

환경 모니터링을 위한 EDA 기반 데이터 분석

강윤희*, 조재혁**

*백석대학교 ICT학부

**숭실대학교 전자정보공학부

yhkang@bu.ac.kr, chojh@ssu.ac.kr

EDA based Data Analysis for Environmental Monitoring

Yunhee Kang*, JaeHyuk Cho**

*Div. of ICT, Baekseok Univsersity

**Dept. of Electronic Eng., Soongsil Univsersity

요약

최근 센서데이터 활용 영역이 넓어지면서 데이터 분석 서비스가 활성화되고, 분석을 용이하게 할 수 있는 환경으로 진화하고 있다. 이에 따라 센서데이터의 신뢰성 보장이 필요하다. 신뢰성을 갖는 환경모니터링을 위해서는 센서로부터 수집된 환경 데이터의 분포 및 값을 살펴본 후 데이터가 표현하는 현상을 더 잘 이해하고, 센서 및 센서데이터에 대한 잠재적인 특이점을 발견을 실행하여야 한다. 이를 위해 EDA를 통해 수집된 센서 값을 시각화하고 분석에 주어진 데이터의 개별 속성의 특징 및 상관관계를 도출한다. 본 연구의 EDA 분석 결과는 센서데이터의 신뢰성을 평가하기 위해 사용한다.

1. 서론

최근 센서 데이터 활용 영역이 넓어지면서 데이터 분석 서비스가 활성화되고, 분석을 용이하게 할 수 있는 환경으로 진화하고 있다[1]. 아마존 AWS IoT은 데이터 수집을 지원하는 플랫폼으로 데이터 분석 활용 영역이 넓어지면서 빅데이터 분석 서비스가 활성화되고, 분석이 용이해지고 한다.

탐색적 데이터 분석(EDA, Exploratory Data Analysis)은 데이터의 구조적 특성을 알아내기 위한 기법이다[2,3]. EDA는 수집한 데이터에 대한 다양한 각도에서 관찰하고 이해하는 과정으로 상세한 데이터를 분석하기 전에 그래프나 통계적인 방법으로 자료를 직관적으로 바라보는 과정이다[4].

기존 센서평가는 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 환경모니터링을 위해서는 센서로부터 수집된 환경 데이터의 분포 및 값을 살펴본 후 데이터가 표현하는 현상을 더 잘 이해하고, 센서 및 센서데이터에 대한 잠재적인 특이점을 발견을 실행

하여야 한다.

이 논문에서는 신뢰성 평가에 필요한 센서특성을 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 동일공간 및 동일시간대의 다른 센서에 대한 출력을 비교하여 비교군 센서의 특성을 관찰함으로써 EDA 분석 결과는 센서데이터의 신뢰성을 평가하기 위해 사용한다.

2. 실험결과 및 분석

본 실험은 실내 공기질 모니터링을 위해 2020년 4월 2일부터 2020년 4월 19일까지 PM2.5, PM10.0, 온도, 습도, TVOC, CO2 센서로부터 전달된 센서데이터(sensory data)를 기반으로 한다. 미세먼지 측정을 위해 PM2.5와 PM10.0을 사용하며, 공기품질 측정을 위해 TVOC, CO2를 사용한다.

EDA를 수행한 후 센서간의 상관관계를 도출한다. 기초적인 데이터 분석은 파이썬의 Pandas 라이브러리를 사용하여 수행한다. 표 1은 실험에 사용한 센서 데이터 중 습도와 CO2의 통계정보 (descriptive

statistics)를 보인 것이다. 225919의 전체 관측치를 가지며, 1초에 6개의 센서 데이터를 획득하여 수집서버에 전달한다.

<표 1> 데이터셋의 통계 값

	습도	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

<표 2>와 <표 3>은 5개의 최저 및 최고 습도 측정값에 따른 다른 5 센서값과의 측정값을 기술한 것으로 습도값은 CO2 과 연관성이 있음을 관찰할 수 있다.

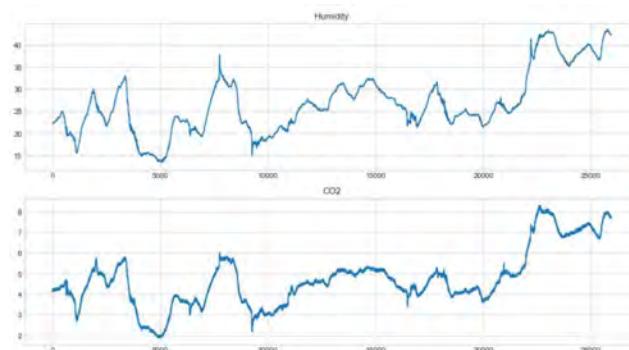
<표 2> 최저습도값의 타 센서 측정값

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

<표 3> 최고습도값의 타 센서 측정값

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 1>은 습도와 CO2의 측정값을 시각화하여 연관성을 플롯팅한 것을 보인 것으로 시각화를 통해 두 센서의 변화추이를 관찰할 수 있다.



<그림 1> 습도(Humidity) 와 CO2 관측값 플롯팅

<표 4>은 5개 센서값을 기반으로 센서간의 연관성

을 측정하였다. S4(습도)는 S6(CO2)와 상호 높은 연관성(0.978920)을 갖고 있음을 보인다.

<표 4> 센서값의 연관성(correlation) 결과

	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

3. 결론 및 향후 연구

이 논문에서는 실내 환경 모니터링을 위해 PM2.5, PM10.0, 온도, 습도, TVOC, CO2 센서로부터 측정값을 획득하였으며 센서특성을 이해하였다. 이를 위해 EDA를 통해 수집된 센서 값을 시작적으로 표현한 후 분석에 주어진 데이터의 개별 속성에 서로 의미 있는 상관관계를 도출한다. EDA 분석 결과는 센서데이터의 신뢰성을 평가하기 위해 사용한 후 스마트 시티, 스마트 국가 등 대규모 영역의 복잡한 사회 인프라를 최적화하기 위한 기술로서 활용예정이다.

감사의 글

본 논문은 과학기술정보통신부(정보통신기획평가원, 2019-0-00136) ICT 혁신선도 인프라 구축 사업의 ICT 기반 환경 모니터링 센서 검증 플랫폼 과제의 지원을 받아 수행된 연구임

참고문현

- [1] 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)
- [2] Tukey J.W. “Exploratory Data Analysis. Addison Wesley, Reading” MA
- [3] Andrienko, N & Andrienko, G “Exploratory Analysis of Spatial and Temporal Data. A Systematic Approach” Springer.
- [4] Chen, Daniel Y. “Pandas for Everyone : Python Data Analysis” Addison-Wesley