

예방해요! 유행성질병

- 국민건강알림서비스

국민건강보험공단

서울1(수목)-03 11조



분석개요

- **추진배경 및 필요성**
: 주요질병의 위험도 동향과 알림 제공 필요
- **추진목적**
: 예측모델의 정확도를 향상시키고 사용자화면을 개선, 정확하고 친절한 국민건강알람서비스를 구축하여 국민건강증진
- **성과**
: 7개권역별로 동일한 기상변수 및 환경변수를 사용하는 예측모델을 개발하여 질병 별 변수 확인.
위험도 모델링을 통해 질병 별 위험도 구간을 세분화하여 위험도 구간 설정
- **활용방안**
: 5개질병에 대한 실시간 전국 국민건강알람서비스 제공

활용데이터

- **활용데이터**

데이터 종류	수집 채널	수집 방법	수집 간격
정형	국민건강보험공간	관리도구 FTP	일별
	식약처	관리도구	수시
	기상청	API	시간별, 일별
	국립환경과학원	API	시간별, 일별
비정형	트위터	API	일별
	블로그	API	일별
	뉴스	API	일별
	검색 트렌드	API	주별

분석모델링

• 분석 프로세스

정형데이터와 비정형데이터를 이용, 질병의 진료건수를 예측하는 모형으로 적합한 음이항 회귀모형을 선택

- 고도화 알람 서비스

진료건수에 유의하게 영향을 미치는 변수들(기상변수,환경변수, 소셜 데이터)의 데이터를 자연스럽게 결합한 예측모형

- 식중독서비스

식약청의 확진 건수 정보, 기상, 환경, 소셜 데이터를 융합한 예측모델 기반

○ 감기 진료발생건수 예측모델

$y = \exp(8.890e+00 - 2.447e-03 \cdot \text{최저기온} + 5.481e-03 \cdot \text{일교차} + 6.264e-04 \cdot \text{습도} + 8.171e-04 \cdot \text{뉴스} + 5.290e-06 \cdot \text{트위터} + \text{의료급여} + \text{요일효과} + \text{지역가중치})$

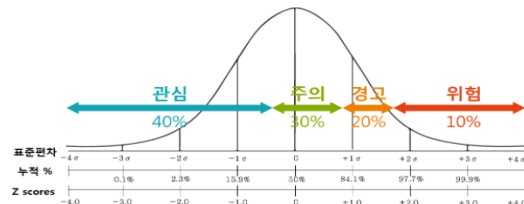
- y : 예측진료건수
- e : 십의 자승을 의미합니다. (예: $1.0E+03 = 1000$ / $1.0E-02 = 0.01$)

○ 감기 위험도 예측

예측된 질병 발생 건수를 과거 2년 평균 질병 발생 건수와 비교하여, 평균으로부터 벗어나 있는 정도를 기준으로 해당값의 분포도에 따라 위험도를 관심, 주의, 경고, 위험 4단계로 구분합니다.

단계	구간	구간 확률	누적 확률
● 관심	$-\infty \leq Z \leq -0.11$	40.00%	40.00%
● 주의	$-0.11 < Z \leq 0.58$	30.00%	70.00%
● 경고	$0.58 < Z \leq 1.14$	20.00%	90.00%
● 위험	$1.14 < Z \leq \infty$	10.00%	100%

• $Z = (X - \mu) / \sigma$ where μ =평균, σ =표준편차



2. 데이터 수집 - 프로세스

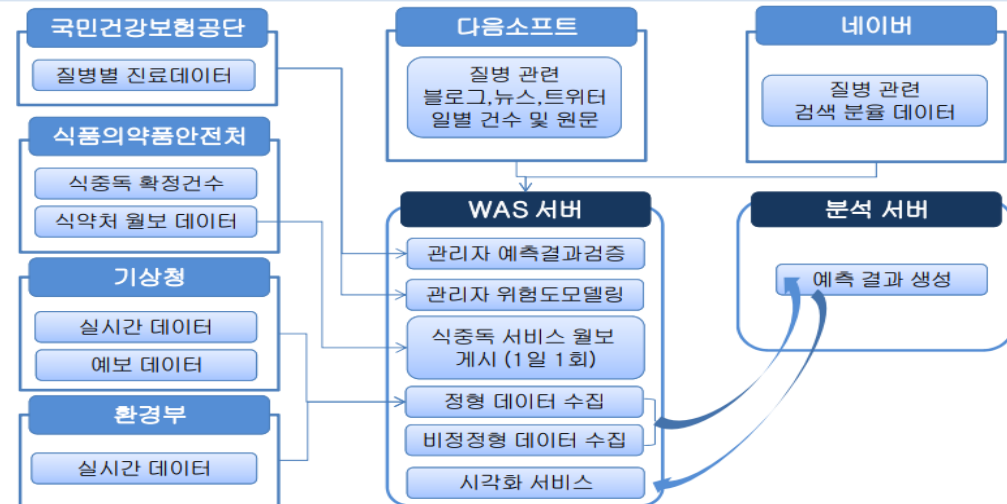
국민건강 알람서비스 고도화 사업

II. 사업 추진 내역



알람 대상 질병 확대 → 데이터 수집 → 질병 예측 모델 → 위험도 모델링 → 서비스 시각화

유관기관의 정형 데이터, 민간의 다양한 비정형 데이터를 일별, 시간별로 수집하여 예측 결과 생성



분석결과

• 분석결과

- 전국 예측모델과 지역적 유사성을 가지고 7개 권역별로 권역별로 동일한 기상변수 및
- 환경변수를 사용하는 예측모델을 개발
- 예측모델 구동결과 실제 값과 예측 값 사이에 0.90이상의 높은 상관도 존재

전국기준					
구분	감기	눈병	식중독	피부염	천식
상관도	0.912	0.97	0.9157	0.971	0.933

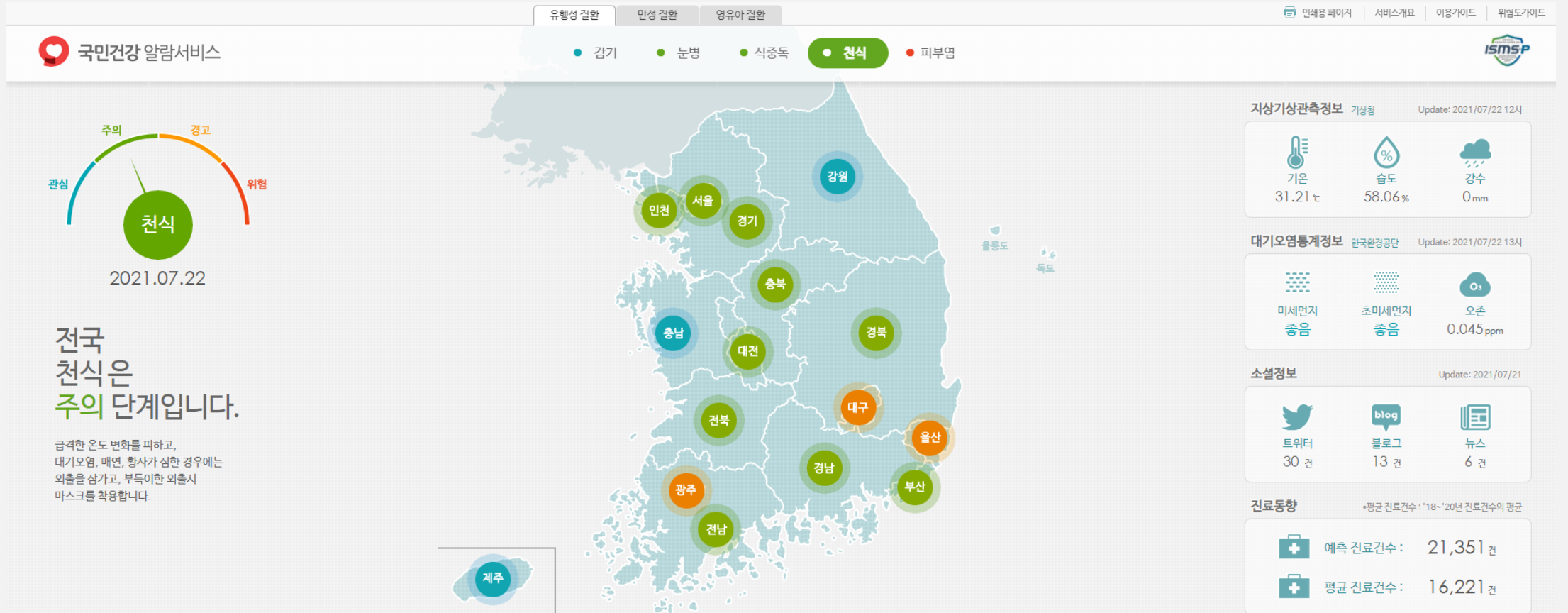
• 분석결과 활용 측면

- AS-IS : 질병특성 고려, 질병 별 3단계 위험도 구간 알림 서비스/알람 대상 질병 (9개)
- TO-BE : 정확하고 친절한 알람 서비스 구축
 1. 유관기관 협업
 2. 정형/비정형 데이터 보강
 3. 사용자 화면 개선
 4. 대상질병 확대 및 예측 정확도 향상

분석결과 및 시각화 예시

- 국민건강 알람서비스 홈페이지 화면

국민건강 알람서비스 (nhis.or.kr)



분석 사례 개선사항 도출

분석 모델 설계 관점

개선 필요점

5대 질병 외에 특정 시기에
더 주의가 요구되는 만성질환
(심혈관계 질병, 고혈압 등)에 대한
알람 서비스 항목의 한계

해결 방안

만성질환 서비스 항목 다양화 및
기상청 데이터를 활용하여 계절에
따른 위험도 알림 서비스 추가
(ex. 만성질환-심뇌혈관 환자들 온열질환
에 취약하여 여름철에 주의 필요)

예측변수로 사용된 변수 중
예측성 변수 부족

감기 진료발생건수 예측모델

$y = \exp(8.890e+00 - 2.447e-03 \cdot \text{최저기온} + 5.481e-03 \cdot \text{일교차} + 6.264e-04 \cdot \text{습도} + 8.171e-04 \cdot \text{뉴스} + 5.290e-06 \cdot \text{트위터} + \text{의료급여} + \text{요일효과} + \text{지역가중치})$

* y = 예측진료건수

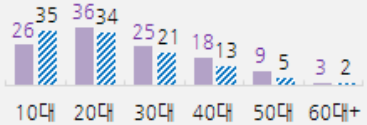
* e : 십의 자승을 의미합니다. (예 : $1.0E+03 = 1000$ / $1.0E-02 = 0.01$)

전년도의 time-series 상관성을 기반으로
전년도 발병현황 등 데이터 활용, 가중
치를 분석하여 이를 올해 예측모델에 변
수로 추가하거나 가중치 조정.

분석 사례 개선사항 도출

데이터 수집/활용 관점

트위터



개선 필요점

@YoungpyoHc 우한 폐렴으로 국민들은 고통을 호소하는 판
국에 대법원의 판결까지 성치권에서 논해야 되겠소? 자 생각해 봄
시다. 드루킹이 아무것도 ...

비정형 데이터 정제 미흡 및 데이터 추가 수집 필요
(ex. 영유아 폐렴에서 예측데이터로 트위터와
블로그의 자료 사용, “폐렴” 키워드만으로 자료가 잡
혀 영유아 폐렴과 성인 폐렴 유행 구분 불가)
(ex. 각 질병에 대한 동의어 질병명 처리 미흡)

해결 방안

비정형 데이터(소셜정보) 수집 채널 다양화
(네이버 카페, 지식인, 인스타그램 등등)
및 키워드 세분화하여 선정 필요
각 질병에 대한 동의어 질병명 처리 분류
체계의 주기적인 관리가 필요함

기상 관측 데이터가 실시간으로
반영되지 않음(최소 2시간 간격)

지상기상관측정보 기상청

Update: 2021/07/22 15시

기온
30.45℃

습도
61.89%

강수
0mm

최근 시간 대 별로 온도변화 폭이 더 큰
양상을 보여 즉각적인 데이터 업데이트가
요구됨.

분석 사례 개선사항 도출

활용된 분석모델 관점

개선 필요점

예측기간이 2일로 짧게 제공됨
감기, 식중독, 피부염만 제공



해결 방안

예측기간 증가 및
다른 질병에도 예측정보 제공

분석 사례 개선사항 도출

전국
감기는
관심 단계입니다.

환기를 자주 시켜 깨끗한 환경을
유지하고, 외출 후에는 반드시
손을 씻는 등 평소 손 씻기를
생활화합니다.

개선 필요점



행동요령 정보가 현재
위험 단계에 대한 것만
제공됨

해결 방안

모든 행동요령을 보여주고
현재 단계에 맞는 요령에
볼드를 해서 보여주는 등
으로 개선

시각화 및 결과 활용 서비스 관점

해당 서비스의 홍보 부족으로
서비스 인지도 낮음.

만성질환 알리기 캠페인에서 기상
서비스와 함께 해당 서비스 홍보
질병관리청 만성질환데이터 통계
사이트, 감염병 포털 사이트와 연결
하여 흩어져 있는 질병관련제공 데
이터 통합

자신이 속한 지역 내의 전체
질병을 한번에 보기 힘들

보고 싶은 구를 선택하고,
해당 시(군,구)의 질병정보
를 모아서 보여주는 시각화
서비스 추가 필요함

질병 별로 같은 정보가 제공
되지 않음(수족구병의 진료
동향이 제공되지 않음)

시행되고 있는 서비스가 제
대로 시행되고 있는지 점검
이 필요함.

