 방법

2.1. 중량농도법

중량농도법에서는 기설정된 양의 공기를 필터에 통과시켜서 필터에 묻은 먼지의 무게를 측정한다. 먼지의 무게가 필터의 무게에 비해서 너무 작으므로 초정밀 저울이 필요하고 다량의 공기를 이용해서 많은 먼지를 필터에 흡착시켜야 측정할 수 있다.

이 방법은 측정 원리가 간단하고 측정값이 정확하다. 하지만, 새 필터 질량 측정, 필터 설치, 먼지 입자 포집, 필터 수집, 수집된 필터 질량 측정이 일련의 수작업으로 이루어지고, 12시간 또는 24시간 이상 공기를 필터에 통과해야만 측정 가능한 미세먼지의 질량이 얻어진다.

2.2. 베타선법

중량농도법의 수작업성과 느린 측정 주기를 개선하는 방법으로서, 베타선을 이용한 간접적인 무게 측정 방법은 테이프처럼 동작하는 필터에 의해서 일련의 과정이 자동으로 수행되고, 1시간 단위로 미세먼지 농도를 측정할 수 있다. 중량농도법보다 나은 편리성 때문에 국가측정소에서는 이 방법을 사용한다.

이 무게 측정 원리는 먼지가 포집된 필터에 베타선

을 투과하여 베타선의 감쇄량을 측정하여 간접적으로 무게를 측정한다[16].

베타선법 측정 장치들은 입경분리기 및 노즐 등 다양한 부품에 대한 빈번한 청소 관리, 한 시간 주기 측정, 큰 부피, 주기적 필터 관리, 높은 전력 소비와 관련된 문제를 가지고 있다.

2.3. 광산란법

먼지입자에 레이저를 노출하면 빛이 산란하는데, 입자 크기가 작을수록 빛의 산란이 많이 된다. 이 원리를 이용하여 각 먼지입자의 크기와 개수를 측정하고, 질량 보정계수로써 간접적으로 질량을 변환하는 방식이 광산란법이다[17-19]. 또한, 소량의 공기만을 사용하기 때문에 작은 크기로 설계할 수 있다.

이 방식으로 초 단위의 측정이 가능하고 휴대할 수 있는 크기로 설계할 수 있다. 하지만 먼지입자의 구성 물질에 따라서 밀도가 다르므로 먼지입자의 조성에 따라 오차가 발생한다.

3. 설계 및 구현

3.1 황사와 초미세먼지

황사는 주로 PM_{10} (미세먼지)으로 구성되어 있으며, 황사주의보 발생 시에 PM_{10} 의 농도가 크게 증가한다. 하지만 황사에는 일부 입경이 작은 $PM_{2.5}$ (초미세먼지)도 포함되어 있다. 황사 주의보나 경보가 발생할 때는 황사 성분이 미세먼지 및 초미세먼지 농도의 대부분을 차지하게 된다.

광산란법 측정 장치는 입자의 크기를 실시간으로 측정하기 때문에 다양한 입경 별 농도를 측정할 수 있다. 본 논문에서는 환경부 미세먼지 간이측정기인증에서 1등급 인증을 받은 공감센서의 GGS727 모델을 사용하였다. 이 장치는 $PM_{1.0}$, $PM_{2.5}$, $PM_{4.0}$, 및 PM_{10} 을 계측할 수 있으며 내장된 소형 에어로졸 컨디셔너를 통하여 습도에 따른 오차를 극복하였다.

광산란법 초미세먼지 측정 장치는 황사 발생 시에 큰 오차를 발생시킨다. 이 오차를 줄이는 기본 방법은 황사 발생 시에 PM_{10} 의 농도가 더 작은 먼지입자들의 농도보다 매우 더 높다는 성질을 이용하였다.

2021년 3월 1일부터 2021년 3월 31일 동안 $PM_{1.0}$, $PM_{2.5}$, $PM_{4.0}$, 및 PM_{10} 을 광산란법 초미세먼지 측정기로

계측하였다. Fig. 1은 황사가 없을 때와 있을 때, 해당 측정 기간 동안 입경 크기들에 대한 평균 비율을 보여준다.

황사가 없을 때는 PM_{10} 과 $PM_{1.0}$ 의 비율이 약 1.6:1이지만, 황사가 발생하면 PM_{10} 과 $PM_{1.0}$ 의 비율 약 4.1:1로 크게 증가하였다.

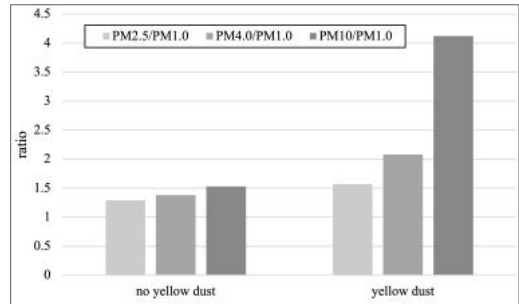


Fig. 1. The ratios of various particulate matter sizes

3.2 설계

황사는 먼지입자의 입경 분포의 변화를 발생시키기 때문에 다양한 입경 별 농도를 통한 기계학습으로 초미세먼지의 농도를 보정할 수 있다.

광산란법 초미세먼지 측정기가 $PM_{1.0}$, $PM_{2.5}$, $PM_{4.0}$, 및 PM_{10} 을 실시간에 측정하고 이 측정값들을 독립변수(x)로 입력받는 기계학습을 통하여 보정된 $PM_{2.5}$ (초미세먼지) 또는 PM_{10} (미세먼지)를 구할 수 있다.

기계학습의 참조데이터(y)는 베타선법 또는 중량농도법 측정기로 얻은 측정값을 사용하여, 황사의 유무에 상관없이 광산란 측정기의 목표데이터들(x)로부터 참조데이터(y)에 근사한 결과가 나오도록 학습시킨다.

이 기계학습은 황사가 발생하였을 때 참조데이터와 유사한 결과로 보정된 값을 출력하지만, 황사가 없을 때는 영향을 주지 않아야 한다.

본 논문에서는 다중 선형 회귀 (multiple linear regression) 기법의 기계학습을 이용하였다. 다른 기계학습 기법을 사용할 수도 있지만, 본 논문은 간단한 기계학습 기법만으로도 다수의 입경 크기별 농도를 통하여 황사에 따른 보정을 할 수 있음을 보이고 황사가 없을 때는 영향을 주지 않을 수 있음을 제시한다. 또한, 네 가지 크기별 농도 ($PM_{1.0}$, $PM_{2.5}$, $PM_{4.0}$, 및 PM_{10}) 뿐만 아니라, 이 중 세 가지 또는 두 가지 계측값만 사용하였을 때에 기계학습이 가능함을 실험 결과에서 보인다.

다중 선형 회귀 모델은 여러 독립 변수(x)로써 종속

변수(h)를 예측하는 선형 방정식을 이용한다. 선형 회귀 모델 방정식은 수학적 (1)과 같다.

$$h = w_0x_0 + w_1x_1 + w_2x_2 + w_3x_3 + b \quad (1)$$

종속 변수 h는 황사 보정된 초미세먼지 농도이며, $x_0, x_1, x_2,$ 및 x_3 는 각각 $PM_{1.0}, PM_{2.5}, PM_{4.0}$ 및 PM_{10} 농도의 측정값들이다. 회귀 계수 $w_0, w_1, w_2, w_3,$ 및 b 는 베타선법 측정기에서 얻은 정확한 참조데이터로 훈련(학습)하여 얻어진다.

훈련을 통하여 회귀계수들은 결정이 되고, 이 결정된 회귀계수들은 광산란 초미세먼지 측정기에 입력되어 실시간에 보정된 초미세먼지 측정값이 수학적 (1)에 의하여 예측된다.

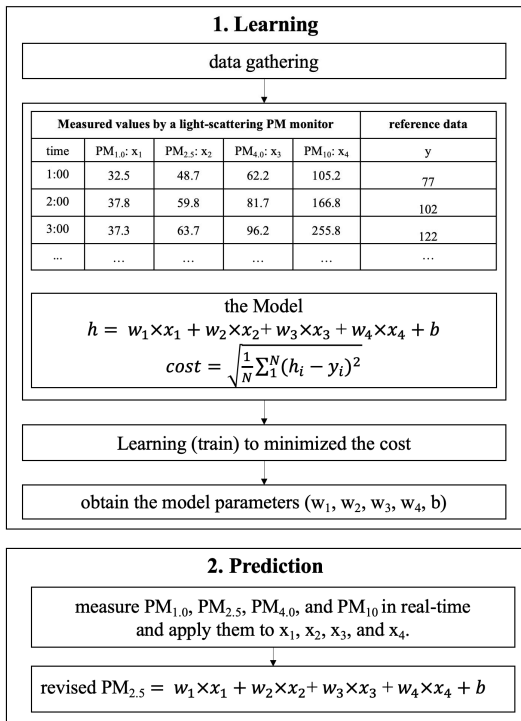


Fig. 2. The flow diagram of the machine learning (multiple linear regression).

Fig. 2는 다중 선형 회귀를 통한 기계학습의 흐름도를 보여준다. 가장 먼저 해야 할 것은 학습(훈련)에 필요한 데이터집합을 수집하는 것이다. 수집하는 데이터는 네 가지 크기의 PM농도($PM_{1.0}, PM_{2.5}, PM_{4.0}, PM_{10}$)

와 정확한 측정값으로 여겨지는 $PM_{2.5}$ 의 참조데이터(y)이다. 실험 결과에서는 GGS727 광산란법 측정 장치에서 네 가지 PM농도를 독립변수(x)로, 중부권 대기환경 연구소에서 측정한 $PM_{2.5}$ 값을 참조데이터(y)로 가져왔다. 훈련에 사용될 데이터는 황사가 있는 경우와 없는 경우가 충분히 섞여 있어야 한다. 실험 결과에 나타난 데이터집합은 2021년 3월 동안 한 시간 단위로 수집되었으며 744개의 집합으로 구성되어 있다.

크기별 PM 농도의 데이터집합을 수학적 (1)의 $x_1, x_2, x_3,$ 및 x_4 에 대입한 종속변수(h)와 참조데이터(y)의 차이에 대한 비용(제공 평균 제공근)이 최소가 되는 회귀계수 $w_1, w_2, w_3, w_4,$ 및 b 를 찾는 훈련을 수행한다.

훈련을 통하여 산출한 회귀계수들(w_1, w_2, w_3, w_4, b)은 광산란 측정 장치에 저장되어, 실시간에 측정된 $PM_{1.0}, PM_{2.5}, PM_{4.0}, PM_{10}$ 과 산출한 회귀계수들을 수학적 (1)에 대입하여 보정된 $PM_{2.5}(h)$ 를 얻는다.

데이터집합에 황사가 있는 경우와 없는 경우 모두가 포함되어 있으므로 황사가 없는 경우에도 기계학습은 올바른 데이터를 출력할 수 있다. 하지만 기계학습 보정 데이터는 원본 데이터보다 약간의 오차를 가질 수 있다. 그래서 황사가 없을 때는 기계학습 보정을 하지 않는 필터링 기법을 사용할 수 있다.

필터링 기법에 있어서, 황사가 있는지를 판단하는 방법으로서 PM_{10} 과 $PM_{2.5}$ 의 비율이 높으면 황사가 있다고 판단할 수 있다. 예를 들어 PM_{10} 농도/ $PM_{2.5}$ 농도가 기설정된 상수 이상이면 황사가 있고, 그렇지 않으면 없다고 판단할 수 있다.

4. 실험 결과

대전광역시 문화동에 있는 국립환경과학원(NIER) 중부권대기환경연구소의 측정값을 기계학습의 참조데이터(y)로 사용하였다. 이 한 시간 주기의 측정값은 에어코리아를 통하여 공개된다. 광산란법 초미세먼지 측정기인 GGS727은 이 측정소에서 가까운 곳에 설치하였다.

실험에 사용된 독립변수를 위한 데이터집합은 Fig. 3에 나타나 있다. 2021년 3월 1일에서 31일까지 한 달 동안 한 시간 간격으로 GGS727에서 수집한 측정값을 사용되었다. 이 기간에 2차례의 6일간의 황사가 발생하였다. 369에서 404시에 약한 황사가 나타났으며, 674에서 715시에 큰 황사가 발생하였다. 700시경에 가장 높은 황사농도가 발생하였는데 이때 PM_{10} 이 $PM_{4.0}$

이 비해서 유난히 높다는 것을 알 수 있다. 2021월 20 일에서 31일간의 데이터로 회귀계수를 구했으며, 전 구 간의 데이터로 성능 평가하였다.

NIER의 중부권대기환경연구소는 PM₁₀은 공개하지 않고 PM_{2.5}만 공개하였기에 우리는 PM_{2.5}에 대해서 기계학습 보정을 시행하였다.

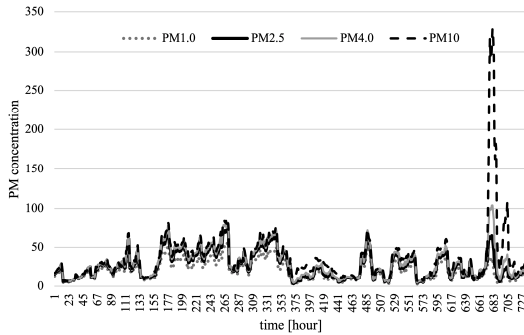


Fig. 3. The dataset for x1, x2, x3, and x4 acquired in Munhwa-dong Daejeon, South Korea during March 2021.

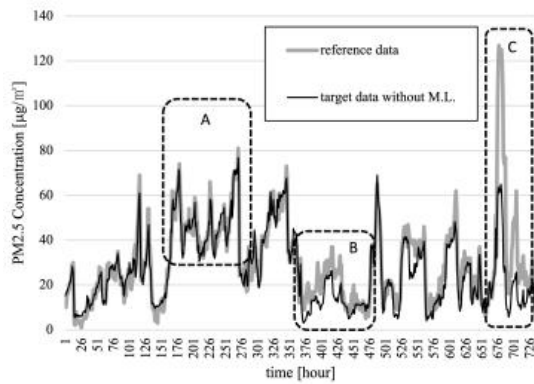


Fig. 4. The reference data of national institute of environmental research (NIER) and the target data of the light-scattering PM monitor GGS20 during March 2021.

Fig. 4는 해당 시험 기간의 NIER의 초미세먼지 참조데이터(y)와 GGS727의 초미세먼지 목표데이터(x)를 한 시간 간격으로 비교한 그래프이다. 이 그림의 구간 B와 구간 C에서 황사가 발생하였으며 구간 C는 황사 주의보가 발령되었었다. 이 황사 구간 동안 GGS727에서 얻은 목표데이터(검정선)는 중부권대기환경연구소의 계측값인 참조데이터(회색선)보다 저평가되어 있음을 알 수 있다.

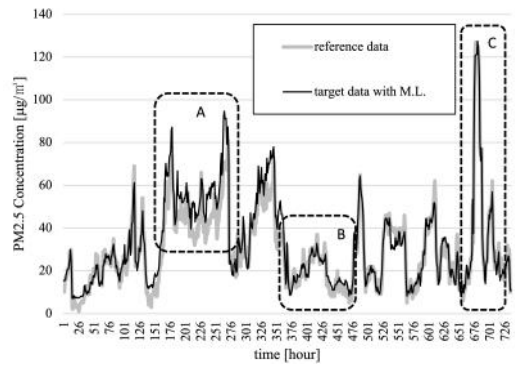


Fig. 5. The reference data of NIER and the target data that is compensated by machine learning.

Fig. 5는 기계학습을 적용하여 보정한 PM_{2.5} 농도(목표데이터)와 참조데이터의 PM_{2.5} 농도를 보여준다. 황사와 관계없이 3월 20일부터 31일까지의 데이터로 기계학습을 수행하였으며, 3월 1일부터 31일간의 데이터에 대해서 보정이 적용되었다.

이 기계학습은 python을 이용한 tensorflow[20]로 구현하였으며, 이 훈련을 통하여 얻은 회귀계수 w₁, w₂, w₃, w₄, 및 b는 각각 -2.543, 0.414, 2.775, -0.312, 2.708이었다.

황사가 있는 구간 B와 구간 C에서 참조데이터가 목표데이터와 매우 유사하게 되었음을 알 수 있다. 즉 기계학습으로 잘 보정이 되었다. 하지만 Fig. 5의 구간 A에서는 황사가 없는 구간인데, Fig. 4와 비교하면 기계학습을 하지 않은 것이 더 좋다는 것을 알 수 있다.

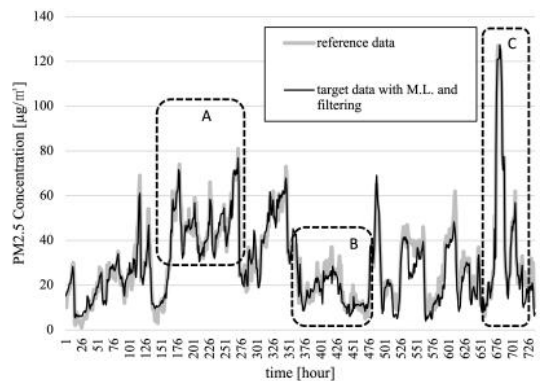


Fig. 6. The reference data of NIER and the target data of GGS20 with the filtering scheme and machine learning.

그래서 우리는 필터링 기법을 사용하여 황사가 없을 때는 기계학습을 적용하지 않을 수 있다. Fig. 6은 기계 학습과 필터링 기법으로 보정한 시간에 따른 PM_{2.5}의 농도를 보여준다. Fig. 6의 구간 A는 Fig. 5의 구간 A 보다 확연히 좋아진 것을 확인할 수 있다.

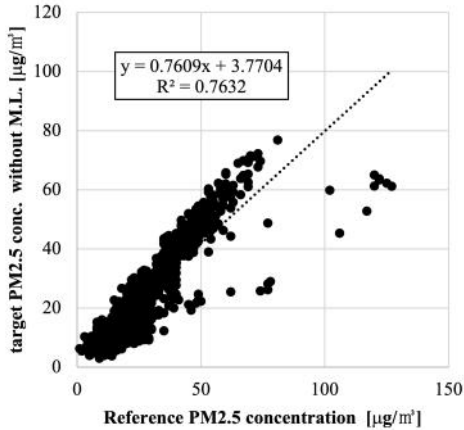


Fig. 7. The comparison of the reference data and the target data without machine learning.

시험 기간 중 황사가 없는 구간에 대한 참조데이터와 목표데이터의 상관계수(R²)는 0.927이었다. Fig. 7은 황사를 포함한 전체 시험 기간의 참조데이터와 목표데이터의 상관 관계를 보여준다. 황사가 있는 포인트가 추세선에서 크게 이탈하였으며 0.763의 낮은 상관계수가 나타났다.

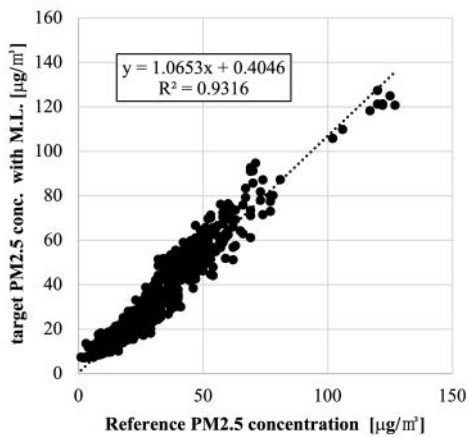


Fig. 8. The comparison of the reference data and the target data with machine learning.

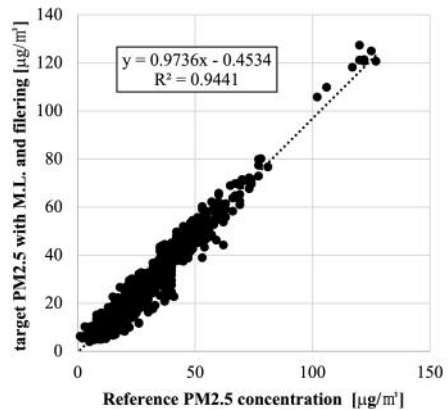


Fig. 9. The comparison of the reference data and the target data with machine learning and the filtering scheme.

Fig. 8과 Fig. 9는 각각 기계학습만을 적용하여 보정한 데이터와, 기계학습과 필터링을 적용하여 보정한 데이터에 대해서 참조데이터와 비교한 상관관계를 보여준다. Fig. 8은 기계학습이 적용되어 보정한 PM_{2.5} 농도(목표데이터)와 참조데이터를 비교하였다. Fig. 9는 황사 발생을 스스로 감지하는 필터링 기법을 사용하여 황사 구간에만 보정이 적용되었다.

적용된 필터링 기법에서는 PM₁₀/PM_{2.5}가 1.7 이상이면 황사가 있다고 판단하고 기계학습으로 PM_{2.5}를 보정하였다.

Fig. 8에서 볼 수 있듯이 기계학습이 있는 경우에는 기울기가 약 1.065, 결정계수가 0.932로 기계학습이 없는 경우(Fig. 7의 결과)에 비해서 크게 개선되었다.

Fig. 9는 필터링까지 적용하여 보정한 PM_{2.5} 농도와 NIER의 참조데이터를 비교한 상관관계를 보여주고 있다. 기울기는 0.974, 절편은 -0.453이었고, 결정계수는 0.944로써 기계학습만 한 것(Fig. 8의 결과)보다 더 개선되었다. Fig. 9에서 0.944의 결정계수(R²)는 0.972의 상관계수(R)로 변환할 수 있다. 미국 EPA의 광산란식 측정기(class III)의 형식승인기준에서는 상관계수(R)가 0.95 이상, 기울기가 0.88에서 1.12 사이 이내이고 절편은 -7에서 7사이여야 한다[21]. Fig. 9의 결과는 이 기준 안에 들어간다.

이상의 기계학습 결과에서는 PM_{1.0}, PM_{2.5}, PM_{4.0}, 및 PM₁₀이 입력으로 사용되었다. 하지만 이 중에서 세 개 또는 두 개만을 적용한 기계학습으로도 황사 보정이 가능할 수 있다.

Table 1은 여러 가지 입력 조건에 대한 추세선 및 결정계수를 보여준다. 기계학습 및 필터링을 모두 사용한 경우가 가장 좋은 결과를 보였다. 세 가지 입력을 사용한 기계학습은 0.942에서 0.944의 결정계수를 보임으로써 0.932의 결정계수를 가지는 네 가지 입력을 사용하는 것보다 좋은 결과를 가졌다. 세 개의 입력이 네 개의 입력보다 빠르고 쉽게 훈련되기 때문이었다.

하지만, 두 가지 입력만을 사용한 경우 0.913에서 0.939의 결정계수를 보임으로써 황사 보정이 효과가 있음이 나타났다. 하지만, 서너 가지 입력을 사용한 것 보다는 조금 낮은 결정계수를 보인다.

Table 1. The trend lines and R square of the machine learning with various inputs.

	slope	intercept	R ²
without M.L.	0.761	3.770	0.763
M.L. and filtering with four PMs	0.974	-0.453	0.944
M.L. with four PMs	1.065	0.405	0.932
M.L. with (PM _{1.0} , PM _{2.5} , PM _{4.0})	1.062	-0.308	0.944
M.L. with (PM _{1.0} , PM _{2.5} , PM ₁₀)	1.030	0.458	0.942
M.L. with (PM _{2.5} , PM _{4.0} , PM ₁₀)	1.028	0.33	0.944
M.L. and filtering with (PM _{2.5} , PM _{4.0} , PM ₁₀)	1.034	-0/805	0.945
M.L. with (PM _{1.0} , PM _{2.5})	1.115	-0.552	0.913
M.L. with (PM _{2.5} , PM _{4.0})	1.036	0.005	0.927
M.L. with (PM _{2.5} , PM ₁₀)	1.009	0.729	0.939

5. 결론

일반 미세먼지보다 밀도가 높은 황사는 먼지입자의 부피를 측정하여 무게를 추정하는 광산란법 초미세먼지 측정 장치에 큰 오차를 발생시킨다. 또한, 황사가 발생하였을 때는 PM₁₀ 농도가 PM_{2.5}보다 매우 높은 특징을 가지고 있다.

광산란법 초미세먼지 측정 장치는 하나의 센서로 다양한 입경 크기별로 농도를 측정할 수 있다. PM_{1.0}, PM_{2.5}, PM_{4.0}, 및 PM₁₀ 과 같은 입경 크기별 농도를 독립변수로 입력받는 기계학습을 수행하여 황사 발생 시 저평가되는 PM_{2.5}의 값을 보정할 수 있음을 보였다. 또한, 황사가 없을 때는 큰 영향을 받지 않음을 실측된 실험 결과에서 확인하였다. 황사 발생 여부를 스스로 판별하는 필터링 기법을 통하여 황사가 없을 때의 오차를 향상할 수 있었다.

황사 발생이 있었던 2021년 3월 대전 문화동에서 광산란 기기로 측정된 데이터와 중부권대기환경연구

소(문화동)의 PM_{2.5} 측정결과의 상관관계(R²)가 황사 때문에 0.763에 불과한 것을 본 논문에서 제시한 기법으로 0.944로 개선할 수 있었다.

세 가지 종류의 PM 크기 별 측정값을 독립변수로 입력받는 기계학습과 필터링 기법을 사용하는 경우가 가장 좋은 결과를 얻었으며, 두 가지 종류의 독립변수로도 상당히 좋은 결과를 얻을 수 있었다.

REFERENCES

- [1] J. Lelieveld, K. Klingmüller, A. Pozzer, U. Pöschl, M. Fnais, A. Daiber & T. Münzel. (2019). Cardiovascular disease burden from ambient air pollution in Europe reassessed using novel hazard ratio functions. *European heart journal*, 40(20), 1590-1596. DOI : 10.1093/eurheartj/ehz135
- [2] B. Choubin, M. Abdolshahnejad, E. Moradi, X. Querol, A. Mosavi, S. Shamshirband & P. Ghamisi. (2020). Spatial hazard assessment of the PM10 using machine learning models in Barcelona, Spain. *Science of The Total Environment*, 701, 134474. DOI : 10.1016/j.scitotenv.2019.134474
- [3] Kampa, M., & Castanas, E. (2008). Human health effects of air pollution. *Environmental pollution*, 151(2), 362-367. DOI : 10.1016/j.envpol.2007.06.012
- [4] X. Wu, R. C. Nethery, B. M. Sabath, D. Braun & F. Dominici. (2020). *Exposure to air pollution and COVID-19 mortality in the United States*. MedRxiv.
- [5] Teledyne Advanced Pollution Instrumentation. (2018). User manual model T640 PM mass monitor. Teledyne API [Online], <http://www.teledyne-api.com/>
- [6] Grimm Aerosol. (2012). GRIMM EDM 180 dust monitor, NAQC-EPA Monitoring Conference Denver. Colorado May 14th.
- [7] D. S. Kang, J. E. Oh, S. Y. Lee, H. J. Shin, H. K. Bong & D. S. Kim. (2018). Development and performance evaluation of a real-time PM monitor based on optical scattering method. *Particle and Aerosol Research*, 14, 107-119. DOI : 10.11629/jpaar.2018.14.4.107
- [8] S. Son. & J. Kim, (2020). Evaluation and Predicting PM 10 Concentration Using Multiple Linear Regression and Machine Learning. *Korean Journal of Remote Sensing*, 36(6_3), 1711-1720. DOI : 10.7780/kjrs.2020.36.6.3.7

- [9] K. S. Harishkumar, K. M. Yogesh & I. Gad. (2020). Forecasting air pollution particulate matter (PM2.5) using machine learning regression models. *Procedia Computer Science*, 171, 2057-2066. DOI : 10.1016/j.procs.2020.04.221
- [10] S. Abdullah, N. N. L. M. Napi, A. N. Ahmed, W. N. W. Mansor, A. A. Mansor, M. Ismail & Z. T. A. Ramly. (2020). Development of multiple linear regression for particulate matter (PM10) forecasting during episodic transboundary haze event in Malaysia. *Atmosphere*, 11(3), 289. DOI : 10.3390/atmos11030289
- [11] L. A. Díaz-Robles, J. C. Ortega, J. S. Fu, G. D. Reed, J. C. Chow, J. G. Watson & J. A. Moncada-Herrera, (2008). A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: The case of Temuco, Chile. *Atmospheric Environment*, 42(35), 8331-8340. DOI : 10.1016/j.atmosenv.2008.07.020
- [12] M. D. Mallet. (2021). Meteorological normalisation of PM10 using machine learning reveals distinct increases of nearby source emissions in the Australian mining town of Moranbah. *Atmospheric pollution research*, 12(1), 23-35. DOI : 10.1016/j.apr.2020.08.001
- [13] J. M. Lim. (2019). An Estimation Model of Fine Dust Concentration Using Meteorological Environment Data and Machine Learning. *Journal of Information Technology Services*, 18(1), 173-186. DOI : 10.9716/KITS.2019.18.1.173
- [14] A. Ibrir, Y. Kerchich, N. Hadidi, H. Merabet & M. Hentabli. (2021). Prediction of the concentrations of PM1, PM2.5, PM4, and PM10 by using the hybrid dragonfly-SVM algorithm. *Air Quality, Atmosphere & Health*, 14(3), 313-323. DOI: 10.1007/s11869-020-00936-1
- [15] S. H. Sani, A. Rony, F. I. Karim, M. F. Mridha & M. A. Hamid, (2021). Evaluate and Predict Concentration of Particulate Matter (PM 2.5) Using Machine Learning Approach. In *International Conference on Innovative Computing and Communications*. Springer, Singapore. 771-785. DOI: 10.1007/978-981-15-5148-2_67
- [16] K. Yoo, H. Yoo, J. M. Lee, S. K. Shukla & J. Park. (2018). Classification and regression tree approach for prediction of potential hazards of urban airborne bacteria during Asian dust events. *Scientific reports*, 8(1), 1-11. DOI : 10.1038/s41598-018-29796-7
- [17] Y. Chun, J. Kim, J. C. Choi, K. O. Boo, S. N. Oh & M. Lee. (2001). Characteristic number size distribution of aerosol during Asian dust period in Korea. *Atmospheric Environment*, 35(15), 2715-2721. DOI : 10.1016/S1352-2310(00)00404-0
- [18] W. Chueinta & P. K. Hopke. (2001). Beta gauge for aerosol mass measurement. *Aerosol Science & Technology*, 35(4), 840-843. DOI : 10.1080/027868201753227398
- [19] D. S. Kim, Y. K. Cho & Y. H. Yoon. (2014). Development of a real-time monitoring device for measuring particulate matter, *Particle and Aerosol Research*, 10, 1-8. DOI : 10.11629/jpaar.2014.10.1.001
- [20] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean & X. Zheng. (2016). Tensorflow: A system for large-scale machine learning. In *12th USENIX symposium on operating systems design and implementation (OSDI 16)*, 265-283.
- [21] Environmental Protection Agency. (2007). *Test procedure for class II and class III methods for PM 2.5 and PM*, Legal Information Institute, 2007.

백 승 훈(Sung Hoon Baek)

[정회원]



- 1997년 2월 : 경북대학교 전자공학과(공학사)
- 1999년 2월 : 한국과학기술원 전기 및 전자공학과(공학석사)
- 1999년 1월 ~ 2005년 1월 : 한국전자통신연구원 연구원
- 2008년 8월 : 한국과학기술원 전기 및 전자공학전공(공학박사)
- 2008년 9월 ~ 2011년 2월 : (주) 삼성전자 책임
- 2011년 3월 ~ 현재 : 중원대학교 컴퓨터공학과 교수
- 관심분야 : 초미세먼지 측정, 운영체제, 컴퓨터 스토리지, 머신러닝
- E-Mail : shbaek@jwu.ac.kr