머신러닝 기법을 활용한

미세먼지 영향 변수 평가 연구

- PROGRESS REPORT -

김 진 형

2017.05.24

주요 연구내용 및 연구추진방법

데이터 수집 및 클리닝
- 미세먼지 농도에 영향을 미치는 요인에 대한 데이터 수집
- 비정상·결측값 처리 등 데이터 클리닝 작업
- 시공간 해상도 통일, 데이터 조인 등 알고리즘 적용을 위한 변수화

머신러닝 알고리즘 적용
- 앞서 구축한 데이터에 머신러닝 알고리즘 적용
- 알고리즘 변형을 통한 정확도 또는 안정성 향상

결과 해석 및 정책적 제언
- 분석 결과를 통해 미세먼지의 농도에 대한 설명력
- 해당 변수를 바탕으로 정책적 시사점 제안

주요 연구내용 및 연구추진방법

데이터 수집 및 클리닝

- 미세먼지 관련 문헌 및 이론 분석을 통한 변수 후보군 선정

변수 분류	변수	
발생원인 변수	직접 발생 원인	NOx, SOx 등
	간접 발생 원인	VOCs, O3, NH3, 빛에너지 등
기후기상요인 변수	기온, 기압, 강수량, 바람 등	
사회경제적 변수	가계	인구, 인구밀도, 가계 소득, 난방비 등
	기업	대기오염 배출 사업장 및 배출량정보 등
외부요인 변수	중국 동북부 미세먼지 농도, 황사발생일수 등	

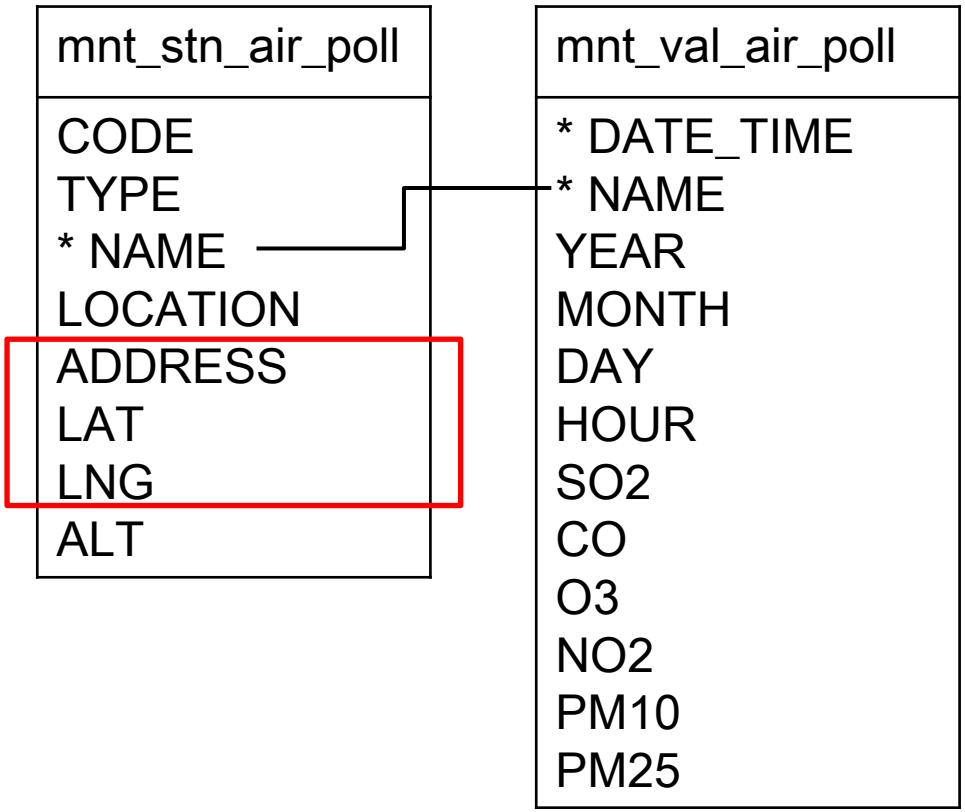
- 데이터 수집(각종 오픈 데이터)

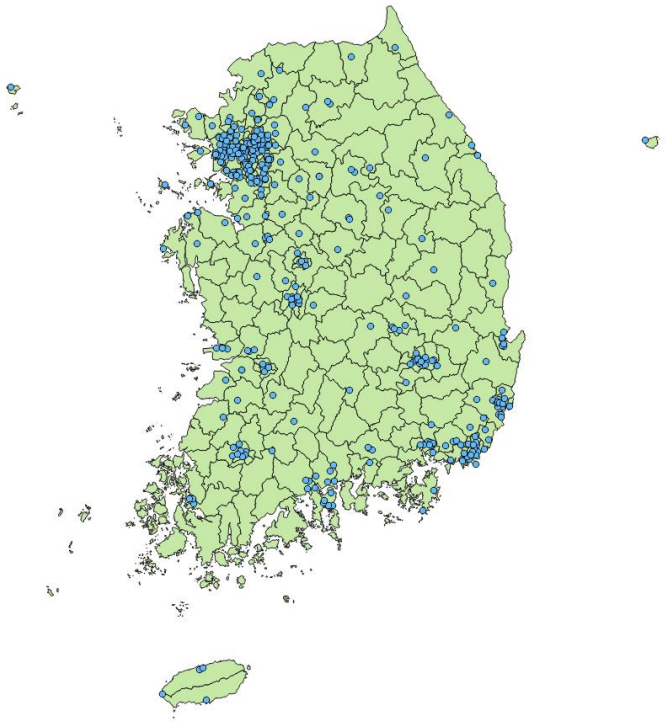
데이터 출처	URL	변수
기상청 국가기후데이터센터	http://sts.kma.go.kr/jsp/home/contents/main/main.do	기후 기상요인 변수
기상자료개방포털	https://data.kma.go.kr/cmmn/main.do	
에어코리아	http://www.airkorea.or.kr/index	발생원인 변수
국가통계포털	http://kosis.kr/	사회경제적 변수
환경공간정보서비스	https://egis.me.go.kr/main.do	
World Air Quality Index Sitemap	http://aqicn.org/map/china/kr/	외부요인 변수

이름	출처	공간해상도	시간해상도	시작	끝
대기오염물질	에어코리아(http://www.airkorea.or.kr)	측정망(Point)	1h	2001010101	2016123123
기상기후	기상자료개방포털(https://data.kma.go.kr)	측정망(Point)	1d, 1M	200101	201612
대기오염물질 배출량	국가 대기오염물질 배출량 서비스(http://airemiss.nier.go.kr)	시군구	1y	2001	2013
도시가스 수급	국가에너지통계종합정보시스템(http://www.kesis.net/)	전국	1m	201112	201612
도시가스 소비	국가에너지통계종합정보시스템(http://www.kesis.net/)	시도	1y	2001	2015
인구(11-16)	국가통계포털(http://kosis.kr/)	시군구	1m	201101	201612
인구(01-10)	국가통계포털(http://kosis.kr/)	시군구	1y	2001	2010
황사일수	기상청 국가기후데이터센터(http://sts.kma.go.kr/)	측정망(Point)			

국가대기오염물질 배출량 서비스	http://airemiss.nier.go.kr/mbshome/mbs/airemiss/index.do
에너지경제연구원	http://www.keei.re.kr/main.nsf/index.html
국가에너지통계종합정보시스템	http://www.kesis.net/
전력통계정보시스템(EPSIS)	http://epsis.kpx.or.kr/epsisnew/selectMain.do?locale=
kepco	http://home.kepco.co.kr/kepco/main.do
한전경제경영연구원	http://home.kepco.co.kr/kepco/KR/kemri/main/kemriMain.do
서울통계정보시스템	http://stat.seoul.go.kr/jsp3/index.j네
환경부	
주민등록 인구통계	http://rcps.egov.go.kr:8081/jsp/stat/ppl_stat_jf.jsp
서울연구데이터서비스	http://data.si.re.kr/
국가교통DB	https://www.ktdb.go.kr/www/index.do
전력 빅데이터 센터	http://home.kepco.co.kr/kepco/BD/bigData/main/bigDataMain.do
국가전력자원맵	http://nrgmap.kepco.co.kr/introEinfo.jsp
한국에너지기술연구원	http://www.kier.re.kr/
신재생에너지 데이터센터	http://kredc.kier.re.kr/kier/
한국에너지공단	http://www.energy.or.kr/web/kem_home_new/new_main.asp
대기배출원관리시스템(SEMS)	http://sodac.nier.go.kr/main/intro.jsp
사업장대기오염물질관리시스템	https://www.stacknsky.or.kr/stacknsky/main.j네
대기총량관리시스템	https://www.n-sky.or.kr/
환경정보공개시스템	http://www.env-info.kr/member/main/main.do
환경경영정보포털	http://www.gmi.go.kr/
국가주요지표	http://www.index.go.kr/main.do

대기오염물질



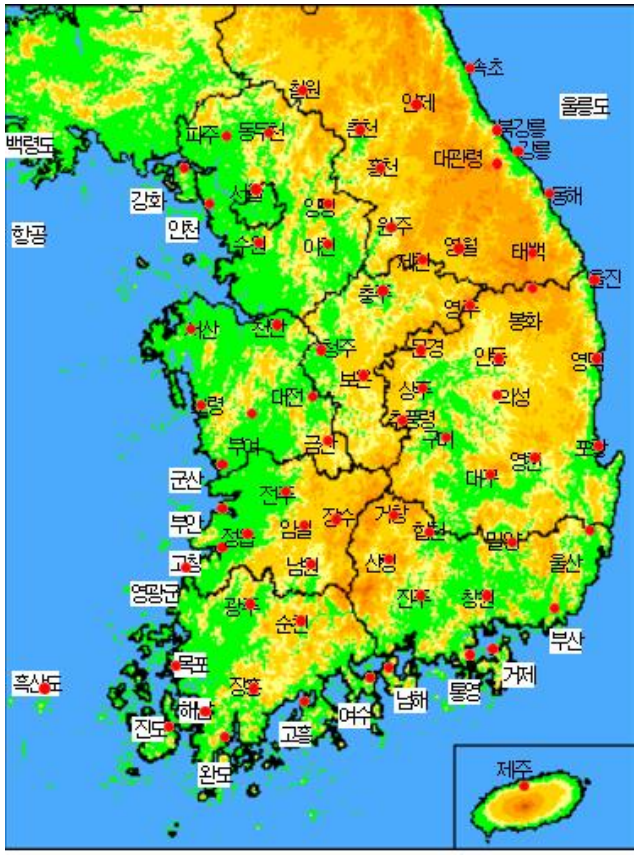


LOCATION	NAME	CODE	ADDRESS	LAT	LNG
서울	강남구	111261	서울 강남구 학동로 426 강남구청 별관 1동	37.51754	127.048
서울	강남대로	111264	서울 서초구 강남대로 201 서초구민회관 앞 중앙차로	37.48181	127.0357
서울	강동구	111274	서울 강동구 구천면로 42길 59 천호1동 주민센터	37.54498	127.1368
서울	강변북로	111143	서울 성동구 강변북로 257 한강사업본부 옆	37.53889	127.0416
서울	강북구	111291	서울 강북구 덕릉로41길 74 번1동 주민센터	37.63786	127.0289
서울	강서구	111212	서울 강서로45다길 71 화곡3동 푸른들청소년도서관	37.54465	126.8352
서울	공항대로	111213	서울 강서구 마곡동 727-1091 마곡역 중앙차로정류장 옆	37.57649	126.8307
서울	관악구	111251	서울 관악구 신림동길 14 신림동 주민센터	37.48738	126.9271
서울	광진구	111141	서울 광진구 광나루로 571 구의아리수정수센터	37.54508	127.0932
서울	구로구	111221	서울 구로구 가마산로 27길 45 (구로고등학교)	37.49826	126.8901
서울	금천구	111281	서울 금천구 금하로21길 20 (시흥5동 주민센터)	37.45235	126.9083
서울	노원구	111311	서울 노원구 상계로 118 상계2동 주민센터	37.65878	127.0685
서울	도봉구	111171	서울 도봉구 시루봉로2길 34 쌍문동청소년문화의 집	37.65414	127.029

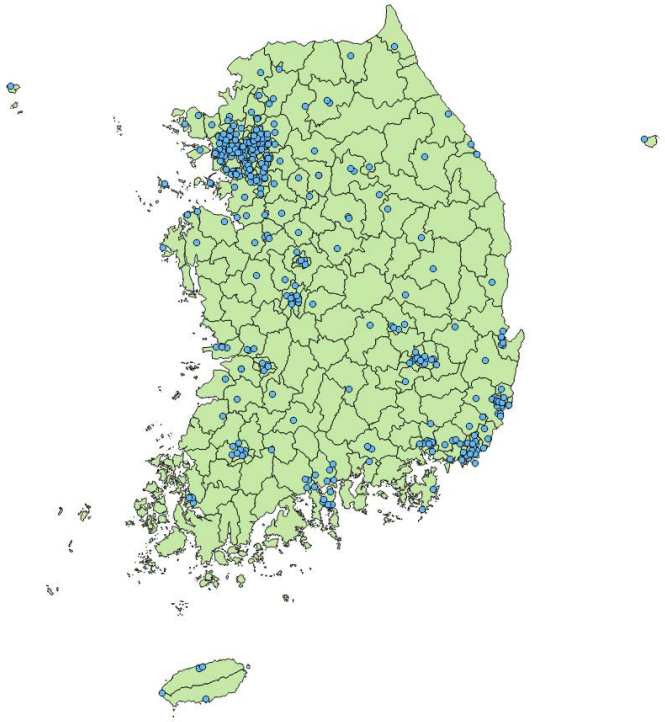
기상기후

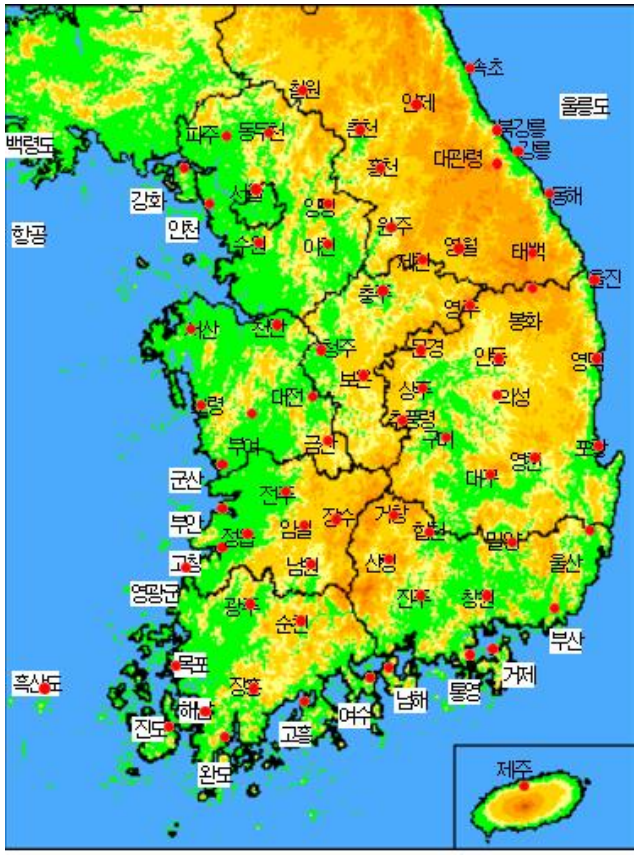
mnt_stn_weather
* CODE TYPE NAME LOCATION ADDRESS LAT LNG ALT

mnt_val_weather
* DATE_TIME * CODE YEAR MONTH HOUR TEMP ⋮



LOCATION	NAME	CODE	ADDRESS
서울	관악산	116	
서울	서울	108	
부산	부산	159	
대구	대구	143	
대구	대구(기)	176	
인천	강화	201	
인천	백령도	102	
인천	인천	112	
광주	광주	156	
대전	대전	133	
울산	울산	152	





문제점

- 대기오염물질 측정소와 기상기후 측정소 간의 차이

해결방안

- 기존 PM10 측정소를 기준으로 데이터를 매칭시키려 했으나 지역(시군구) 단위로 데이터를 매칭시키는 것을 고려 중

대기오염물질배출량

시도	시군구	배출원대분류	배출원중분류	배출원소분류	연료대분류	연료소분류	연도							
							CO	NOx	SOx	TSP	PM10	PM2.5	VOC	NH3

특징
- 배출원 별, 연료 별로 변수를 생성할 수 있음
- 시군구 단위로 데이터가 제공됨

문제점
- 연 단위 데이터임

해결방안
- 도시가스 수급 데이터를 통해 월단위로 나눔

도시가스 수급

단위 : 한국도시가스협회

	원료		생산(million ㎥)	소비(million ㎥)					
	프로판(1,000 ton)	LNG(million ㎥)		합계	산업	수송	가정	상업	공공
2011-12	7,714	3,092,206	3,101,137	2,571	815	100	1,319	326	12
2012-01	9,149	3,319,361	3,329,954	3,103	810	93	1,713	464	23
2012-02	8,106	3,121,209	3,130,594	3,017	832	91	1,625	446	22
2012-03	6,738	2,604,847	2,612,648	2,553	775	98	1,304	359	17
2012-04	4,665	1,815,238	1,820,639	1,925	691	94	883	243	14
2012-05	2,864	1,292,986	1,296,302	1,398	670	100	429	192	6
2012-06	2,168	1,161,47	1,163,98	1,198	641	101	257	193	6
2012-07	3,593	1,204,932	1,209,092	1,225	675	107	212	225	6

인구

행정구역(시	항목	단위	2011. 01 월	2011. 02 월	2011. 03 월	2011. 04 월	2011. 05 월	2011. 06 월
전국	세대수	세대	19869849	19891086	19906696	19923412	19935590	19945050
서울특별시	세대수	세대	4223301	4225512	4221613	4219198	4214866	4210104
종로구	세대수	세대	77569	77598	77380	77298	77224	77121
중구	세대수	세대	61423	61500	61419	61381	61285	61110
용산구	세대수	세대	113154	113151	113264	113255	113412	113330
성동구	세대수	세대	129388	129255	128922	128572	128319	128052
광진구	세대수	세대	159671	159706	159580	159447	159263	159091
동대문구	세대수	세대	158661	158954	158890	158874	158719	158336
중랑구	세대수	세대	176074	176110	176003	175949	175845	175699
성북구	세대수	세대	201402	201664	201594	201424	201097	200761
강북구	세대수	세대	142793	142894	142892	142880	142735	142626
도봉구	세대수	세대	140628	140584	140438	140402	140286	140159
노원구	세대수	세대	227678	227497	226913	226546	226258	225975
은평구	세대수	세대	196659	197094	197298	197507	197352	197209

특징

- 시군구 단위로 데이터가 제공됨

문제점

- 연 단위 데이터를 월별로 나눌 예정

황사일수

특징
- 시군구 단위로 데이터가 제공됨

문제점
- 모든 측정소의 데이터가 제공되지 않음
- 모든 시기의 데이터가 제공되지 않음
- 장소별로 파일이 따로 존재
- 황사 발생일로 데이터가 제공됨

조회기간 : 2001년 01월 01일 ~ 2016년 12월 31일
요소 : 황사
지점 : 강릉

날짜	시작	종료	시작	종료	지속시간
2001-01-02	8:45	24:00:00			15:15
2001-01-03	0:00	0:50			0:50
2001-03-03	7:40	19:35			11:55
2001-03-04	1:10	8:40			7:30
2001-03-05	8:30	22:10			13:40
2001-03-06	9:10	24:00:00			14:50
2001-03-07	0:00	17:20			17:20
2001-03-20	8:05	24:00:00			15:55
2001-03-21	0:00	9:50			9:50
2001-03-22	6:10	24:00:00			17:50
2001-03-23	0:00	8:40	17:20	21:50	13:10
2001-04-09	17:20	24:00:00			6:40
2001-04-10	0:00	24:00:00			24:00:00
2001-04-11	0:00	24:00:00			24:00:00
2001-04-12	0:00	10:30			10:30
2001-04-24	10:10	24:00:00			13:50

해결방안
- 파일 합치기
- 월별 데이터로 병합하기

향후계획

데이터 수집

- 토지이용도, 지형정보
- 중국 황사정보

데이터 전처리

- 시군구 단위로 데이터 매칭하기

Thank you!