

딥러닝을 이용하여 기후 및 대기 요소에 따른 전염성  
질환 발생패턴 분석

빅데이터팀 강선아  
2017. 5. 23

# 목차

---

I. 데이터 수집 현황 및 전처리

II. DB구축 예정

# 진행상황

## 연구계획과 진행상황 비교

구분	3월	4월	5월	6월	7월	8월	9월	10월	11월
◦ 선행연구 분석 및 데이터 변수 선정									
◦ tensorflow설치 및 데이터 수집/학습									
◦ 기후데이터와 질병 간의 상관관계 분석									
◦ 기후와 질병 데이터 간의 시·도별 패턴분석(GAM 이용)									
◦ 질병 예측을 위한 RNN 알고리즘 구축 및 예측									
◦ 보고서 작성									수정



업무	진행상황	비고
◦ 데이터 변수 선정	선정완료	◦ 기후, 대기, 인구통계 데이터 ◦ 국민건강보험공단 표본코호트 DB
◦ 데이터 수집	수집중	◦ 기후, 대기, 인구통계 데이터 수집완료 ◦ 국민건강보험공단 표본코호트 DB 심의완료 후 공단에서 자료구축 중
◦ 데이터 전처리	전처리 일부 완료	◦ 기후, 대기, 인구통계 데이터 기초 전처리 완료, 향후 표본코호트 DB 연계를 위해 추가 전처리 필요

# 1. 데이터 수집 현황 및 전처리

## 1 데이터 수집

- 기후, 대기, 인구통계 데이터 수집

데이터	설명
기후 데이터	<ul style="list-style-type: none"><li>- 데이터 수집: 기상자료개방포털</li><li>- 변수설명: 2003-2012년 기온(최고, 평균, 최저), 상대습도(최고, 평균, 최저), 월합강수량, 일조율, 적설량</li><li>- 101개의 측정소에서 수집</li></ul>
대기 데이터	<ul style="list-style-type: none"><li>- 데이터 수집: 에어코리아</li><li>- 504개의 측정소에서 수집</li><li>- 변수설명: 2003-2012년 SO2, CO, O3, NO2, PM10</li></ul>
인구통계 데이터	<ul style="list-style-type: none"><li>- 데이터 수집: KOSIS(국가통계포털)</li><li>- 변수설명: 연령별 시군구별 인구데이터(년기준), 시군구별 인구이동(월기준), 인구밀도(5년기준), 지역 및 성별 외국인 현황(년기준)</li></ul>

# 기후 데이터

지점	일시	평균기온(°C)	평균최고기온(°C)	평균최저기온(°C)	최고기온(°C)	최저기온(°C)	최고기온(°C)	최저기온(°C)	평균현지기온(°C)	평균해면기온(°C)	최고해면기온(°C)
90	Jan-00	-0.4	3.2	-4	9.9	-10.9	20000106	20000126	1021.7	1023.9	1037.5
90	Feb-00	-0.1	4.1	-4.4	9.2	-8.6	20000212	20000201	1017.4	1019.5	1034.1
90	Mar-00	6.9	11.5	2.1	18.5	-4.2	20000318	20000308	1012.9	1015	1028.3
90	Apr-00	11.2	16	6.5	24.1	2.4	20000413	20000403	1008.4	1010.4	1025.4
90	May-00	14.3	18.8	10.7	30.5	6.2	20000529	20000507	1007.3	1009.3	1018.9
90	Jun-00	19.2	23.2	15.6	33.3	9.7	20000606	20000610	1005.6	1007.6	1015.7
90	Jul-00	24.8	28.8	21.5	33.6	17	20000718	20000709	1002.8	1004.7	1015.7
90	Aug-00	24	27.2	21.6	33.7	17.9	20000811	20000801	1007.3	1009.3	1016.5
90	Sep-00	18.6	22.1	15.5	32	11.2	20000902	20000927	1011.6	1013.6	1024.8
90	Oct-00	14.5	18.5	10.3	23	5.1	20001008	20001031	1016.6	1018.7	1029.6
90	Nov-00	7.8	11.7	3.3	19.9	-2.1	20001104	20001122	1021.1	1023.2	1034.5
90	Dec-00	3.2	7.4	-1.3	14.5	-9.9	20001202	20001226	1019.1	1021.2	1031.7
90	01-Jan	-2.3	1.2	-5.8	6.2	-14.6	20010121	20010115	1019.3	1021.4	1035.8
90	01-Feb	1.2	5.3	-2.7	16.2	-8.4	20010222	20010202	1019.7	1021.9	1033.2
90	01-Mar	5.6	10.1	1.1	19.9	-6	20010320	20010311	1010.5	1012.6	1025.5
90	01-Apr	12.3	17.2	7.1	25.2	1.2	20010419	20010403	1012.3	1014.3	1024.7
90	01-May	17.2	22.2	13.5	31.2	6.5	20010513	20010503	1007.2	1009.2	1020.5
90	01-Jun	20	23.6	17	34.2	12.9	20010603	20010602	1004.9	1006.9	1015
90	01-Jul	23.4	27.1	20.1	31.3	15.7	20010713	20010707	1004.4	1006.3	1012.4
90	01-Aug	22.4	26.2	18.8	30.0	14.6	20010807	20010826	1008.5	1010.5	1018.2

# 시도, 연령별 인구 데이터

= 시군구/성/연령(5세)별 주민등록연앙인구					2003년	2004년	2005년
행정구역(시군구)별	성별	연령별	항목	단위			
전국	계	계	주민등록연앙인구	명	48308385.5	48485314	48683039.5
전국	계	0 - 4세	주민등록연앙인구	명	2852509	2705352	2531164.5
전국	계	5 - 9세	주민등록연앙인구	명	3467506	3381814	3297888
전국	계	10 - 14세	주민등록연앙인구	명	3374487.5	3457590	3532629.5
전국	계	15 - 19세	주민등록연앙인구	명	3239195.5	3159628	3137402.5
전국	계	20 - 24세	주민등록연앙인구	명	4057274.5	3981199	3802950.5
전국	계	25 - 29세	주민등록연앙인구	명	4003477.5	3890299.5	3894503.5
전국	계	30 - 34세	주민등록연앙인구	명	4612819	4546133.5	4440949.5
전국	계	35 - 39세	주민등록연앙인구	명	4261933	4353161	4418296
전국	계	40 - 44세	주민등록연앙인구	명	4436490	4404869.5	4338463
전국	계	45 - 49세	주민등록연앙인구	명	3563675	3785353.5	3986582
전국	계	50 - 54세	주민등록연앙인구	명	2543477.5	2635665	2829603.5
전국	계	55 - 59세	주민등록연앙인구	명	2065926	2149594.5	2249969.5
전국	계	60 - 64세	주민등록연앙인구	명	2020810	2019689.5	1997902.5
전국	계	65 - 69세	주민등록연앙인구	명	1544634.5	1621190.5	1676881
전국	계	70 - 74세	주민등록연앙인구	명	1026261	1085010.5	1162612
전국	계	75 - 79세	주민등록연앙인구	명	644337.5	681460.5	727178.5
전국	계	80세 이상	주민등록연앙인구	명	593572	627303.5	658063.5
전국	계	80 - 84세	주민등록연앙인구	명	375701.5	399144	416340
전국	계	85세 이상	주민등록연앙인구	명	217870.5	228159.5	241723.5
전국	계	85 - 89세	주민등록연앙인구	명	156994	163755.5	173749.5
전국	계	90 - 94세	주민등록연앙인구	명	49742.5	52453.5	55065
전국	계	95 - 99세	주민등록연앙인구	명	9108	10017	10985
전국	계	100세 이상	주민등록연앙인구	명	2026	1933.5	1924
전국	남자	계	주민등록연앙인구	명	24235167.5	24316613	24409658.5
전국	남자	0 - 4세	주민등록연앙인구	명	1490623	1411847	1318437.5
전국	남자	5 - 9세	주민등록연앙인구	명	1831718.5	1778589	1727771.5
전국	남자	10 - 14세	주민등록연앙인구	명	1789104	1838148	1879134.5
전국	남자	15 - 19세	주민등록연앙인구	명	1686289.5	1651535	1646915.5

# 인구 전출입 데이터

= 시군구별 이동자수 (명)							
행정구역(시군구)별	항목	단위	2003. 01 월	2003. 02 월	2003. 03 월	2003. 04 월	2003. 05 월
전국	총전입	명	670655	874247	881150	825703	822174
전국	총전출	명	670655	874247	881150	825703	822174
전국	순이동	명	0	0	0	0	0
전국	시도내이동-시군구내	명	230140	292912	327828	322664	318347
전국	시도내이동-시군구간	명	212099	275421	269364	254159	260373
전국	시도내이동-시군구간	명	212099	275421	269364	254159	260373
전국	시도간전입	명	228416	305914	283958	248880	243454
전국	시도간전출	명	228416	305914	283958	248880	243454
서울특별시	총전입	명	139391	192362	213244	200954	197717
서울특별시	총전출	명	137834	187734	217783	208301	205279
서울특별시	순이동	명	1557	4628	-4539	-7347	-7562
서울특별시	시도내이동-시군구내	명	38095	52370	71400	71428	68275
서울특별시	시도내이동-시군구간	명	52952	73252	81157	76140	76383
서울특별시	시도내이동-시군구간	명	52952	73252	81157	76140	76383
서울특별시	시도간전입	명	48344	66740	60687	53386	53059
서울특별시	시도간전출	명	46787	62112	65226	60733	60621
경로구	총전입	명	2588	3433	3841	3439	3502
경로구	총전출	명	2714	3392	4350	3978	3727
경로구	순이동	명	-126	41	-509	-539	-225
경로구	시도내이동-시군구내	명	528	618	1051	945	866
경로구	시도내이동-시군구간	명	1233	1677	1731	1634	1762
경로구	시도내이동-시군구간	명	1406	1909	2282	2116	1998
경로구	시도간전입	명	827	1138	1059	860	874
경로구	시도간전출	명	780	865	1017	917	863
경구	총전입	명	1651	2294	2703	2384	2577
경구	총전출	명	1917	2409	3119	2674	2928
경구	순이동	명	-266	-115	-416	-290	-351
경구	시도내이동-시군구내	명	295	338	508	539	571
경구	시도내이동-시군구간	명	801	1190	1398	1198	1306
경구	시도내이동-시군구간	명	1071	1343	1792	1475	1642
경구	시도간전입	명	555	766	797	647	700
경구	시도간전출	명	551	728	819	660	715
영산구	총전입	명	2809	3894	4397	3940	4185
영산구	총전출	명	3157	4225	4944	4421	4470
영산구	순이동	명	-348	-331	-547	-481	-285

# 외국인 현황 데이터

☐ 지역 및 성별 등록외국인 현황 ☐

행정구역(시군구)별	성별	항목	단위	2000년	2001년	2002년	2003년
총계	총계	지역민성별등록외국인	명	244172	267630	287923	437014
총계	남자	지역민성별등록외국인	명	143177	153449	159356	257628
총계	여자	지역민성별등록외국인	명	100995	114181	128567	179386
서울특별시	총계	지역민성별등록외국인	명	61920	67908	73228	102882
서울특별시	남자	지역민성별등록외국인	명	31581	34045	35869	48786
서울특별시	여자	지역민성별등록외국인	명	30339	33863	37359	54096
강남구	총계	지역민성별등록외국인	명	5150	5398	5675	7054
강남구	남자	지역민성별등록외국인	명	2826	2957	3150	3092
강남구	여자	지역민성별등록외국인	명	2324	2441	2525	3962
강동구	총계	지역민성별등록외국인	명	1204	1344	1534	2040
강동구	남자	지역민성별등록외국인	명	486	527	576	834
강동구	여자	지역민성별등록외국인	명	718	817	958	1206
강북구	총계	지역민성별등록외국인	명	777	913	1010	1431
강북구	남자	지역민성별등록외국인	명	307	345	360	548
강북구	여자	지역민성별등록외국인	명	470	568	650	883
강서구	총계	지역민성별등록외국인	명	1731	2053	1848	2670
강서구	남자	지역민성별등록외국인	명	886	1073	879	1307
강서구	여자	지역민성별등록외국인	명	845	980	969	1363
관악구	총계	지역민성별등록외국인	명	1522	1977	2353	4515
관악구	남자	지역민성별등록외국인	명	717	911	905	2052
관악구	여자	지역민성별등록외국인	명	805	1066	1448	2463
광진구	총계	지역민성별등록외국인	명	1746	1981	1967	3730
광진구	남자	지역민성별등록외국인	명	861	932	909	1710
광진구	여자	지역민성별등록외국인	명	885	1049	1058	2020
구로구	총계	지역민성별등록외국인	명	1689	2123	2426	6593
구로구	남자	지역민성별등록외국인	명	820	1036	1159	3530
구로구	여자	지역민성별등록외국인	명	869	1087	1267	3063
금천구	총계	지역민성별등록외국인	명	2337	2585	2591	6131
금천구	남자	지역민성별등록외국인	명	1113	1265	1300	3428
금천구	여자	지역민성별등록외국인	명	1224	1320	1291	2703

# 1. 데이터 수집 현황 및 전처리

## 1 데이터 수집

- 국민건강보험공단 표본코호트 DB

심의결과

변경신청

### 연구지원결정통지서

신청구분	연구지원신청		
연구명	환경 빅데이터 분석 및 서비스 개발-딥러닝을 이용한 기후 및 대기오염에 따른 전염성 질환 발생 패턴 분석		
연구책임자	강성원	연구지원유형	정책과제
심의결과	승인	결정일자	2017-05-23
안내	<연구지원신청건:승인> 수수료 80% 감면대상		
접수부서	빅데이터운영실	접수일자	2017-05-23
담당자	김현지	전화번호	039-736-2432

\* 신청 접수확인서의 접수내역이 다를 경우 공단에 통보하여 주십시오.

목록

신청번호 (연구관리번호)	신청데이터 구분	연구목적	연구명	신청일	상태
REQ0000008185	표본코호트DB	정책연구	환경 빅데이터 분석 및 서비스 개발-딥러닝을 이용한 기후 및 대기오염에 따른 전염성 질환 발생 패턴 분석	2017-05-15	자료구축중

# 1. 데이터 수집 현황 및 전처리

## 1 데이터 수집

- 국민건강보험공단 표본코호트 DB 수집 절차

<p>보건복지부 지정 공용기관생명윤리위원회</p>	<p>연구계획서 작성 및 심의신청-심의완료(승인번호 부여)</p>
<p>국민건강보험공단</p>	<p>제공절차</p>  <pre> graph LR     A[서류신청 (신청일)] -- 25일 --&gt; B[서류심의결정 (신청일부터 25일)]     B -- 15일 --&gt; C[비용납부 (신청일부터 40일)]     C -- 5일 --&gt; D[자료수령 (신청일부터 45일)]     </pre> <p>* 신청건수가 많을 경우 차기 심의일로 이관</p>

# 1. 데이터 수집 현황 및 전처리

## 2 데이터 전처리

- 대기데이터

지역	측정소명	측정소코드	측정일시	SO2	CO	O3	NO2	PM10	주소
서울	강서구	111212	2003030904	0.002	1.4	0.002	0.041	61	서울 강서구 화곡동
서울	강서구	111212	2003030906	0.002	1.4	0.002	0.044	60	서울 강서구 화곡동
서울	강서구	111212	2003030907	0.002	1.6	0.002	0.041	65	서울 강서구 화곡동
서울	강서구	111212	2003030908	0.003	1.7	0.003	0.048	105	서울 강서구 화곡동
서울	강서구	111212	2003030909	0.003	1.8	0.003	0.048	97	서울 강서구 화곡동
서울	강서구	111212	2003030911	0.005	1.8	0.004	0.064	124	서울 강서구 화곡동
서울	강서구	111212	2003030912	0.004	1	0.008	0.059	90	서울 강서구 화곡동
서울	강서구	111212	2003030913	0.004	0.6	0.016	0.037	59	서울 강서구 화곡동

- 지역, 측정소별, 월별 데이터 추출
- 결측치 제외 후 각 변수별 최대, 최소, 평균 도출

month	지역	측정소	평균					최대					최소				
			SO2	CO	O3	NO2	PM10	SO2	CO	O3	NO2	PM10	SO2	CO	O3	NO2	PM10
1	강원 강릉	옥천동	0.006011	0.707527	0.023589	0.008341	52.73805	0.02	2.3	0.051	0.018	155	0.001	0.4	0.005	0.004	8
1	강원 고성	간성읍	0.006343	NA	0.05144	0.006685	26.34993	0.021	NA	0.07	0.058	92	0.003	NA	0.011	0.001	1
1	강원 양구	방산면	0.00507	0.711809	0.034881	0.006353	32.27459	0.017	1.7	0.059	0.028	120	0.001	0.1	0.006	0.002	3
1	강원 원주	명륜동	0.006929	0.701766	0.012103	0.029102	93.28552	0.027	5.2	0.042	0.069	272	0	0.1	0.001	0.003	15
1	강원 원주	중앙동(원)	0.009867	0.871528	0.007821	0.032008	75.91159	0.029	3.8	0.033	0.072	222	0.001	0	0.001	0.002	21
1	강원 춘천	중앙로	0.003489	1.323077	0.01853	0.012337	99.98514	0.013	5.2	0.044	0.037	619	0.001	0.1	0.007	0.001	3

## 2. DB 구축

### 1 DB 관계도

기후데이터

지역명	
일시	최고상대습도
최저기온	평균강수량
최고기온	일최다강수량
평균기온	월합강수량
평균상대습도	일조시간합
최소상대습도	일조율
	월적설량합

표본코호트 DB

개인일련번호	
기준년도	소득분위
시군구코드	장애중증도구분
성	장애유형구분
연령대	장애등록년월
사망년월	
사망원인	
가입자구분	

대기데이터

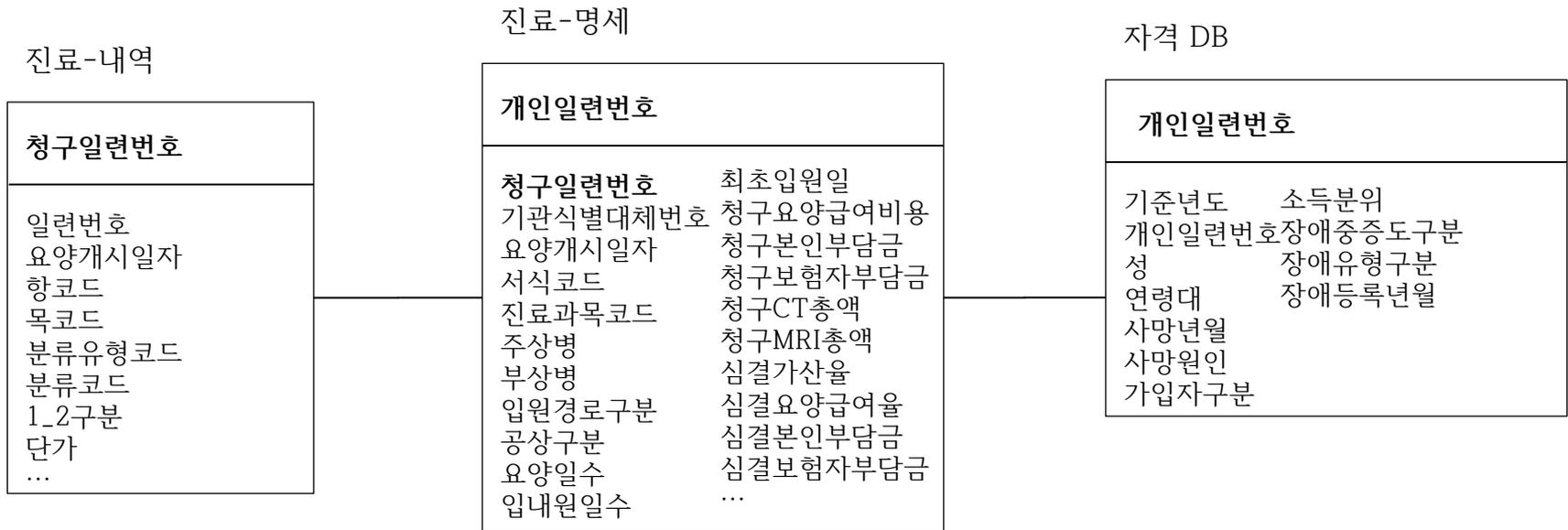
지역명	
일시	최고 O3
평균 SO2	최고 NO2
평균 CO	최고 PM10
평균 O3	최저 SO2
평균 NO2	최저 CO
평균 PM10	최저 O3
최고 SO2	최저 NO2
최고 CO	최저 PM10

인구통계데이터

지역명	
연도	
월	
연령	
인구수	
인구밀도	
총전입	
총전출	
외국인 현황	

## 2. DB 구축

### 1 DB 관계도



감사합니다