

환경 모니터링을 위한 EDA 기반 데이터 분석

홍명우¹ · 강윤희^{2*}

¹우송정보대학 스마트팩토리기술과 교수

^{2*}백석대학교 컴퓨터공학부 교수

EDA based Data Analysis for Environmental Monitoring

Myoung-Woo Hong¹ · Yun-Hee Kang^{2*}

¹Professor, Department of Smart Factory Technology, Woosong College, Daejeon 34518, Korea

^{2*}Professor, Division of ICT, Baekseok University, Chunan 31065, Korea

[요약]

최근 센서 데이터 활용 영역이 넓어지면서 데이터 분석 서비스가 활성화되고, 분석을 용이하게 할 수 있는 환경으로 진화하고 있다. 이에 따라 센서 데이터의 신뢰성 보장이 필요하다. 신뢰성을 갖는 환경 모니터링을 위해서는 센서로부터 수집된 환경 데이터의 분포 및 값을 살펴본 후 데이터가 표현하는 현상을 더 잘 이해하고, 센서 및 센서 데이터에 대한 잠재적인 특이점의 발견을 선행하여야 한다. 이를 위해 EDA를 통해 수집된 센서 값을 시각화하고 분석에 주어진 데이터의 개별 속성의 특징 및 상관관계를 도출한다. 이 실험을 통해, 온도 측정값의 이상치, CO₂와 습도 사이의 양의 상관관계, PM_{2.5}와 PM_{10.0}의 센서 값의 패턴 유사성, TVOC의 요일별 특성을 발견하였다. 본 연구의 EDA 분석 결과는 센서 데이터의 신뢰성을 평가하기 위해 사용할 수 있으며 실내 환경에 미치는 상관관계를 활용한 환경관리 제어 시스템 개발에 활용할 수 있다.

[Abstract]

Recently, as the field of sensor data utilization has been expanded, data analysis services are being activated, and it is evolving into an environment that can facilitate analysis. Accordingly, it is necessary to guarantee the reliability of the sensor data. For reliable environmental monitoring, after examining the distribution and value of environmental data collected from sensors, it is necessary to better understand the phenomena expressed by the data, and to discover potential singularities for sensors and sensor data. To this end, we visualize the sensor values collected through EDA and derive the characteristics and correlations of the individual properties of the data given to the analysis. Through this experiment, we found anomaly in temperature measurement values, positive correlation between CO₂ and humidity, pattern similarity between PM_{2.5} and PM_{10.0} sensor values, and daily characteristics of TVOC. The results of the EDA analysis of this study can be used to evaluate the reliability of sensor data and can be used to develop an environmental management control system utilizing the correlation on the indoor environment.

색인어 : 탐색적 데이터 분석, 환경 모니터링, 센서, 센서 데이터, 사물인터넷

Keyword : EDA(Exploratory Data Analysis), Environmental Monitoring, Sensor, Sensor Data, IoT

<http://dx.doi.org/10.9728/dcs.2022.23.7.1289>



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.

Received 09 June 2022; **Revised** 05 July 2022

Accepted 20 July 2022

***Corresponding Author; Yun-Hee Kang**

Tel: +82-41-550-0504

E-mail: yhkang@bu.ac.kr

I. 서론

사물 인터넷 (IoT; Internet of Things)은 사물에 센서와 통신 기능을 내장하여 인터넷에 연결하는 기술로써, 자율주행 자동차, 스마트공장, 스마트홈 등의 분야에서 스마트기기를 인터넷에 연결하여 통합하는 기반 환경 구축으로 활용되고 있다[1-5]. 이러한 IoT 환경 구축은 가정, 도시, 공장 등 어느 곳에서나 적용가능하게 됨에 따라 센서 활용 또한 상당히 증가하고 있다[6]. 이에 다양한 분야에서 센서를 활용한 모니터링이 지속적으로 이루어지고, 이러한 모니터링을 위한 센서 데이터 활용 영역이 넓어지면서 데이터 분석 서비스가 활성화되고 있다[7].

데이터 분석 서비스가 활성화됨에 따라, 최근에는 센서 데이터의 분석을 용이하게 할 수 있는 환경으로 진화하고 있다 [8]. 아마존 AWS IoT는 데이터 수집을 지원하는 플랫폼으로 데이터 분석 활용 영역이 넓어지면서 빅데이터 분석 서비스가 활성화되고, 그에 따른 분석도 용이해지고 있다. 데이터 환경이 바뀌면서 클라우드를 중심으로 한 중앙집중식 컴퓨팅으로 진화하고 있으나, 데이터 센서 수집단에서 센서 데이터에 대한 신뢰성을 평가하는 것은 데이터 전송에 따른 높은 비용의 제약점을 갖는 문제가 있다.

이를 해결하기 위해 엣지 컴퓨팅 (edge computing)에 대한 기술이 다양한 분야에서 적용되고 있다. 엣지 컴퓨팅은 데이터와 가장 가까운 곳에서 컴퓨팅을 수행하도록 함으로서 많은 데이터를 중앙 클라우드까지 전송하지 않고 데이터가 생성된 곳에서 바로 처리하도록 하며, 이를 통해 짧은 지연시간과 대역폭 효율성의 이점을 얻을 수 있다. 이와 함께 데이터의 구조적 특성을 알아내기 위한 기법인 탐색적 데이터 분석 (EDA; exploratory data analysis)을 활용할 수 있다. EDA는 수집한 데이터에 대해 다양한 각도에서 관찰하고 이해하는 과정으로 상세한 데이터를 분석하기 전에 그래프나 통계적인 방법으로 자료를 직관적으로 바라보는 과정이다[9-11].

기존 센서 평가는 주로 내구성 시험과 전자파 인증 및 친환경 제품 인증과 같은 위해성 검사에 한정되어 있으며 센서 데이터의 신뢰성 검사는 제한적이다. 이를 해결하기 위해서는 ICT 기반 센서 데이터의 신뢰성 제고를 위한 환경구축이 요구된다.

신뢰성을 갖는 환경 모니터링을 위해서는 센서로부터 수집된 환경 데이터의 분포 및 값을 살펴본 후 데이터가 표현하는 현상을 더 잘 이해하고, 센서 및 센서 데이터에 대한 잠재적인 특이점을 발견하는 과정이 선행하여야 한다.

본 논문에서는 센서 수집을 담당하는 엣지단에서 획득된 센서 데이터를 기반으로 신뢰성 평가에 필요한 센서 특성을 이해하는 것을 목적으로 한다. 이를 위해 EDA를 통해 수집된 센서 값을 시각화하고 분석에 주어진 데이터의 개별 속성의 특징 및 상관관계를 도출한다. 이는 측정인자에 따른 출력 특성을 알고 있는 센서를 이용하여 현재 시간과 공간에서 측정하고자 하는 환경인자 값을 도출하고, 동일 공간 및 동일 시간대의 다른 센서에 대한 출력을 비교하여 비교군 센서의 특

성을 관찰한다. 이때 EDA 분석 결과는 센서 데이터의 신뢰성을 평가하기 위해 사용한다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구로써 EDA, 표준값 z 및 상관관계를 소개하였다. 3장에서는 본 연구에서 설계한 데이터 분석 및 실험환경을 기술하고 이상치 분석과 상관관계 분석 등의 실험결과를 보여 준다. 마지막으로 4장에서 결론과 향후 연구를 제시한다.

II. 관련연구

2-1 EDA

EDA는 데이터의 구조적 특성을 알아내기 위한 기법으로 이것의 목표는 통계적 가설을 확인하는 것이 아니라 탐색, 조사 및 통찰에 있다[12]. 이때 구체적으로 데이터 시각화, 결측치 (missing value) 처리, 이상치 (anomaly) 검출, 파생변수 (derived variable) 생성과 같은 프로세스가 이루어진다[13].

결측치란 응답자의 실수 혹은 고의로 데이터를 입력하지 않았을 때 발생하는 값으로 처리하는 방법에는 2가지가 있는데, 결측치를 제거하는 방법과 제거하지 않고 다른 값으로 대체하는 방법이다. 일반적으로 생각할 때 결측치는 제거하는 것이 낫지만, 대체하는 방법을 적용하기도 하는데 그 이유는 결측치를 제거하면 중요한 데이터가 손실될 수도 있기 때문이다. 따라서 결측치를 특정값, 평균 및 최빈값으로 대체한다.

이상치란 정상적인 분포에서 벗어나는 값으로 이것을 그대로 두면 정상적인 분석 결과를 도출하는 과정에서 오류를 만들어낼 수 있다. 이상치 처리는 범주형 변수와 연속형 변수로 구분하여 이루어진다. 범주형 변수는 성별과 같은 범주로 구분할 수 있는 변수이고, 연속형 변수는 수치로 측정할 수 있는 수치 변수이다.

2-2 표준값 z

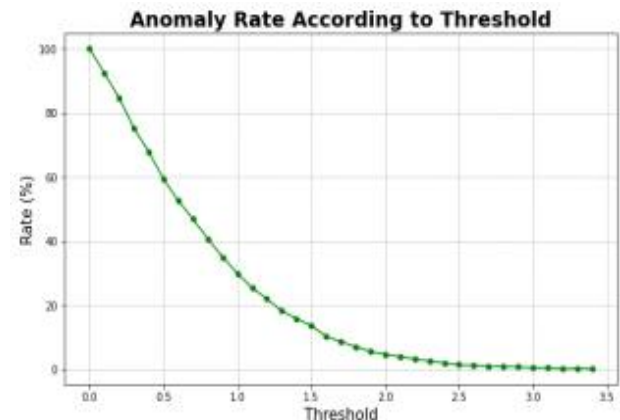


그림 1. 임계값에 따른 이상데이터 비율
Fig. 1. Anomaly data rate according to Threshold

온도센서와 압력센서를 통해 센서데이터를 주기적으로 얻고 센서데이터가 이상데이터인지 정상데이터인지 판단하도록 하였다. 센서 종류에 따라 센서데이터 값의 범위가 서로 다르기 때문에 먼저 센서데이터에 대한 표준값을 정해야 한다.

본 연구에서는 측정된 센서데이터를 x 라 가정할 때 평균과 편차를 이용하여 다음 수식 1과 같이 표준값 z 를 정의하였다.

$$z = \frac{(x - \mu)}{\sigma} \quad (1)$$

센서데이터를 표준화하면 표준값은 0을 중심으로 분포하게 된다. 이와 같이 표준화한 이후에 절대값의 threshold 값을 결정하여 [-threshold, +threshold] 범위에서 벗어나는 센서데이터를 이상치로 정하면 된다. 이 때 Threshold 값을 얼마로 정하는 것이 적당한지 결정해야 한다. Threshold 값에 따라 이상치 데이터의 양이 늘어날 수도 있고 줄어들 수도 있다.

그림 1은 1970년부터 현재까지 매일 측정된 온도센서 데이터에 대해서 threshold 값에 따라 이상데이터의 비율을 나타낸 것이다.

Threshold 값이 3일 때 이상데이터의 비율은 전체 데이터의 약 0.5%가 되기 때문에 적절하다. 이상 데이터를 감지하기 위해 threshold 값을 3으로 정하면 센서로부터 측정된 데이터가 정상범위에 속하는지 이상범위에 속하는지 감지가 가능하게 된다.

2-3 상관관계

상관관계는 2개 변수가 선형 관계에 있는 범위를 표현하는 통계적 측도이다. 원인과 결과에 관한 표현 없이 간단한 관계를 설명하는 일반적인 도구이다[14]. 상관관계는 연속적 속성을 갖는 두 변인들 간 상호 연관성에 대한 기술통계 정보를 제공해 줄 뿐만 아니라, 두 변인 간의 상호 연관성에 대한 통계적 유의성을 검증해 주는 통계분석 기법이다.

따라서 상관관계는 연구자가 관심을 갖고 연구하고자 하는 현상들 간 관계, 이를 통계학 용어를 사용해 변인들 간의 관계를 나타낸다. 만일 변인들 간의 상호 연관성만을 간략하게 설명한다면 상관관계분석은 기술통계가 되지만, 만일 변인들 간의 상호 관계에 대한 통계적 유의성을 검증함으로써 모집단의 속성을 추론하고자 한다면, 상관관계분석은 추론통계가 된다.

변인들 간의 관계는 크게 두 가지 속성으로 이해할 수 있다. 하나는 변인들 간의 상호 연관성 강도이고, 또 다른 하나는 변인들 간의 연관성 방향이다.

본 실험에서는 습도와 CO2 및 PM (particulate matter)2.5와 PM10의 각각의 측정값들의 상관관계를 시각화하여 표현하고자 한다. 각각의 두 가지 변수를 비교하여 선형관계를 한 눈에 비교할 수 있도록 하는 것이다.

III. EDA 기반 데이터 분석

3-1 데이터 분석 및 실험 환경

본 실험은 실내 공기질 모니터링을 위해 2020년 4월 2일부터 2020년 4월 19일까지 PM2.5, PM10.0, 온도, 습도, TVOC, CO2 센서로부터 전달된 센싱 데이터를 기반으로 한다. 이를 위해 입자의 크기가 2.5 μ m 미만인 먼지인 초미세먼지 측정을 위해 PM2.5와 입자의 크기가 10 μ m 미만인 먼지 PM10.0 센서를 사용한다. 공기질 측정을 위해 대기 중에 액체 및 기체 유기화합물을 총칭하는 TVOC (total volatile organic compounds)와 실내의 유해상태를 표현하는 CO2를 값을 획득하기 위한 센서를 사용한다.

수집된 센서값은 개별 센서의 통계적 특성과 센서값 간의 특징을 탐색하기 위해 EDA를 수행한 후 센서 간의 상관관계를 도출한다. 기초적인 데이터 분석은 파이썬의 Pandas 라이브러리를 사용하여 수행한다.

3-2 실험 및 분석 결과

1) 이상치 분석

그림 2는 수집된 온도 정보의 변화를 나타낸 그래프이다. 온도 정보의 평균값은 23.85, 최소값은 19.6, 최고값은 27.3이다.

온도 정보를 사분위로 분류하는 박스플롯으로 나타내면 그림 2와 같이 나타낼 수 있는데 25%인 Q1은 23.3, 50%인 Q2는 23.8, 75%인 Q3은 24.5를 갖는다. 본 연구에서는 이를 기반으로 하위 25%에 속한 데이터와 상위 25%에 속한 데이터를 대상으로 이상 값, 즉 이상치 검출을 시행하였다. 검출에는 수식 2를 이용하였다.

이상 값을 추출하였을 때 전체 자료의 0.72%가 이상 값으로 검출되었다. 그렇지만 사분위 값을 기준으로 하면 전체 샘플의 자료 분포에 대한 상대적인 자료특성을 반영하지 못하는 제약점을 가진다.

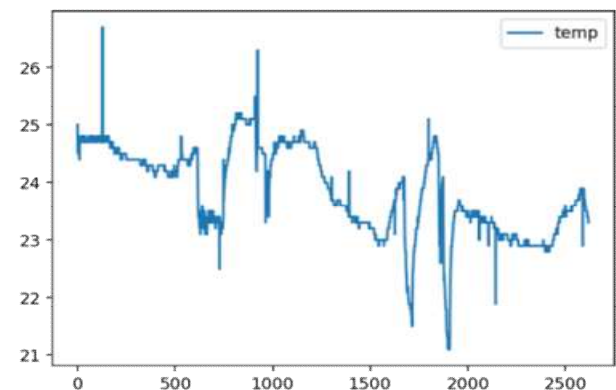


그림 2. 온도 관측 값 플롯팅
Fig. 2. Plotting Temperature Observations

$$Temp < (Q1 - 1.5 * IQR) \text{ or } Temp > (Q3 + 1.5 * IQR) \quad (2)$$

where $IQR = Q3 - Q1$

표 1. 임계값 변화에 따른 이상 값 비율
Table 1. Percentage of anomaly data with Threshold Changes

Threshold value	Percentage of anomaly data
1	33%
1.5	7%
2	2.5%
2.5	1.7%
3	0.76%
3.5	0.34%

표 2. 데이터 셋의 통계 값
Table 2. Statistical values in the data set

	Humidity	CO2
count	225919	225919
mean	26.715375	4.709173
std	13.300000	1.380514
min	19.600000	1.880000
25%	22.300000	3.870000
50%	25.500000	4.540000
75%	30.500000	5.180000
max	43.400000	8.310000

이에 본 연구에서는 z-value를 사용하여 임계값을 조정하여 이상 값을 검출하도록 하였다. 이상 값을 얻기 위한 임계값을 1에서 3.5 사이의 값으로 정했을 때 이상 값의 비율은 표 1과 같다.

임계값이 3인 경우 0.76%로 사분위 값을 기반으로 한 이상 값 추출 비율 0.72와 유사함을 보였다. z-value를 기반으로 한 이상 값의 검출은 유효성을 제공하며, 다른 센서와의 융합을 통해 임계값을 구할 수 있는 장점이 있다.

2) 상관관계 분석

표 2는 실험에 사용한 센서 데이터 중 습도와 CO2의 통계 정보 (descriptive statistics)를 보인 것이다. 225919의 전체 관측치를 가지며, 1초에 6개의 센서 데이터를 획득하여 수 집서버에 전달한다.

그림3과 그림 4는 5개의 최저 및 최고 습도 측정값에 따른 다른 5개의 센서 값과의 측정값을 기술한 것으로 습도 값은 CO2와 연관성이 있음을 관찰할 수 있다.

	REG_DATE	PM25	PM100	Temperature	Humidity	TVOC	CO2
5071	2020-04-05 12:31	15.38	18.57	19.5	13.3	1.55	1.92
5072	2020-04-05 12:32	16.54	18.57	19.5	13.4	1.56	1.98
5073	2020-04-05 12:33	15.67	18.57	19.6	13.4	1.53	1.90
5074	2020-04-05 12:34	15.09	12.38	19.6	13.4	1.55	1.98
5070	2020-04-05 12:30	15.67	18.57	19.5	13.5	1.54	1.98

그림 3. 최저 습도 값의 타 센서 측정 값
Fig. 3. Other sensor measurements of the lowest humidity value

	REG_DATE	PM25	PM100	Temperature	Humidity	TVOC	CO2
25735	2020-04-19 20:55	10.25	10.25	20.7	43.4	1.88	7.95
25736	2020-04-19 20:56	10.01	10.01	20.7	43.4	1.88	7.98
25737	2020-04-19 20:57	9.38	9.38	20.7	43.4	1.86	7.97
25740	2020-04-19 21:00	8.97	8.97	20.7	43.4	1.89	7.92
25741	2020-04-19 21:01	9.67	9.67	20.7	43.4	1.87	7.93

그림 4. 최고 습도 값의 타 센서 측정 값
Fig. 4. Other sensor measurements of the highest humidity values

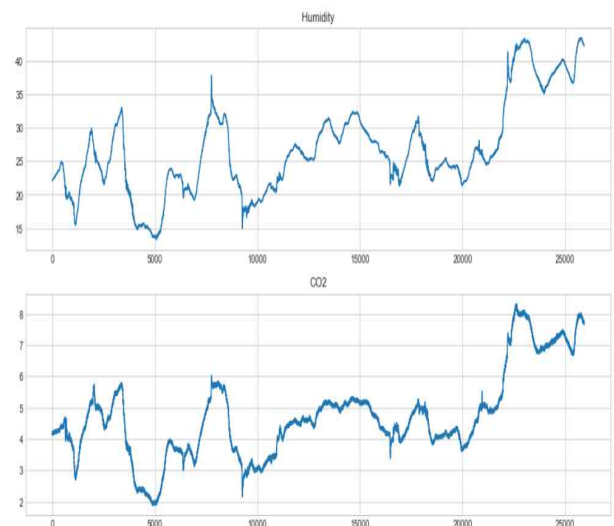


그림 5. 습도와 CO2 관측 값 플롯팅
Fig. 5. Plotting Humidity and CO2 Observations

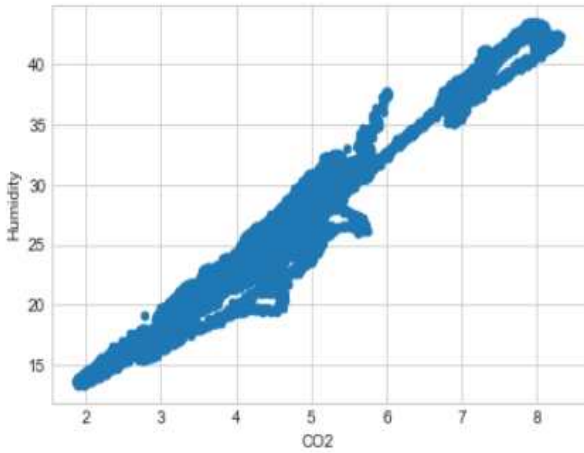


그림 6. 습도와 CO2 상관관계
Fig. 6. Humidity and CO2 Correlation

그림5는 습도와 CO2의 측정값을 시각화하여 연관성을 플롯팅한 것으로 시각화를 통해 두 센서의 변화추이를 관찰할 수 있다. 관찰되어진 내용을 볼 때, 실내에 사람이 있는 경우 호흡을 통해 CO2의 증가가 발생하며, 이는 습도 값에도 영향이 있음을 분석할 수 있다.

그림 6은 피어슨 상관관계 (pearson correlation)를 보인 것으로 두 변수 간의 선형 상관관계를 표현한다. 이는 센서 자료인 습도와 CO2의 측정값을 기반으로 표현된 상관관계를 보인 것으로 두 센서 값은 양의 상관관계를 갖고 있음을 관찰할 수 있다.

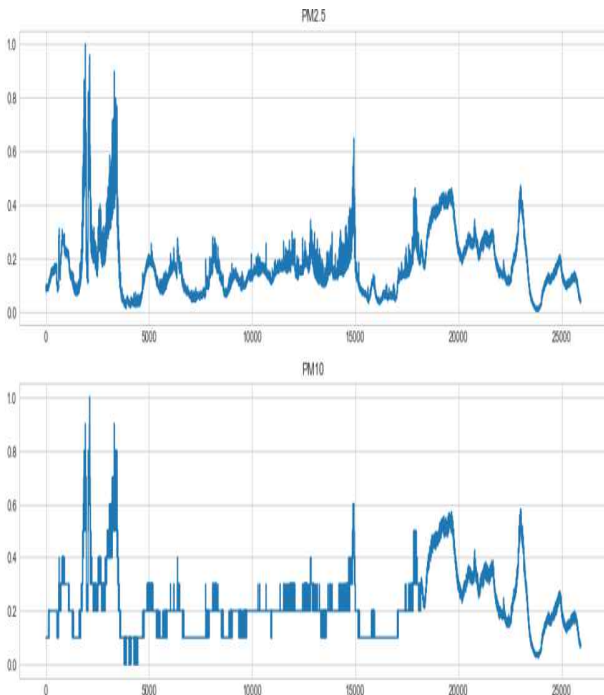


그림 7. PM2.5와 PM10.0 센서 값
Fig. 7. PM2.5 and PM10.0 sensor values

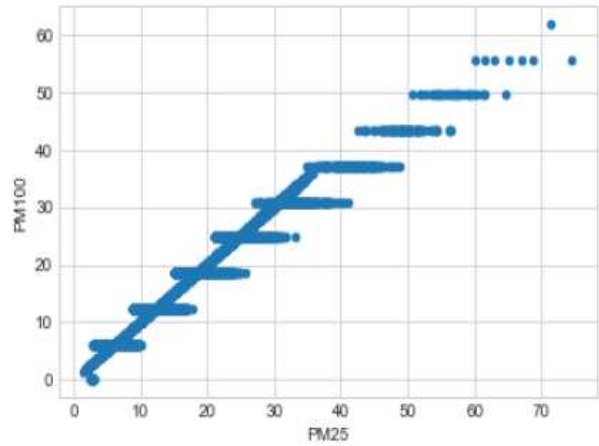


그림 8. PM2.5와 PM10.0 센서 값 플롯팅
Fig. 8. Plotting PM2.5 and PM10.0 sensor values

	S1	S2	S3	S4	S5	S6
S1	1.000000	0.978462	0.126186	0.086916	0.355466	0.106732
S2	0.978462	1.000000	0.121493	0.071399	0.346467	0.091029
S3	0.126186	0.121493	1.000000	0.179200	0.549721	0.369453
S4	0.086916	0.071399	0.179200	1.000000	0.250780	0.978920
S5	0.355466	0.346467	0.549721	0.250780	1.000000	0.349379
S6	0.106732	0.091029	0.369453	0.978920	0.349379	1.000000

그림 9. 센서 값의 연관성 결과
Fig. 9. Correlation result of sensor values

그림 7은 PM2.5와 PM10.0 센서의 측정결과를 보인 것으로 센서 값의 패턴의 유사성을 볼 수 있으며 관찰 값의 구간의 특징을 통해 PM2.5가 정밀한 센서 값을 보이는 것을 관찰할 수 있다. 두 센서 PM2.5와 PM10.0의 센서 값의 패턴 유사성을 통해 해당 센서 값의 추세를 통해 센서 값의 신뢰성을 평가하기 위한 지표로 활용할 수 있다.

그림 8은 PM2.5와 PM10.0의 플롯팅을 보인 것으로 이들 센서 값은 선형적인 특징을 보이며 PM2.5 변화 구역에서 PM10.0이 같은 값을 보인 것을 관찰할 수 있다.

그림 9에서 S1~S6은 각각 PM2.5, PM10.0, 온도, 습도, TVOC, CO2 센서를 나타내며, 5개 센서 값을 기반으로 센서 간의 연관성을 측정된 것이다. S4 (습도)는 S6 (CO2)과 상호 높은 연관성(0.978920)을 갖고 있음을 보이고 있다. 또한 S1 (PM2.5)과 S2 (PM10)는 정밀도의 차이를 갖는 동형의 센서 간의 연관성을 보인 것으로 두 센서의 상호연관성은 0.978462로 높은 상관성을 갖는다.

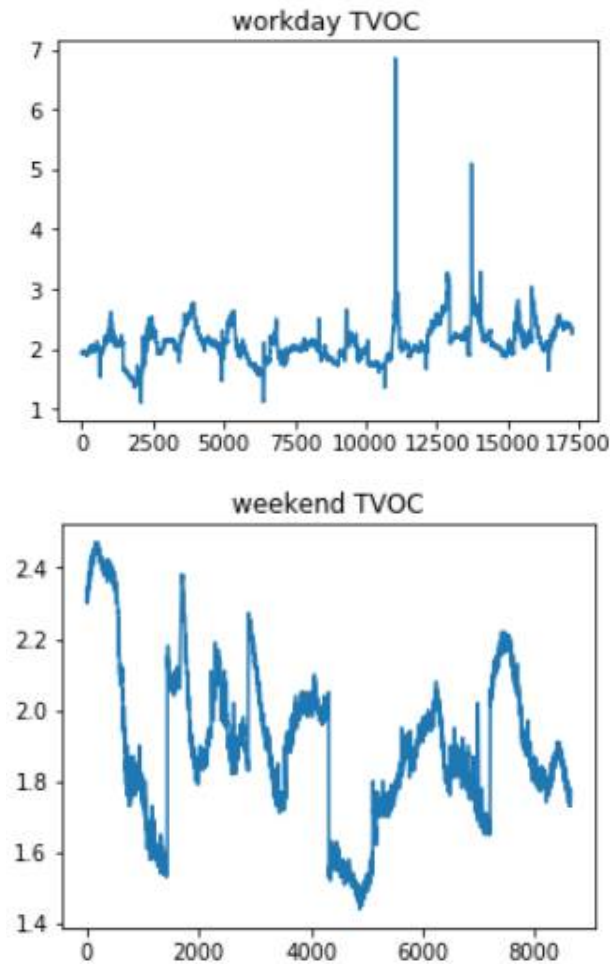


그림 10. 요일별 TVOC 센서 특성 및 플로팅
 Fig. 10. TVOC sensor characteristics and plotting by day of the week

그림 10은 TVOC의 요일별 특성을 보인 것으로 주중과 주말의 센서 값의 차이가 큰 것을 관찰할 수 있었다. 이를 통해 실내 환경에서의 유해물질이 주말에 비해 높은 것을 확인할 수 있다. 실내 TVOC 값의 증가는 휘발성 유기 화합물질의 증가를 의미하는 것으로 실내 유아와 소아의 건강에 미치는 영향이 크다. 관찰된 요일별 특성을 볼 때, 실내에서 사람의 이동에 의한 TVOC 유해물질 증가를 해결하기 위해서는 환기주기를 짧게 하도록 환경관리에 대한 제어가 요구된다.

IV. 결론 및 향후 연구

본 논문에서는 실내 환경 모니터링을 위해 PM2.5, PM10.0, 온도, 습도, TVOC, CO2 센서로부터 측정값을 획득하였으며 센서 특성 기반의 측정값 상관관계를 이해하였다. 이를 위해 EDA를 통해 수집된 센서 값을 시각적으로 표현한 후, 분석에 주어진 데이터의 개별 속성에 서로 의미 있는 상

관관계를 분석하였다. 분석결과로는 온도측정값의 이상치, CO2와 습도 사이의 양의 상관관계, PM2.5와 PM10.0의 센서 값의 패턴 유사성, TVOC의 요일별 특성을 발견하였다.

EDA 분석 결과는 실내 환경에 미치는 상관관계를 활용한 환경관리 제어 시스템 개발에 활용될 수 있으며, 같은 목적의 환경 센서의 센서 값의 특성을 통해 센서 값의 신뢰성을 평가하기 위한 지표로 활용할 수 있을 것이다. 추가적으로 실내 환경의 환경 센서 데이터의 특성을 활용하여 높은 신뢰성을 요구하는 스마트 시티, 스마트 국가 등 대규모 영역의 복잡한 사회 인프라를 최적화하기 위한 기술에 대해 연구할 예정이다.

참고문헌

[1] K. R. Kishore, J. Vishal, B. Suvadeep, B. Lakshmi, "IoT based smart security and home automation system", *2016 International Conference on Computing, Communication and Automation (ICCCA)*, pp.1286-1289, 2016. <https://doi.org/10.1109/CCAA.2016.7813916>

[2] M. Yu, D. Zhang, Y. Cheng, M. Wang, "An RFID electronic tag based automatic vehicle identification system for traffic iot applications", in *Chinese Control and Decision Conference*, Mianyang, pp.4192-1497, 2011. <https://doi.org/10.1109/CCDC.2011.5968962>

[3] D. Lu, T. Liu, "The application of IOT in medical system", in *International Symposium on IT in Medicine and Education*, Cuangzhou, pp.272-275, 2011. <https://doi.org/10.1109/ITiME.2011.6130831>

[4] V. C. Gungor, D. Sahin, T. Kocak, S. Ergut, C. Buccella, C. Cecati, G. P. Hancke, "Smart Grid Technologies: Communication Technologies and Standards", *IEEE Transactions on Industrial Informatics*, Vol.7, Issue.4, pp.529-539, 2011. <https://doi.org/10.1109/TII.2011.2166794>

[5] D. E. Bakken, R. Rameswaran, D. M. Blough, A. A. Franz, T. J. Palmer, "Data obfuscation: anonymity and desensitization of usable data sets", *IEEE Security and Privacy Magazine*, Vol.2, Issue.6, pp.34-41, 2004. <https://doi.org/10.1109/MSP.2004.97>

[6] Luigi Atzori, Antonio Iera, and Giacomo Morabito, "The Internet of Things: A survey," *Computer Networks: The International Journal of Computer and Telecommunications Networking*, vol. 54, no. 15, pp. 2787-2805, 2010. <https://doi.org/10.1016/j.comnet.2010.05.010>

[7] J. Gubbi, R. Buyya, S. Marusic, M. Palaniswami, "Internet of Things (IoT): A vision, architectural elements, and future directions", *Elsevier Future Generation Computer Systems*, Vol.29, Issue.7, pp.1645-1660, 2013. <https://doi.org/10.1016/j.future.2013.01.010>

- [8] Ali, H., Soe, J. K., and Weller, S. R., "A real-time ambient air quality monitoring wireless sensor network for schools in smart cities," in *Proceedings of the 2015 IEEE First International Smart Cities Conference (ISC2)*.
<https://doi.org/10.1109/ISC2.2015.7366163>
- [9] Tukey, John Wilder, *Exploratory Data Analysis*. Addison Wesley, Reading MA
- [10] Andrienko, N. & Andrienko, G., *Exploratory Analysis of Spatial and Temporal Data: A Systematic Approach*, Springer, 2016.
- [11] Daniel Y. Chen, *Pandas for Everyone : Python Data Analysis*, Addison-Wesley, 2018.
- [12] Gravetter, F. J. & Wallnau, L. B., *Statistics for the behavioral sciences*, 9th ed. Belmont, CA: Wadsworth, Cengage Learning, 2013.
- [13] j. s. Kim, *R programming for big data analysis*, 2nd ed. Kame, 2020.
- [14] jmp Statistical Discovery, Relation [Internet]. Available: <http://jmp.com/>.



홍명우(Myoung-Woo Hong)

1995년 : 충남대학교 대학원 (공학석사)
 1997년 : 고려대학교 대학원 (박사수료)
 2010년 : 안동대학교 대학원 (공학박사-멀티미디어공학)

1993년~1996년: 시스템공학연구소
 1999년~현 재: 우송정보대학 스마트팩토리기술과 교수
 ※관심분야 : 유비쿼터스 컴퓨팅, 사물인터넷, 영상처리

강윤희(Yun-Hee Kang)



1993년 : 동국대학교 대학원 (공학석사)
 2002년 : 고려대학교 대학원 (이학박사-컴퓨터과학)

1997년~2000년: (주) 오름정보 개발부 과장
 2010년~2011년: 인디애나 대학교 PTI 연구소 방문연구원
 2000년~현 재: 백석대학교 컴퓨터공학부 교수
 ※관심분야 : 분산시스템, 인공지능, 클라우드컴퓨팅