

유해화학물질의 분광 스펙트럼 파형 그래프를 통한 AI 학습 및 분류 방법

유성민, 조숙경*, 백성하**, 김경배*

(주)우경정보기술, *서원대학교, **인하공업전문대학,

smin9830@gmail.com, *skyindb@naver.com **bshzeratul@gmail.com, *gbkim@seowon.ac.kr

AI Learning and Classification using Spectral Waveform Graph of Harzardous Chemicals

Ryoo Seong Min, Cho Sook Kyoung*, Baek Seong Ha**, Kim Gyoung Bae*

WooKyoung IT, *Seowon Univ., **Inha Technical College

요약

본 논문에서는 시각적 구별이 어려운 유해화학물질의 종류를 파악하기 위해 분광 스펙트럼 데이터를 이용하여 AI융합 유해화학물질 관독시스템을 학습 시킨다. 해당 데이터는 분광 카메라로 촬영된 데이터 파일에서 학습에 적합한 형태로 군집화를 통해서 추출한 것으로 해당 데이터를 분광 관독 모델에 학습시킨다. 해당 모델의 정확도 판별을 위해 7종류 물질의 데이터를 통해서 해당 모델의 물질 분류 시험을 시행하며, 해당 시험을 통해서 모델이 학습된 물질에 대해서 어느 정도의 정확도를 가지고, 어떤 문제점이 존재하는지 확인한다.

I. 서론

유해화학물질 관련 사고의 증가에 따라서 사고 대응을 위한 물질 종류의 신속한 판독이 중요해지고 있다[1,2]. 유해화학물질 사고는 위험성이 크기에 직접 접근하는 것이 매우 어렵고, 원거리에서 영상만으로 판별하는 것도 시각적 특징이 없는 경우에는 어려움이 따른다[2, 3, 5]. 본 논문에서는 시각적 특징이 없는 물질의 경우에도 판독이 가능한 AI 판독시스템의 학습 데이터로써 분광 스펙트럼 데이터를 선택했다[4, 6, 7].

분광 카메라로 유해화학물질을 촬영하게 된다면 물질의 구성성분, 구조 등에 따라서 서로 다른 스펙트럼 그래프 파형을 출력하게 되고, 해당 그래프는 화학물질의 종류를 구별할 수 있는 지표로써 사용될 수 있다[6, 7]. 해당 데이터는 카메라를 통해서 데이터 파일의 형식으로 저장되고, 기본적으로는 한 장의 이미지 형식이며, 해당 이미지의 모든 픽셀마다 해당 지점의 분광 스펙트럼 그래프가 저장되어 있는 형식이다[7]. 해당 스펙트럼 그래프들은 모두 상대적인 값을 가지고 있긴 하지만 같은 물질에 대해서는 동일한 파형의 그래프를 가지게 된다[7]. 데이터 파일 형식으로 이미지가 저장되어 있기 때문에 기본 프로그램으로는 내부의 데이터를 확인 및 추출할 수가 없고, 분광 스펙트럼 데이터를 추출하기 위해서는 binary 형식으로 저장된 데이터를 해석해서 해당 파일의 데이터들을 추출해야 한다[7]. 데이터의 특성상 스펙트럼 파형 그 자체의 데이터가 필요하기에 이미지 파일과 동일한 방식으로 라벨링을 하는 것에 어려움이 있어서 해당 데이터 파일에서 AI 데이터 학습에 적합한 데이터 형식으로 스펙트럼 그래프를 추출할 필요가 있다[7].

프로그램을 통해서 물질이 존재하는 영역의 분광 스펙트럼 데이터를 픽셀별 각각 하나의 AI 학습용 텍스트 파일로 추출 해냈다[7]. 본 논문에서는 해당 데이터들을 이용하여 AI 모델을 학습시키고, 해당 모델이 실제로 분광 스펙트럼 데이터를 입력받았을 때 얼마나 정확하게 데이터를 인식하고, 분류하는지 결과를 확인한다. 또한, 잘못 분류하는 경우를 분석하며, 해당 방법을 통해서 유해화학물질 관독용 AI의 정확성을 높이는 것이 가능한지 확인하려 한다.

II. 본론

분광 카메라 데이터에서 추출한 데이터들은 AI 학습을 위해서 전부 텍스트 파일 형식으로 추출된다. 해당 파일의 제목은 원본이 되는 데이터 파일의 파일명과 해당 데이터 파일의 어느 부분에서 추출되었는지 표시하는 samples와 lines의 값으로 이루어져 있다. 파일의 내부에는 spectrum, wavelengths 쌍으로 이루어져 있으며 파일 하나당 228쌍의 데이터가 포함되어 있다. 해당 데이터는 그림 1과 같이 스펙트럼 그래프의 특징을 나타내는 극대 및 극소점의 수, 극대 및 극소점의 wavelength, 평균, 최대, 최소 신호 값의 강도 등의 특징을 통해서 비슷한 특징을 지닌 그래프끼리 군집화를 해서 추출한 데이터이다.

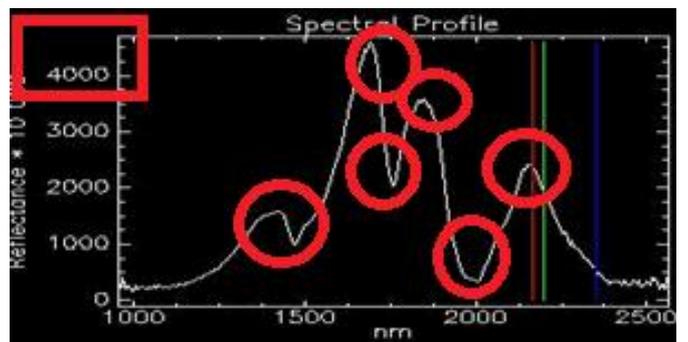


그림 1 스펙트럼 그래프 추출 특징

분광 스펙트럼 데이터에서 3개의 영역으로 군집화를 했기 때문에 실질적으로 물질이 존재하지 않는 곳의 극소량의 데이터가 함께 물질의 데이터로 분류된 경우가 존재하고, 물질을 담은 유리병의 특성상 물질이 담긴 병 내부에도 빛의 반사로 인해 제대로 측정되지 않은 데이터와 다른 영역과의 경계면으로 인해서 오염된 데이터가 포함되어 있다. 그렇지만 그러한 경우는 극소수이고, 물질로 분류된 영역에서 추출된 데이터는 대부분이 데이터를 통해서 AI 모델을 학습시키고, 해당 모델을 통해서 7개의 물질 데이터를 각 100개씩 추출해 물질 분류실험을 거쳐서 성능을 시

험했다. 사용한 물질 데이터는 황산, 암모니아, 등유, 시너, 플루오르화칼륨 수용액, 톨루엔, 메탄올의 7종류이다. 검증에 사용한 물질 데이터의 경우는 학습에 사용된 데이터들과 동일한 데이터 파일에서 추출하였으나 해당 데이터들은 학습에 사용된 데이터들과 중복 되지 않는 텍스트 데이터로 선정했다.

위의 물질들은 모두 눈으로는 판별하기 힘든 무색을 띄고 있는 물질들이며, 물질의 냄새, 불꽃색, 연기 등의 다른 특징을 띄는 물질들이 몇 가지 섞여 있기는 하지만 시각적인 특징들이 매우 적기 때문에 본 논문에서는 해당 물질들의 데이터를 통해서 학습을 시키고, 분류 실험을 거치는 것을 통해 시각적 특징이 없는 유해화학물질에 대해서 AI 판독모델이 얼마나 정확히 분류할 수 있는지 시험했다.

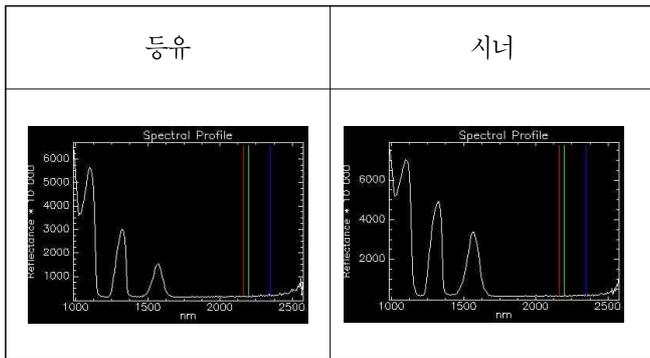


그림 2 등유와 시너의 스펙트럼 그래프 비교

7가지 물질 중에서 등유와 시너의 경우는 분광 스펙트럼 그래프를 그려보면 그림 2와 같이 스펙트럼 파형이 거의 일치하는 모습을 볼 수 있다. 스펙트럼 측정에 사용한 시너가 등유와 성분이 거의 똑같아서 두 스펙트럼 그래프가 거의 동일하게 그려진 것이다. 해당 두 물질은 학습 후 분류 실험에서 낮은 수치의 정확도를 기록할 것으로 예상 되었다.

경남:/home/vai/Workspace/HCS/hc_AI_Data_forLearn/video0sec2/hc13.등유_swir/ONU_10_2023-09-26_05-11-28_refl_Line_139_Sample_209.txt
 물질명: [hc13.등유], 예측물질명: [hc13.등유]
 경남:/home/vai/Workspace/HCS/hc_AI_Data_forLearn/video0sec2/hc13.등유_swir/ONU_10_2023-09-26_05-11-28_refl_Line_140_Sample_199.txt
 물질명: [hc13.등유], 예측물질명: [hc13.등유]
 경남:/home/vai/Workspace/HCS/hc_AI_Data_forLearn/video0sec2/hc13.등유_swir/ONU_10_2023-09-26_05-11-28_refl_Line_145_Sample_111.txt
 물질명: [hc13.등유], 예측물질명: [hc13.등유]
 경남:/home/vai/Workspace/HCS/hc_AI_Data_forLearn/video0sec2/hc13.등유_swir/ONU_10_2023-09-26_05-11-28_refl_Line_145_Sample_211.txt
 물질명: [hc13.등유], 예측물질명: [hc13.등유]
 경남:/home/vai/Workspace/HCS/hc_AI_Data_forLearn/video0sec2/hc13.등유_swir/ONU_10_2023-09-26_05-11-28_refl_Line_149_Sample_209.txt
 물질명: [hc13.등유], 예측물질명: [hc13.등유]
 경남:/home/vai/Workspace/HCS/hc_AI_Data_forLearn/video0sec2/hc13.등유_swir/ONU_10_2023-09-26_05-11-28_refl_Line_149_Sample_232.txt
 물질명: [hc13.등유], 예측물질명: [hc13.등유]
 경남:/home/vai/Workspace/HCS/hc_AI_Data_forLearn/video0sec2/hc13.등유_swir/ONU_10_2023-09-26_05-11-28_refl_Line_151_Sample_256.txt
 물질명: [hc13.등유], 예측물질명: [hc13.등유]
 경남:/home/vai/Workspace/HCS/hc_AI_Data_forLearn/video0sec2/hc13.등유_swir/ONU_10_2023-09-26_05-11-28_refl_Line_152_Sample_162.txt
 물질명: [hc13.등유], 예측물질명: [hc13.등유]
 경남:/home/vai/Workspace/HCS/hc_AI_Data_forLearn/video0sec2/hc13.등유_swir/ONU_10_2023-09-26_05-11-28_refl_Line_157_Sample_117.txt
 물질명: [hc13.등유], 예측물질명: [hc13.등유]

그림 3 분류 결과

그림 3의 경우는 학습시킨 AI 모델에 위에서 언급한 물질 데이터 700개를 분류 시험을 한 결과이다. 해당 모델이 대상 데이터들에 대해서 어떤 물질로 예측을 했고, 실제 물질명은 무엇이며, 그 값이 일치하는지에 대해 결과창으로 표현하며, 모델이 제대로 분류하지 못했을 경우 해당 모델이 한 물질을 어떤 다른 물질로 예측했는지 확인할 수 있다. 그림 3에서도 실제 물질과 예측물질이 일치하지 않는 경우가 하나 보이는데 해당 경우는 오답으로 표기되어 있다.

그림 4의 모델의 결과를 보면 f1-score가 낮게 나온 두 물질의 경우는 앞서 말했듯이 구성성분이 거의 비슷하기 때문에 분류할 때 등유를 시너로, 시너를 등유로 분류하는 경우가 점수 감소의 주된 원인이었으며, 그 외의 다른 물질들은 전부 0.7 이상을 기록했다. 전체적으로 학습 데이터로 학습 된 모델이 물질을 잘 분류하고 있음을 확인할 수 있다.

| 물질명 | precision | recall | f1-score |
|---------------|-----------|--------|----------|
| hc02_황산 | 0.72 | 0.93 | 0.81 |
| hc12_암모니아 | 0.87 | 0.68 | 0.76 |
| hc13_등유 | 0.76 | 0.37 | 0.5 |
| hc14_시너 | 0.53 | 0.59 | 0.56 |
| hc21_플루오라이드칼륨 | 0.65 | 0.88 | 0.75 |
| hc24_톨루엔 | 0.75 | 0.93 | 0.83 |
| hc26_메탄올 | 0.89 | 0.64 | 0.74 |

그림 4 모델 결과

분광 스펙트럼 데이터 그래프 파형의 차이를 통해 학습시킨 AI 모델을 통해서 스펙트럼 그래프 데이터로 학습시킨 모델을 이용하면 AI가 분광 스펙트럼 데이터를 통해서 해당 물질이 어떤 물질에 해당 하는지 높은 정확도로 분류하는 것이 가능하다.

III. 결론

본 논문에서는 분광 카메라를 통해 얻은 AI 학습용 데이터를 통해서 AI 모델을 학습시키고, 해당 모델이 유해화학물질의 종류를 얼마나 잘 판독 하고, 분류하는지에 대해서 확인했다. 분류 시험은 7개의 시각적 특징이 없는 물질에 대해서 진행되었고, 구성성분이 거의 같은 물질 2종류를 포함 하고 있다. 해당 시험으로 분광 스펙트럼 파형이 다를 경우는 높은 정확도로 해당 물질들을 분류하는 것이 가능하고, 기존에 학습시킨 그대로 해당 물질의 종류를 판독하는 것을 확인할 수 있었다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부와 정보통신산업진흥원 주관으로 소방청 컨소시엄에서 수행하는 “AI융합 유해화학물질 판독시스템 사업”(2022~2024)의 지원을 받았음

참고 문헌

- [1] 화학물질안전원 (<https://nics.me.go.kr/>)
- [2] 김연진, 박봉섭, 김경배, “인공지능기술 기반의 유해화학물질 사고 대응에 관한 연구,” 한국통신학회 하계종합학술발표회, pp.359-360, June, 2022.
- [3] 조숙경, 백성하, 박봉섭, 김경배, “인공지능기술을 활용한 유해화학물질 탐지 방법,” 한국통신학회 추계종합학술발표회, pp.264-265, Nov., 2022.
- [4] 한국전자통신연구원, “초분광 이미징 기술동향”, 2019 Electronics and Telecommunications Trends, pp86-97
- [5] 조숙경, 김희훈, 백성하, 박봉섭, 김경배 “인공지능 시각화 기법을 이용한 유해화학물질 판독 프로토타입 시스템”, 한국통신학회 하계종합학술발표회, pp101-102, June, 2023.
- [6] 국가연구시설장비진흥센터, PRISM MANUAL 제 12호 푸리에변환적외선분광기(FT-IR)의 이해, 2022
- [7] 유성민, 윤아롱, 심보리, 김연진, 백성규, 조숙경, 백성하, 박봉섭, 김경배 ‘AI 학습을 위한 분광 스펙트럼 라벨링 데이터 추출 방법’, 한국통신학회 추계종합학술발표회