

## 위치 기반의 UCI Sensor 시계열 데이터 분석

\*장일식 \*\*박구만

\*서울과학기술대학교 나노IT디자인융합대학원 정보통신미디어공학전공

\*\*서울과학기술대학교 전자미디어IT공학과

\*foreverme@naver.com

## Location-based UCI Sensor time series data analysis

\*Chang, Il-Sik \*\*Park, Goo-man

\*Dept. of Information Technology and Media Engineering

\*\*Dept. of Electronics and IT Media Engineering

Seoul National University of Science and Technology

### 요약

인공지능 기술과 서비스는 딥러닝을 중심으로 한 기계학습 기술의 급속한 발전에서 원인을 둔다. 딥러닝 발전 요인으로 GPU 등 하드웨어 발전, 기술 공유, 대규모 학습데이터 구축 및 공개를 들 수 있다. 데이터 셋에 관련하여 센서를 이용한 데이터셋의 경우 단순히 많은 데이터셋의 확보뿐 아니라 적절한 위치 및 환경에 따른 고려가 필요하다. 본 논문에서는 UCI의 화학 가스의 데이터셋을 이용하여 위치별 시계열 데이터를 딥러닝을 이용하여 분석하고, 위치별 정확도와 손실을 계산한다. 또한 계산된 결과를 히트맵을 통하여 시각화하여 직관적인 이해를 높인다. 또한 위치별 정확도가 높은 상위 5개의 위치에서 앙상블 방법을 통한 성능의 향상을 확인 하였다.

### 1. 서론

데이터 시각화(data visualization)는 데이터 분석 결과를 쉽게 이해할 수 있도록 시각적으로 표현하고 전달되는 과정을 말한다. 데이터 시각화의 목적은 도표라는 수단을 통해 정보를 명확하고 효과적으로 전달하는 것이다. UCI의 화학 가스의 데이터셋[1]에 대한 시각화를 위한 분석은 Inhibitory Support Vector Machines(ISVM)을 통한 가스 센서데이터 분석[2], 딥러닝 학습 후 마지막 레이어를 시각화하는 방법[3] 등이 있다. 본 논문에서는 위치별 시계열 가스데이터를 딥러닝을 통해 학습하고, 학습된 모델을 통한 정확도 및 손실을 히트맵을 통해 직관적으로 이해가 가능하도록 시각적으로 보여준다. 또한 위치별 정확도가 높은 상위 몇 개의 위치에서 앙상블 방법을 이용하여 성능향상을 확인 하였다. 본 논문의 구성은 다음과 같다. 2장에서는 UCI 센서 데이터셋에 관한 설명을 기술하고, 3장은 위치별 UCI 센서 데이터 분석한다. 마지막으로 4장에서 결론을 맺는다.

### 2. UCI 센서 데이터셋

UCI Sensor의 종류는 MOX 센서로 총 8개의 어레이로 구성된다.[1] 각 센서의 반응은 화학 물질에 따라 다르게 나타낸다. 그림1[2]은 센서 데이터 수집을 보여준다.

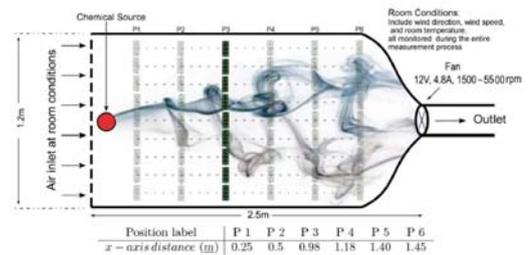


그림 1. 센서 데이터 수집

그림 1에서 가로는 P1~P6으로 총 6개의 위치에 세로로 총 9개의 MOX 센서가 위치 한다. 위치별로 보면 9X6의 2차원 위치로 볼 수 있다. 각각의 MOX 센서는 8개의 센서값을 포함하기 때문에 위치별로 8차원의 데이터가 된다. 데이터 셋에서는 fan 속도를 1500rpm, 3900rpm, 5500rpm 으로 변경하여 값을 측정하였고, 화학 센서의 내장 발열체에 적용되는 전압 값은 4V, 4.5V, 5V, 5.5V, 6V로 구분하였다. 본 논문에서는 fan 속도 및 화학 센서의 내장 발열체의 전압 값은 고려하지 않고, 단순히 위치별 데이터를 고려하였다. 그림 2[2]는 센서 데이터 방출시 MOX 가스 센서의 값을 나타 낸다. UCI Dataset은 10~20msec 단위로 0 ~ 260 Sec 동안 저장되어 있다. 본 논문에서는 0~70초 사이의 값을 사용한다. 실제 샘플링 주기는 500msec로 총 140x8의 시계열 데이터가 되지만 실제 사용값은 137x8을 사용하였다.

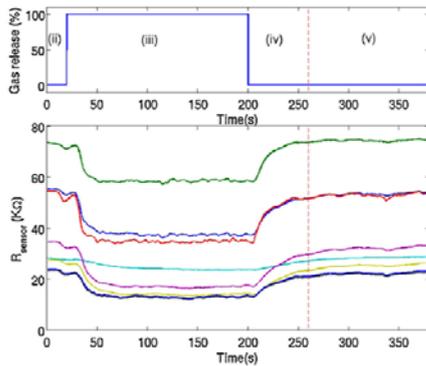


그림 2. MOX 가스 센서 값

데이터 셋에서는 총 11개의 클래스가 있으나 본 논문에서는 총 9개의 클래스로 분류한다. 각각의 클래스 이름은 "Acetaldehyde\_500", "Acetone\_2500", "Ammonia\_10000", "Benzene\_200", "CO\_4000", "Ethylene\_500", "Methane\_1000", "Methanol\_200", "Toluene\_200" 이다.

### 3. 위치별 UCI 센서 데이터 분석

위치별 시계열 데이터는 그림 3과 같이 딥러닝 모델을 구성하였다.

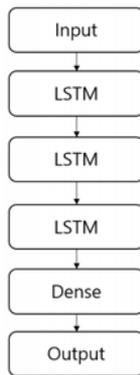


그림 3. 시계열 데이터 분석 모델

그림3에서 LSTM은 dropout은 0.1, recurrent\_dropout 0.1로 설정하였고, 처음 2개의 LSTM의 출력 차원은 64, 3번째 LSTM의 출력 차원은 32로 설정하였다. 옵티마이저는 adam을 사용하였다. 그림 4는 위치별 학습을 통한 정확도 및 손실을 히트맵을 통하여 시각적으로 보여 준다. 그림4의 히트맵은 총 5번의 테스트 결과에 대한 평균값을 시각화한 것이다.

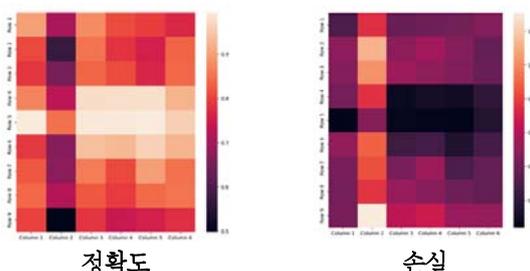


그림 4. 위치별 정확도 및 손실

본 논문에서는 위치별 정확도를 기준으로 가장 높은값을 나타내는 위치 5곳과 가장 낮은 위치 5곳을 선택하여 앙상블 방법을 사용하여 분석하였다. 즉 5개의 모델에 대한 추론 결과가 정답 클래스가 3개이상인 경우 정상인식으로 판별하였다. 표 1은 위치별 앙상블 방법을 적용 하였을 경우의 정확도를 나타낸다.

표 1. Sensor 종류

	높은 정확도	낮은 정확도
앙상블	1.0	0.71
위치 1	0.995	0.491
위치 2	0.990	0.609
위치 3	0.980	0.645
위치 4	0.985	0.627
위치 5	0.980	0.725

표1은 가스 센서데이터의 수집의 위치가 중요함을 보여준다. 각위치별 정확도 및 앙상블 방법을 사용하였을 경우 정확성의 효과가 좋아짐을 볼수 있었다. 또한 앙상블의 개수를 조절해 가면서 정확도를 확인함으로써 센서데이터의 수집위치 및 위치 갯수를 효과적으로 결정 할 수 있다.

### 3. 결론

본 논문에서는 UCI 센서 데이터셋을 이용하여 시계열 센서 데이터를 학습하였으며, 위치별 데이터를 학습하여, 위치별 정확도 및 손실을 계산하여 히트맵을 통하여 시각적으로 표시하였다. 센서 데이터의 수집을 위해선 위치에 대한 선택 및 몇 개의 위치에서 센서 데이터를 수집할 것인가에 대한 선택이 어렵다. 본 논문에서는 히트맵을 통하여 직관적인 이해와 여러 위치에서의 선택에 대한 앙상블 방법을 적용함으로써 성능이 향상됨을 볼 수 있었다. 향후 연구로는 설명가능한 인공지능 기법을 적용하여 위치별 중요도에 대한 연구를 진행할 예정이다.

### Acknowledgement

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2019-0-01106, Sub-ppb급 가스성분 감지를 위한 후각지능 기술 개발)

### 참고문헌

[1] Gas sensor arrays in open sampling settings Data Set <http://archive.ics.uci.edu/ml/datasets/Gas+sensor+arrays+in+open+sampling+settings?ref=datanews.io>

[2] Alexander Vergara, Jordi Fonollosa, Jonas Mahiques, Marco Trincavelli, Nikolai Rulkov, Ramon Huerta, On the performance of gas sensor arrays in open sampling systems using Inhibitory Support Vector Machines, Sensors and Actuators B 185. 2013

[3] 장일식, 최희조, 박구만. 데이터 시각화 기반의 UCI Sensor Data 분석. 한국방송미디어공학회 학술발표대회논문집, 21-24. 2020