

미세먼지 관리를 위한 효율적인 데이터베이스 설계와 시각화 기법의 제안

박소연¹, 박지환^{2*}

¹단국대학교 대학원 인공지능융합학과, ²단국대학교 자유교양대학

An efficient database model for fine dust management and Suggestion of Visualization Techniques

Soyeon Park¹, Jihwan Park^{2*}

¹Department of Artificial Intelligence Convergence, Graduate School, Dankook University
²College of Liberal Arts, Dankook University

요약 최근 미세먼지(PM10)와 초미세먼지(PM2.5)의 위험성이 더욱 강조되고 있는 추세이다. 2016년 세계보건기구(WHO: World Health Organization, 이하 WHO)는 미세먼지로 인해 매년 3백만 명 이상이 조기사망을 한다고 밝혔으며, 그 중에서도 한국은 대기질 하위권 국가로서 증가하는 미세먼지 농도에 대한 대책이 강구되는 국가 중 하나이다. 이러한 문제를 해결하고자, 정부에서는 전 지역의 미세먼지 농도 데이터를 제공하고 있다. 하지만 현재 제공되고 있는 데이터가 각 지역의 '평균' 미세먼지 농도를 반영한다고 보기 어렵고, 방대한 양의 데이터의 경우, 새로운 DB설계에 따라 효율적으로 관리되어야 할 필요성이 요구된다. 이를 위해 본 논문에서는, 지역별 미세먼지 농도를 월평균단위로 분석하여, 새로운 DB설계에 따라 관리하는 방법을 제시하고, 미세먼지 데이터의 효율적인 관리를 위한 DB설계를 제시한다. 그리고 이에 따른 데이터 전처리기법을 소개한다. 더 나아가, 전처리한 데이터를 활용하여 지도 시각화를 수행하고 그 기법을 소개한다. 본 논문에서 제시한 DB설계, 미세먼지 데이터 전처리기법, 시각화 기법은 향후 다양한 미세먼지 관련 서비스에 적합하다.

Abstract Recently, the risk of fine dust (PM10) and ultrafine dust (PM2.5) has been emphasized. In 2016, the World Health Organization reported that more than three million people die early every year due to fine dust. Korea is one of the countries with lower air quality and needs to take measures against increasing fine dust concentrations. The government has provided fine dust concentration data from all regions to solve this problem. On the other hand, it is unclear if the data currently provided reflects the "average" fine dust concentration in each region, and efficient management according to the new database (DB) model is required in the case of vast amounts of data. This study analyzed the concentration of fine dust by region on a monthly average basis. The paper presents a method for managing fine dust data using a new DB model. It also introduces data preprocessing techniques accordingly. Map visualization is performed using preprocessed data, and techniques are introduced. The DB model, fine dust data preprocessing technique, and visualization technique presented in this paper are expected to be useful in providing various fine dust-related services in the future.

Keywords : Fine Dust, Ultrafine Dust, PM10, PM2.5, Fine Particulate Matter, Air Pollution, Air Quality

본 연구는 한국정부(MSIT)의 지원을 받는 한국연구재단(NRF) 연구비의 지원을 받았습니니다(No.NRF-2022R1G1A1011635).

*Corresponding Author : Jihwan Park(Dankook Univ.)

email: bosoagalaxy@gmail.com

Received October 5, 2023

Revised October 27, 2023

Accepted January 5, 2024

Published January 31, 2024

1. 서론

1.1 연구배경

1.1.1 연구배경 및 필요성

미세먼지 입자는 인체의 호흡기에 침투하여 건강에 상당한 영향을 미치는 것으로 입증되어 왔다. 그리고 미세먼지 농도의 증가는 천식에서 심지어는 폐암까지의 심각한 범위에 영향을 미친다[1,2]. 미세먼지는 세계가 집중해야 할 중대한 환경 문제로 손꼽히고 있다. 2016년 세계보건기구(WHO:World Health Organization, 이하 WHO)는 대기오염물질 중 미세먼지(PM10, PM2.5)로 인해 매년 3백만 명 이상이 조기 사망한다고 밝혔다[3]. 심지어 최근 2021년에는 16년만에 미세먼지와 초미세먼지의 권고 수준을 강화했다. 한국은 OECD국가 중 대기질 최하위권 국가로 자리매김하고 있는 실정이다[4]. 심지어 2020년 OECD가 발간한 보고서에 따르면, 한국 인구 중 99.2%가 WHO가 제시한 초미세먼지 농도 권고치에 해당하는 $10\mu\text{g}/\text{m}^3$ 에 노출된 것으로 파악됐다[5]. 따라서 정부 및 민간에서는 이러한 상황을 완화하고자 다양한 연구와 서비스 지원을 수행하고 있다. 특히 정부에서는 전 지역에 걸친 평균 미세먼지 농도 공공데이터를 제공하고 있다. 하지만 공공데이터가 실제적인 '평균' 미세먼지 농도를 보여준다고 하긴 어려운 상황이다. 즉, 각 지역의 미세먼지 농도 차이가 큰 경우에는, 제공된 미세먼지 농도 데이터가 모든 시민에게 적용된다고 보기 어렵다는 것이다[6]. 따라서 지역별 미세먼지 농도를 월평균단위로 파악한 뒤, 이 방대한 데이터를 데이터모델링을 통해 데이터베이스화하여 효율적으로 관리할 필요성이 요구된다. 본 연구에서는, 지역별 미세먼지 농도 측정 데이터를 효과적으로 확인할 수 있는 새로운 데이터모델을 제시하고, 이에 따라 데이터를 전처리하기 위한 기법을 제시한다. 더 나아가, 전처리를 마친 데이터를 시각화하여 세분화된 지역별로 월평균 미세먼지 농도를 손쉽게 파악할 수 있도록 하고자 한다.

2. 관련연구

서울시 초미세먼지(PM2.5) 지역별 극단치 분석[7]: 극단치 이론(EVT)을 활용하여 우리나라의 서울시의 구별 단위로 초미세먼지(PM2.5)의 위험 일 수 추정 및 리스크를 분석한 연구이다. 에어코리아에서 제공하는 국립환경과학원의 최종확정자료를 활용하였고 R studio에서

dplyr, ggplot2 패키지를 사용하여 전처리를 진행하였다[8,9].

한반도 인근의 미세먼지 시공간 농도 변화 연구[10]: 국내에서 측정된 미세먼지 데이터와 상층기류의 데이터를 python을 사용하여 정리 및 분석하고, 한반도 인근 미세먼지의 공간 및 시간별 농도 변화를 보여주었다.

기상데이터를 활용한 LSTM기반 미세먼지 농도 예측 방법 비교[11]: 기상데이터 및 PM10(미세먼지) 이력 데이터를 기반으로 LSTM(Long Short Term Memory)모델을 이용하여 미세먼지 농도 예측모델을 제안하였다. 모델을 구성하기에 앞서, 에어코리아에서 서울지역의 25개의 측정소를 대상으로 수집하고 PM10 데이터만을 필터링하여 일 단위 데이터로 통합하여 사용하였다.

빅데이터 활용 기반을 이용한 미세먼지 대응 방안 연구[12]: 빅데이터를 활용하여 미세먼지 대응방법을 도출하기 위해 딥러닝 통계 분석을 활용하였다. IoT기술을 활용하여 공기질 측정 데이터를 얻고, 이를 기반으로 미세먼지 취약지구 분석 및 미세먼지 대응 서비스에 활용하기 위한 기반 틀을 제시하고 있다.

3. 본론

3.1 데이터 전처리

3.1.1 데이터셋

PM2.5(초미세먼지) 측정자료가 2015년부터 존재하여, 2008년부터 2014년, 2015년부터 2021년과 같이 두 개의 다른 데이터셋을 구하여 차후에 병합하는 과정을 거쳤다.

2008년부터 2014년도의 데이터는 에어코리아의 최종확정 연도별 미세먼지(PM10) 측정자료를 사용하였고, Exel, csv 파일 모두 사용하였다. 결측치는 전처리 과정에서 따로 제외하였다. 주어진 데이터셋은 다음과 같다 (Table 1).

Table 1. 2008-2014 initial data set

...	Measurement date	...	PM10	Location
	2008010101		58	Seoul, Jung-gu, Seosomun-dong

(단위: $\mu\text{g}/\text{m}^3$)

해당 데이터에서 미세먼지만 다를 것이므로 Measurement date(측정일시), PM10, location(주소) column만 뽑아내야 한다. 그리고 측정일시까지 나와 있는 것을 일평균으로 평균을 구한 뒤, 월평균으로 계산하여 년도와 월까지만 표기한 6자리 숫자(ex. 200801)로 표현해야 한다.

2008-2014년도의 초미세먼지(PM2.5)는 2014년 이전에는 PM2.5가 측정되지 않았기 때문에, 미세먼지(PM10) 데이터만 처리하였다.

2015년부터 2021년 2월까지의 데이터는 통계청의 미세먼지(PM10), 초미세먼지(PM2.5) 월별 대기오염도(측정망별, 시도별, 도시별, 측정지점별) 통계자료 csv파일을 사용하였다.

2015년 1월부터 2021년 2월까지의 데이터를 수집하였으며 시-군 혹은 시-구 단위의 월평균데이터로 활용하였다. 주어진 데이터셋은 다음과 같다(Table 2).

Table 2. 2015-2021 initial data set

Loc1	Loc2	Loc3	2015.01	2015.02	...
Seoul	Jung	Seosomun			
Seoul	Jongro	Hyoje			

(단위: $\mu\text{g}/\text{m}^3$)

필요하지 않은 column을 제거하고 여러 column에 행정구역 별로 나누어져 있는 지역명을 하나로 합치는 과정이 필요하다. 향후 전처리 과정에서 측정년도와 월은 2008 - 2014 데이터에 맞춰서 6자리로 표현한다.

3.1.2 데이터전처리 환경

R studio: 2008-2014년도의 데이터, 2015-2021 년도의 데이터를 각각 시,군,구의 월평균 데이터로 전처리할 때 사용하였다.

Google colab: 2008-2014, 2015-2021 데이터를 전처리하여 병합하기 위해 사용하였다.

3.1.3 데이터전처리 기법

2008년부터 2014년까지의 데이터셋 전처리 과정은 다음과 같다.



Fig. 1. Preprocessing Sequence

Step0. 전처리 환경 구성

readxl, dplyr, data.table, tidyverse 패키지를 설치하고 라이브러리를 실행한다[13-15].

Step1. 전처리 데이터 불러오기

exel 파일은 list.files 과 purrr패키지 내의 map함수를 사용하여 불러오고, csv파일은 list.files과 lapply, do.call 함수를 사용하여 불러온다.

Step2. 필요한 column만 뽑아내는 과정

select 함수를 사용하여 필요한 column(주소, 측정일시, PM10)만 뽑아내었다.

Step3. 월평균데이터로 만들기 (전처리)

Step4. 저장(Table 3)

Table 3. 2008-2014 dataset preprocessing result

Location	Measure_date	PM10_average	PM25_average
Seoul-si, Jung-gu Seosomun-dong	200801		NA
	200802		NA
	200803		NA

2015년부터 2021년 2월까지의 데이터셋 전처리과정은 다음과 같다.

Step0. 전처리 환경 구성

readxl, dplyr, tidyr 패키지를 설치하고 라이브러리를 실행한다[16].

setwd 메소드를 활용하여 작업경로를 설정하고 해당 경로에 있는 파일을 list로 저장한다.

Step1. 1차 전처리

3개의 column에 흩어져있는 시군, 구, 동 각각을 paste 메소드로 한 column으로 결합하고 새로운 column으로 추가하는 등 기본적인 전처리를 진행하였다.

이후, tidyr과 dplyr메소드를 사용해 데이터프레임 형태를 전환한다.

Step2. 데이터프레임 정렬 후 병합

arrange메소드로 정렬 후, rbind메소드로 병합한다.

Step3. PM10 데이터프레임 & PM2.5 데이터프레임 병합 후 전처리

Location과 Measure_date column을 기준으로 하나의 데이터프레임으로 병합한 뒤, 불필요한 요소들을 제거한다.

Step4. 저장(Table 4)

Table 4. 2015-2021 dataset preprocessing result

Location	Measure_date	PM10_average	PM25_average
Seoul-si, Jung-gu Seosomun-dong	201501	48	20
	201502	74	24
	201503	72	25

(단위: $\mu\text{g}/\text{m}^3$)

2008년부터 2014년까지의 데이터와 2015년부터 2021년까지의 데이터를 병합하는 과정은 다음과 같다.

데이터 병합 시, 법정동코드를 기준으로 Location_id column을 작성해야 한다. 그런데 두 데이터프레임 사이 혹은 프레임 내의 지역표기에 문제가 있었다. 문제 케이스는 다음과 같다.

- 같은 지역도 다르게 표기한 경우
Ex) 강원 강릉시 옥천동 / 강원 강릉 옥천동
- 같은 시(시 데이터밖에 없는 경우) 혹은 군 혹은 구 안에서 또 세부항목으로 나뉘는 경우
Ex) 강원 동해시 천곡동 / 강원 동해시 동해항강원 양구 방산면 / 강원 양구 양구읍
- 시-군 혹은 시-구로 포함시키기 어려운 데이터
Ex) 서울 서울 / 서울 동작대로 / 서울 청계천

따라서, 위 문제들을 해결하고자 법정동코드(Location_id) 로 데이터를 통일하고, 이에 따른 지역명으로 Location까지 통일하였다. 이후 Location_id는 code로 칼럼명을 변경하였다.

3.2 데이터 시각화

3.2.1 시각화 환경

R studio에서 진행하였으며, htmltools 패키지를 활

용하여 시각화한 결과를 HTML파일로 저장한 뒤 확인할 수 있도록 제작하였다[17]. 그 외에도 sf, tmap, dplyr 패키지를 사용하였다[18,19].

3.2.2 시각화 기법

step1. 데이터 불러오기

위에서 전처리를 마친 미세먼지 데이터파일과 대한민국 시군구 지리정보가 포함된 공간데이터 셋 파일을 불러온다.

step2. 데이터 병합

법정동코드를 기준으로 미세먼지전처리 데이터와 공간데이터셋을 병합한다. 이후, 원하는 기간의 데이터만 추출하여 새로운 데이터프레임을 생성한다.

step3. 유효성검사

지도를 시각화하기 이전에, 데이터의 유효성 검사를 진행하고, 유효하지 않은 데이터는 유효한 데이터로 전환한다.

step4. 지도 시각화

tmap패키지를 활용하여, 'dust_month_sf' 데이터를 사용하여 시군구별 월평균 미세먼지 농도를 지도로 시각화한다.

step5. tmap 객체를 HTML로 변환

tmap_save 함수를 사용하여 tmap객체를 HTML파일로 저장한다.

4. 연구결과

4.1 데이터 전처리 결과

데이터 셋에서 자료획득율이 75% 미만인 데이터는 * 또는 **로 표기가 되어 있었는데, 이를 떼고 평균값을 낸 뒤, 따로 Incomplete 열을 두어 표기하였다. PM10_average 열과 PM25_average 열의 데이터의 단위는 $\mu\text{g}/\text{m}^3$ 이다(Table 5).

Table 5. 2008-2014 example of preprocessing result

Location	Measure_date	PM10_average	PM25_average	Code	Incomplete_10	Incomplete_25
Seoul-si, Jung-gu Seosomun-dong	201501	48	20	1114000000	0	0
Seoul-si, Jung-gu Seosomun-dong	201502	74	24	1114000000	0	0

4.2 데이터 시각화 결과

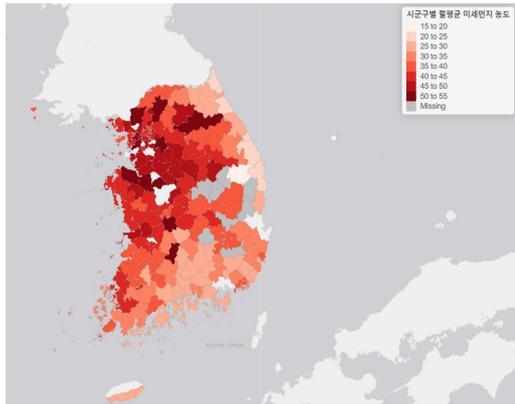


Fig. 2. Visualization of monthly average PM10 (fine dust) concentrations in January 2020

위의 지도 시각화 결과는, 각 지역의 2020년 1월의 월평균 PM10(미세먼지) 농도를 추출한 결과이다.

4.3 DB 모델 설계

DB 모델링은 가상의 PERSON 테이블과 연계하는 가정에 의해 작성되었다. 각각의 미세먼지 측정데이터를 구분하기위해 Dust_event_id를 주요 구분자인 Primary Key로 설정하고, Person_id를 가상의 Person과 연동할 수 있도록 외래키(Foreign -Key)로 설정했다. 현재 만들어진 DB모델(Table 6)은 pm10, pm2.5의 측정데이터를 월별로 정리하여 저장한다. Location_id는 시각화에 사용되는 시군구코드 정보를 저장하고 있으므로,

다른 데이터와 연동하여 분석시 지역별 정보를 구분하는 기준이 된다. DB모델은 지속적 업데이트를 진행하고 있으며, 추가적인 데이터, 예를 들어 일산화탄소(CO) 농도와 같은 정보를 추가하게 될 경우 새로운 필드를 추가하여 확장가능성을 갖는다. DB 모델링을 마친 뒤, Postgre DB에 row데이터를 적재하였다(Table 7).

5. 결론

기준에 진행되었던 서울시 초미세먼지(PM2.5) 극단치 분석은 서울시의 초미세먼지 데이터를 수집하여 전처리하였던 것에 반해, 본 연구는 서울시 뿐만 아니라, 분석범위를 전국으로 확대하여 시군구 단위의 데이터를 분석하였다. 그리고 전처리 결과를 바탕으로, 시각화를 진행하였고, 새로운 DB모델을 제시하는 데에 초점을 두었다.

한반도 인근의 미세먼지 시공간 농도 변화 연구는, 대상지역을 6개지역으로 선정하여 시간별 농도 변화를 추측하는 것에 초점을 두었지만, 본 연구는 다년간의 전국 시군구 단위의 월평균 데이터를 기반으로 지도 시각화를 진행하여 DB모델링을 진행하였다.

기상데이터를 활용한 LSTM기반 미세먼지 농도 예측 방법 비교는, 전처리과정에서 PM10 데이터를 추출하여 사용하였다. 본 연구에서는 PM10(미세먼지)와 함께 PM2.5(초미세먼지) 데이터를 함께 수집하여, 시각화 및 DB모델링에 활용하였다.

Table 6. DB model

Field	User Guide	Data type	Required	Primary -key	Foreign -key	FK Table
Dust_event_id	A unique identifier for each Dust event.	bigint	Yes	Yes	No	
Person_id	The PERSON_ID of the PERSON for whom "the dustevent" is recorded.	bigint	Yes	No	Yes	PERSON
Measurement_month	The year and month of the Measurment of dust.	Date	Yes	No	No	
Location_id	The code of residence of person.	int	Yes	No	No	
PM10_average	Average monthly PM10 measurements ($\mu\text{g}/\text{m}$).	Float	Yes	No	No	
PM2.5_average	Average monthly PM2.5 measurements ($\mu\text{g}/\text{m}$).	Float	Yes	No	No	

Table 7. Data distribution of PM10 and PM2.5

	강원도	경기도	경상남도	경상북도	광주광역시	대구광역시	대전광역시	부산광역시	서울특별시	울산광역시	인천광역시	전라남도	전라북도	제주특별자치도	충청남도	충청북도
n	1752	4746	1538	1987	787	1258	790	2466	3947	790	1595	1541	1429	242	1393	1066
count (%)																
2008	101 (5.8)317 (6.7)	84 (5.5)119 (6.0)	60 (7.6) 96 (7.6)	60 (7.6) 180 (7.3)	297 (7.5) 60 (7.6)	117 (7.3) 72 (4.7)	71 (5.0) 24 (9.9)	72 (5.2) 48 (4.5)								
2009	108 (6.2)324 (6.8)	79 (5.1)116 (5.8)	60 (7.6) 96 (7.6)	60 (7.6) 179 (7.3)	300 (7.6) 60 (7.6)	120 (7.5) 72 (4.7)	71 (5.0) 24 (9.9)	72 (5.2) 48 (4.5)								
2010	108 (6.2)324 (6.8)	84 (5.5)119 (6.0)	60 (7.6) 96 (7.6)	60 (7.6) 179 (7.3)	300 (7.6) 60 (7.6)	120 (7.5) 71 (4.6)	75 (5.2) 24 (9.9)	72 (5.2) 47 (4.4)								
2011	103 (5.9)371 (7.8)	84 (5.5)117 (5.9)	60 (7.6) 95 (7.6)	60 (7.6) 180 (7.3)	300 (7.6) 60 (7.6)	120 (7.5) 72 (4.7)	84 (5.9) 24 (9.9)	70 (5.0) 50 (4.7)								
2012	108 (6.2)372 (7.8)	84 (5.5)124 (6.2)	60 (7.6) 96 (7.6)	60 (7.6) 180 (7.3)	300 (7.6) 60 (7.6)	120 (7.5) 72 (4.7)	82 (5.7) 24 (9.9)	71 (5.1) 60 (5.6)								
2013	108 (6.2)372 (7.8)	93 (6.0)122 (6.1)	60 (7.6) 96 (7.6)	60 (7.6) 192 (7.8)	300 (7.6) 60 (7.6)	120 (7.5) 72 (4.7)	84 (5.9) 24 (9.9)	72 (5.2) 59 (5.5)								
2014	108 (6.2)372 (7.8)	96 (6.2)132 (6.6)	60 (7.6) 91 (7.2)	60 (7.6) 192 (7.8)	300 (7.6) 60 (7.6)	120 (7.5) 72 (4.7)	84 (5.9) 24 (9.9)	70 (5.0) 60 (5.6)								
2015	108 (6.2)372 (7.8)	96 (6.2)132 (6.6)	60 (7.6) 96 (7.6)	60 (7.6) 192 (7.8)	300 (7.6) 60 (7.6)	120 (7.5) 72 (4.7)	104 (7.3) 12 (5.0)	70 (5.2) 70 (6.6)								
2016	112 (6.4)372 (7.8)	104 (6.8)140 (7.0)	60 (7.6) 96 (7.6)	60 (7.6) 192 (7.8)	300 (7.6) 60 (7.6)	120 (7.5) 84 (5.5)	116 (8.1) 12 (5.0)	100 (7.2) 80 (7.5)								
2017	120 (6.8)372 (7.8)	120 (7.8)156 (7.9)	60 (7.6) 96 (7.6)	60 (7.6) 192 (7.8)	300 (7.6) 60 (7.6)	120 (7.5) 108 (7.0)	132 (9.2) 12 (5.0)	156 (11.2) 96 (9.0)								
2018	200 (11.4)372 (7.8)	160 (10.4)176 (8.9)	60 (7.6) 96 (7.6)	60 (7.6) 192 (7.8)	300 (7.6) 60 (7.6)	120 (7.5) 218 (14.1)	162 (11.3) 12 (5.0)	176 (12.6) 136 (12.8)								
2019	216 (12.3)372 (7.8)	188 (12.2)212 (10.7)	60 (7.6) 96 (7.6)	60 (7.6) 192 (7.8)	300 (7.6) 60 (7.6)	124 (7.8) 248 (16.1)	168 (11.8) 12 (5.0)	180 (12.9) 144 (13.5)								
2020	216 (12.3)372 (7.8)	228 (14.8)276 (13.9)	60 (7.6) 96 (7.6)	60 (7.6) 192 (7.8)	300 (7.6) 60 (7.6)	132 (8.3) 264 (17.1)	168 (11.8) 12 (5.0)	180 (12.9) 144 (13.5)								
2021	36 (2.1) 62 (1.3)	38 (2.5) 46 (2.3)	10 (1.3) 16 (1.3)	10 (1.3) 32 (1.3)	50 (1.3) 10 (1.3)	22 (1.4) 44 (2.9)	28 (2.0) 2 (0.8)	30 (2.2) 24 (2.3)								
PM10_ average (mean (SD))	39.81 (15.29)	52.23 (16.77)	40.69 (11.99)	41.01 (13.71)	42.53 (13.21)	44.72 (13.68)	42.77 (13.24)	44.62 (13.12)	46.33 (14.88)	44.65 (11.91)	48.72 (14.80)	36.24 (11.25)	44.77 (14.71)	39.47 (12.28)	43.48 (13.22)	45.60 (16.49)
PM25_ average (mean (SD))	20.42 (8.77)	25.00 (9.59)	19.44 (5.70)	19.93 (7.77)	22.89 (6.87)	22.74 (6.47)	20.83 (7.14)	23.13 (6.87)	23.79 (7.39)	22.02 (5.87)	23.22 (7.18)	19.24 (6.25)	24.78 (9.24)	17.69 (4.21)	23.04 (8.19)	25.30 (9.69)

빅데이터 활용 기반을 이용한 미세먼지 대응 방안 연구는 미세먼지 관련 빅데이터를 활용한 국민이 활용할 수 있는 서비스를 위한 기반 틀을 제시하였다. 본 연구는 더 나아가, 미세먼지 데이터를 어떻게 활용하고 서비스화 시킬 수 있는지 그 과정에 초점을 맞추어 연구를 진행하였다.

본 연구에서는 전국의 2008년부터 2021년 2월까지의 미세먼지, 초미세먼지 농도 데이터를 수집하여 전처리하는 과정을 상세히 다루었다. 이를 통해 지도시각화까지 진행하여 특정지역의 각 년도의 월 평균 미세먼지 농도와 심각성을 파악할 수 있었다. 이는 현 미세먼지의 심각성을 알리거나 해결하기 위한 다양한 서비스에서 유용하게 활용될 수 있다. 더 나아가, 다년간의 월평균 데이터 추이를 파악하여 앞으로의 미세먼지 농도를 예측하는 데에도 기여를 할 것으로 기대한다. 하지만 데이터셋에 결측값이 존재하였기 때문에 특정 지역은 정확한 농도파악이 어려웠다. 따라서 결측값을 최대한 보완하여 모든 지역의 정확한 미세먼지 농도 파악을 용이하게 할 수 있도록 해야 할 것이다.

References

- [1] Kang, Dongmug, and Jong-Eun Kim, "Fine, ultrafine, and yellow dust: emerging health problems in Korea.", Journal of Korean medical science Vol.29(5), pp.621-622, April. 2014. DOI: <https://doi.org/10.3346/jkms.2014.29.5.621>
- [2] Ritz, Beate, Barbara Hoffmann, and Annette Peters, "The effects of fine dust, ozone, and nitrogen dioxide on health.", Deutsches Ärzteblatt International Vol.116(51-52), pp.881-886, Dec. 2019. DOI: <https://doi.org/10.3238/arztebl.2019.0881>
- [3] World Health Organization, "Ambient air pollution: A Global Assessment of Exposure and Burden of Disease. 2016.", Meeting report, World Health Organization, Switzerland, pp.11
- [4] Lanzi, E., et al, The economic consequences of outdoor air pollution, Report, Organization for Economic Cooperation and Development, paris, 2016. DOI: <https://doi.org/10.1787/9789264257474-en>
- [5] van Zanden, Jan Luiten, et al, "How's Life? 2020: Measuring Well-being.", statistical report, OECD (2020), Paris, pp.130-131. DOI: <https://doi.org/10.1787/9870c393-en>
- [6] Chang, S., & Jeong, K, A mobile application for fine

dust monitoring system, In 2017 18th IEEE International Conference on Mobile Data Management (MDM), IEEE, Daejeon, Korea (South), pp.336-339, 2017, May. DOI: <https://doi.org/10.1109/MDM.2017.55>

[7] Jang-Wook Oh, and Tae-Jin lim, "Regional Analysis of Extreme Values by Particulate Matter(PM2.5) Concentration in Seoul, Korea.", Korean Society for quality management, Vol. 47, No. 1, pp.47-57, 2019, Mar. DOI: <https://doi.org/10.7469/KSQM.2019.47.1.47>

[8] Wickham, Hadley, et al, "dplyr: A grammar of data manipulation <https://CRAN.R-project.org/package=dplyr>, 2022.

[9] Wickam, Hadley. "ggplot2: elegant graphics for data analysis." Springer-V erlag. Accessed March 16 (2016): 2021.

[10] Chan-Joo Lee, and Min-Sun Hong, "Spatiotemporal Variations of Fine Particulates in and around the Korean Peninsula", Journal of Korean Society for Atmospheric Environment, Vol. 35, No. 6, pp. 675-682, 2019, Dec. DOI: <https://doi.org/10.5572/KOSAE.2019.35.6.675>

[11] Yang-Mo Seo, and Jae-Hong Yeom. "Comparison of LSTM-based Fine Dust Concentration Prediction Method using Meteorology Data.", Journal of the Korean Society of Metrology, Korean Society of Metrology, pp.117-120, 2019, April.

[12] Dae-sung Son, Young-sun Yoon, Yong-deok Park, and Sung-hyun Kim, "Study on how to respond to fine dust using big data," a collection of academic conferences of the Korean Intelligence Information System Association, Korean Intelligence Information System Association, pp.23-24, 2017, Nov.

[13] Wickham, Hadley, and Jennifer Bryan, "readxl: Read excel files.", R package version 1.1, 785, 2019.

[14] Dowle, Matt, Arun Srinivasan, and Tom Short. "data. table: Extension of 'data. frame'." R package version 1.8, 2019.

[15] Wickham, H., et al. "Welcome to the Tidyverse. J Open Source Softw 4: 1686.", 2019.

[16] Wickham, Hadley, and Lionel Henry. "Tidyr: Tidy messy data." R package version 1.2, 2020.

[17] Cheng, Joe, et al. htmltools: Tools for HTML. 2021.

[18] Pebesma, Edzer J., et al. "Simple features for R: standardized support for spatial vector data." R J, 10.1, pp. 439, 2018.

[19] Tennekes, Martijn. "tmap: Thematic Maps in R." Journal of Statistical Software 84, pp. 1-39, 2018.

박 소 연(Soyeon Park)

[정회원]



- 2023년 2월 : 단국대학교 SW융합 대학 모바일시스템학과(학사)
- 2023년 03월 ~ 현재 단국대학교 대학원 인공지능융합학과 인공지능 융합전공 (석사재학)

<관심분야>

모바일시스템, 정보통신

박 지 환(Jihwan Park)

[정회원]



- 2020년 8월 : 가톨릭대학교 대학원 의생명건강과학과 (의료정보학박사)
- 2023년 3월 ~ 현재 : 단국대학교 자유교양대학 교수

<관심분야>

의료정보, 정보통신