

멀티 센서 기반의 스마트홈 공기질 향상을 위한 모니터링 시스템

안다운, 윤나은, 진나영, 이상금

국립한밭대학교

{ahndaeun122, better5026, jinnayoung03}@gmail.com, sangkeum@hanbat.ac.kr

Monitoring System for Improving Smart Home Air Quality Based on Multi-Sensor

Daeun Ahn, Naeun Yoon, Nayoung Jin, Sangkeum Lee

Hanbat National University

요약

본 논문은 코로나 19 팬데믹으로 인해 스마트홈의 실내에서의 활동이 증가하여 더욱이 중요시되고 있는 실내 공기질에 영향을 미치는 변수들의 상관관계를 센서에서 취득된 빅데이터를 이용해 분석한다. 스마트홈의 CO₂, TVOC, PM₁₀, PM_{2.5} 등의 공기질과 직접적으로 연관이 있는 데이터와 그 외 온도, 습도, 강수량 등 간접적으로 영향을 미치는 데이터의 상관관계를 히트맵(Heatmap)으로 시각화하여 찾아낸다. 공기질을 직관적으로 나타내기 위하여 개별로 존재하는 지표를 룰베이스 기반으로 CO₂, TVOC, PM₁₀, PM_{2.5} 4 가지 데이터가 포함된 복합적인 종합 지표로 구축한다. 비지도 학습 K-means 알고리즘을 통해 실내 공기질을 CO₂, TVOC, PM₁₀, PM_{2.5}의 양에 따라 3 가지 기준으로 분류하였고 최신 데이터에 맞는 새로운 종합 지표를 정립한다. 이를 바탕으로, 공기질 상태를 판단하는 평가 지표를 개발하고 지표를 정규화 및 시각화하여 실내 공기질 분류 모델을 구축한다. 이후 이 모델을 이용하여 온도, 습도, 강수량 데이터만으로 예측하고자 하는 새로운 공기질을 판단할 수 있다.

I. 서론

코로나 19로 실외 활동이 제한되어 실내에서 보내는 시간이 늘어나며 실내공기질의 영향도 같이 증가하였다[1]. 스마트홈 기술은 공기질 모니터링 및 관리에 새로운 가능성을 제공하고 있다. 특히 멀티 센서를 활용한 스마트홈 시스템은 다양한 환경 변수를 실시간으로 감지하여 사용자에게 효과적인 모니터링 및 향상된 제어 기능을 제공한다. 본 연구에서는 멀티 센서를 기반으로 한 스마트홈 공기질 모니터링 시스템을 개발하고, 이를 통해 공기질을 향상시키는 데 기여하는 새로운 방법을 제안한다.

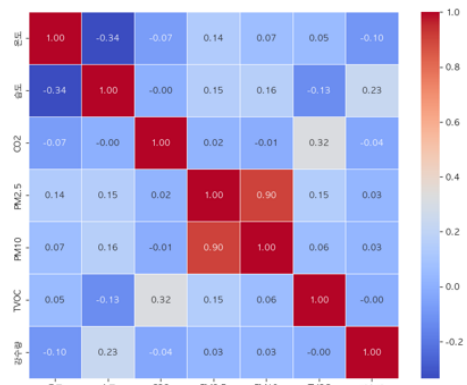
이에 따라 실내공기질 수준을 파악하고 개선하기 위한 실내공기질 지표들이 그림 1과 같이 개발되고 있다[2]. 공기 오염물질의 종류 및 분포가 다양하여 개별 오염물질의 유지기준만으로는 공기질 수준을 전체적으로 평가하기 어렵기에 종합 지표 개발이 필요하다. 기존에 Saad et al. (2017)이 AQI를 응용하여 실내공기질 평가지표(IAQI, Indoor Air Quality Index)와 온열 환경 평가지표(TCI, Thermal Comfort Index)를 도출하고 두 지수를 결합하여 EIAQI(Environment Indoor Air Quality Index) 종합지표를 개발하였다[3]. 그러나 이 지표에서 필요한 CO₂, TVOC, PM_{2.5}, PM₁₀ 등의 데이터는 가정에서 측정하고 조절하기 어렵다.

제안하는 모니터링 시스템은 새로 정립한 종합적인 공기질 판단 지표를 이용하여 사용자의 접근이 쉬운 환경 변수 온도, 습도, 강수량 등의 데이터만으로 종합적인 공기질 상태를 실시간으로 판단한다. 이 정보를 기반으로 사용자가 환기, 온도 조절, 가습기 등의 의사 결정을 스마트하게 수행할 수 있다.

	좋음	보통	나쁨	매우 나쁨	센서 이상
TVOC	0~500	501~1000	1001~2500	2501~3999	4000~
CO ₂	0~500	501~1000	1001~1500	1501~3999	4000~
PM _{2.5}	0~15	16~35	36~75	76~199	200~

-1이면 센서 이상

그림 1. 국립환경과학원 대기질 기준



	온도	TVOC	PM10
습도	반비례	반비례	-
PM2.5	-	비례	비례

그림 2. 공기질 상관관계 Heatmap 시각화

II. 상관관계 분석

데이터를 전처리하여 스마트홈의 공기질 및 환경 변수 데이터를 준비하였다. 미세먼지 및 다양한 센서 데이터를

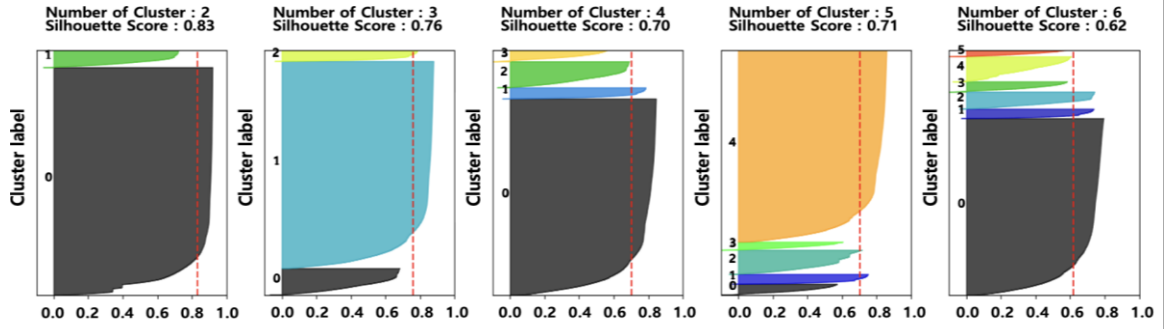


그림 3. 실루엣 스코어 시각화

포함한 모니터링 시스템의 핵심 데이터를 획득하였다. 이때, 시하스 공기 측정기 센서 범위가 아닌 값들은 센서 이상으로 판단하고 제거하였다.

2-1. 공기질 상관관계 히트맵 시각화

수집한 빅데이터 중 공기질에 영향을 미칠 수 있는 특성들의 상관관계를 히트맵으로 시각화한 결과를 그림 2에서 확인해 볼 수 있다.

대부분의 동호수에서 온도, 습도, CO2, PM2.5, PM10, TVOC, 강수량 등 특성 데이터들 간의 두드러지는 상관관계를 파악하였고 그림 3 과 같이 나타내어 진다.

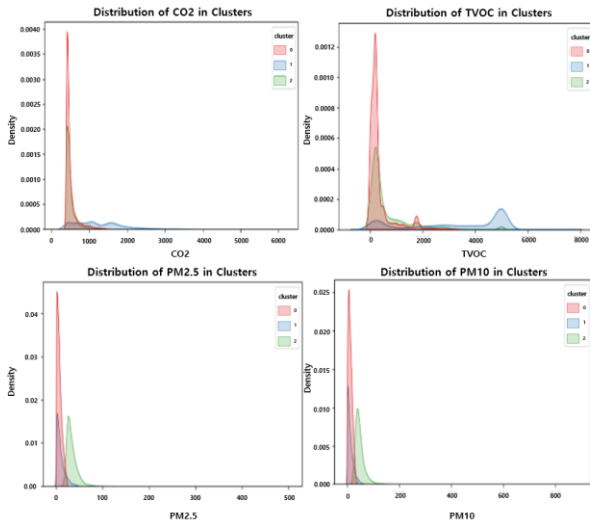


그림 4. K-means 군집별 특성(CO2, TVOC, PM2.5, PM10) 시각화

III. 공기질 종합 지표 정립

3-1. 룰 기반 알고리즘

그림 1 을 바탕으로 TVOC, CO2, PM2.5 등 공기질에 해로운 영향을 미치는 요소들을 종합적으로 판단하는 새로운 지표를 만들었다.

TVOC, CO2, PM2.5 등 각 특성들의 수준에 맞게 점수를 부여(좋음:+3, 보통:+2, 나쁨:+1, 매우 나쁨:+0)하고 전부 합한 score 를 특성으로 추가하였다. TVOC 이 400 ug/m3 이고, CO2 가 600ppm, PM2.5 가 17ug/m3 이면 score 은 7 점이다. score 는 0~9 점이고 score 가 높을수록 좋은 공기질 상태로 판단한다. score 에 따라 전체 데이터 개수가 고르게 분배되도록 새로운 수준 지표 기준을 나누었고, 9 점은 'good', 6~8 점은 'normal',

0~5 점은 'bad'로 설정하였다.

온도, 습도, 강수량 등의 3 가지 데이터를 독립변수로 설정하고, 종합 지표를 종속 변수로 설정하여 테스트를 진행하였다. 랜덤 포레스트 모델은 train, test set 이 각각 79.4%, 62%로 과대 적합이 발생하였다. 그레디언트 부스팅 모델의 대표적인 모델 중 트리 깊이를 최소화하면서 균형 잡힌 트리를 만들어가는 알고리즘인 XGBoost 를 사용하여 과대 적합 없이 가장 높은 63%의 정확도를 기록하였다.

3-2. K-means 알고리즘 & 실루엣 스코어

온도, 습도, 강수량과 실내 공기질 오염도의 상관관계를 더 명확하게 파악할 수 있는 종합 지표를 만들기 위해 비지도 학습을 활용하여 판단 기준을 정립하였다.

주어진 데이터를 k 개의 군집으로 묶어 각 군집과 거리 차이의 분산을 최소화하는 알고리즘인 K-means 를 이용하였다.

엘보우 메소드를 활용하여 군집의 개수(k)를 결정하였다. 엘보우 메소드는 군집 내 제곱 거리의 합(WCSS, Within Cluster Sum of Squares)을 이용하여 최적의 k 값을 도출한다. k=1 부터 k=11 까지의 다양한 k 값에 대한 WCSS 를 시각화를 진행하였다. k=3 일 때, WCSS 의 급격한 감소가 확인된다.

군집화 결과를 토대로 각 군집의 특성을 분석하였고, 0 번 군집의 데이터 중 80%를 랜덤으로 제거하여 데이터의 분포 특성을 명확하게 시각화한 결과를 그림 4 에서 확인할 수 있다. 이를 통해 군집 간의 차이점을 확인하고, 특히 0 번 군집의 특성을 더 자세히 관찰할 수 있다.

생성된 군집을 기반으로 환경 변수들 간의 상관관계를 분석하고 온도, 습도, 강수량 특성들을 기준으로 각 군집 분포를 확인하였다.

군집화 결과를 종합하여 새로운 공기질 판단 기준을 각 군집의 특성과 환경 변수 간의 상관관계를 이용해 평가 지표를 정의하였다. 이를 기반으로 환경 변수와의 복합적인 상호작용을 고려한 종합 판단 기준을 정하였다. 이는 미세먼지 및 기타 대기질 지수 외에도 다양한 환경 요소를 종합적으로 평가하는 데 활용될 수 있다. 이를 토대로 여러 군집의 개수에 대한 군집화가 조화롭게 이루어졌는지 확인하기 위해 시각화한 결과를 그림 3 에서 확인할 수 있다.

k=2 부터 k=6 일 때의 실루엣 스코어가 그림 3 과 같이 나타났으며, 개별 군집의 실루엣 계수 평균값의 편차를 고려하여 최적의 군집 개수를 k=3 으로 판단하였다.

최적의 군집 개수(k=3)를 이용하여 K-means 모델을 학습하고, 각 데이터를 세 개의 군집으로 나누었다. 이를 통해 각 군집은 데이터 간의 유사성을 기반으로

형성되었고, 군집화가 명확하게 되었다는 결과를 얻었다.

최종적으로, 모델의 성능을 높이기 위해 엘보우 메소드 및 실루엣 스코어를 활용하여 군집의 개수를 결정하고, 각 군집의 특성을 분석하여 스마트홈 공기질 모니터링 시스템에 적용할 수 있는 효과적인 판단 기준을 도출하였다.

IV. 결론

본 논문에서는 공기질 향상을 위한 스마트홈 모니터링 시스템 개발을 위해서 공기질에 영향을 미치는 특성의 비례, 반비례 등의 상관관계를 분석하였으며, 이를 토대로 CO₂, TVOC, PM_{2.5} 등의 3 가지 특성을 종합하여 룰 기반 공기질 종합 지표를 정립하였다. 공기질에 직접적으로 영향을 미치는 TVOC, PM_{2.5} 등이 온도, 습도 등과도 연관이 있다는 점을 확인하였다. 이러한 상관관계를 기반으로 온도, 습도, 강수량 등의 데이터만으로도 CO₂, TVOC, PM_{2.5}, PM₁₀ 등의 데이터가 모두 종합된 공기질을 판단할 수 있는 지표를 만들고 모델을 구축하였다. 엘보우 메소드를 이용하여 군집 개수인 k 의 최적 값을 찾았으며, k 별 군집화 결과를 시각화하였다. 제안된 모델은 비지도 학습인 K-means 알고리즘을 활용하여 공기질의 상태를 3 가지의 군집으로 나눈다. 해당 모델의 정확도를 위해 실루엣 스코어를 이용하여 검증하였다. 이러한 모델을 이용하여 사용자가 공기질 개선을 위한 의사 결정을 스마트하게 수행하는 데에 도움을 줄 수 있다.

참 고 문 헌

- [1] 정재원. (2022). 뉴스기사 빅데이터를 활용한 코로나-19 팬데믹 시대 이후 실내공기질에 대한 인식 변화. 한국생태환경건축학회 학술발표대회 논문집, 한국생태환경건축학회.
- [2] 심영선, 최해진. (2022). LSTM 및 회귀모델을 이용한 실내 공기질 예측 및 주거자 행동고지 알고리즘 연구. 대한기계학회 춘추학술대회, 대한기계학회.
- [3] 김현준, 김혜기, 김선숙. (2021). 학교 실내공기질 지표 개발을 위한 오염물질 농도 기준치 설정 방안. 한국건축친환경설비학회 논문집, 15(1), 87-98.