

빅데이터기반 디지털 트윈 활용 폭염 취약계층 예측 시스템의 설계 및 구현

나형선(HyungSun Na)*, 김종인(JongIn Kim)*, 안진현(Jinhyun Ahn)**, 전대성(Daesung Jun)***,
임동혁(Dong-Hyuk Im)****
*호서대학교 컴퓨터공학과
**제주대학교 경영정보학과
**e-mail : jha@jejunu.ac.kr
***전주대학교 행정학과
***e-mail: daesungj@jj.ac.kr
****광운대학교 정보융합학부
****e-mail : dhim@kw.ac.kr

Design and Implementation of Predicting the Heatwave Vulnerable Class Using Digital Twin Based on Big Data

HyungSun Na*, JongIn Kim *, Jinhyun Ahn**, Daesung Jun***, Dong-Hyuk Im****
*Department of Computer Engineering, Hoseo University
**Dept. of Management Information Systems, Jeju National University
***Dept. of Public Administration, Junju University
****School of Information Convergence, Kwang-Woon University

요 약

여름철만 되면 폭염 취약계층의 피해 소식이 꾸준히 발생하고 있다. 본 연구는 폭염 취약계층을 예측하기 위한 방법으로 통신사와 공공데이터에서 유동인구데이터, 전기사용량, 온도데이터, 건물 면적, 병원 접근성 등을 활용하여 분석하였다. 디지털 트윈 기법을 활용해 분석결과 높은 온도대비 면적당 전기사용량이 적으며 동시에 유동인구가 많고 병원 접근성이 떨어질수록 폭염 취약계층일 확률이 높을 것으로 예측하였다.

1. 서론

우리나라의 폭염일수는 80년대 8.2 일에서 점차 증가하여 최근 5년(2016~2020)에는 평균 18.14 일까지 증가하였다. 또한 가장 폭염이 심했던 2018년에는 온열질환자가 4,526명(사망 48명)으로 2011년부터 2017년까지 온열질환자 1,132명(사망 11명)에 비해 압도적인 수치이다. (환경부 보도자료 2019.7.29.) 최근에는 2018년과 대비하면 안정적인 상황으로 볼 수 있지만 지금까지의 통계를 볼 때 언제든 2018년과 같은 폭염재난이 발생할 수 있으므로 이에 관련하여 대비책이 필요하다.

기상학계에서는 폭염은 소리 없는 살인자로 불릴 정도로 폭염일수가 점점 증가함에 따라 선제적인 대응이 요구되고 있는 가운데 다양한 분야에서 폭염의 피해를 예측하고 대응하기 위한 노력이 강제되고 있다. 물론 의료, 복지, 환경 분야 등을 비롯하여 다방면에서 이루어지고 있으나 거기서도 혜택을 받지 못하는 폭염 취약계층이 존재하고 있다.

본 연구에서는 폭염취약계층을 미리 예측하기 위해 전주시를 기반으로 통신사와 공공데이터에서 유동인구데이터, 전기사용량, 온도데이터, 건물 면적, 병원접근성 데이터 등을 활용하여 분석하였다. 그리고 각 데이터 별 어떤 연관이 있는지 분석하였고, 이 결과를 바탕으로 디지털 트윈 기법을 활용해 폭염취약계층을 예측하였다. 이러한 결과는 이후 폭염취약계층을 지원하기 위해 활용될 수 있을 것이다.

2. 관련연구

[1]에서도 폭염지수를 산출하는 시스템을 제안하였으나 [1]에서는 상대습도를 추가하여 폭염지수를 산출하였으며 본 연구는 다양한 공공 데이터를 활용하여 폭염지수를 산출하는 것이 가능하며 특히 각 건물마다 폭염지수를 계산할 수 있는 차이점을 가진다.

[2]에서는 폭염재해 취약지역을 분석하였으나 위성자료를 이용한 지표온도를 기반으로 분석하여 공간적 특성을 관련하여 분석하였으며 본 연구는 각 건물마

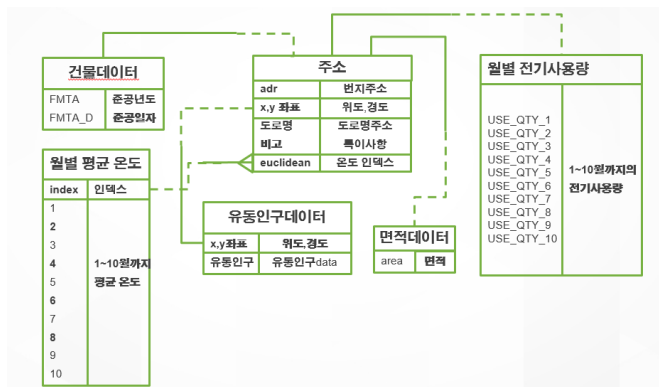
다 폭염지수를 계산할 수 있는 차이점을 가진다. [3]에서는 기후노출, 민감도, 적응능력 등 환경적, 물리적, 사회경제적 요소가 포함된 변수들을 가지고 폭염 취약성의 공간적 효과를 분석한다.[3]연구도 마찬가지로 공간에 중점을 두었다는 차이와 본 연구에서는 전기 사용량을 사용한다는 차이점이 있다.

3. 데이터 전처리

3.1 주 데이터 선정(Main Data Selection)

가장 먼저 폭염취약계층을 예측할 때 중요도가 높은 칼럼들을 보유 데이터를 기반으로 하여 선정하였다. 보유 데이터는 전주시 공공데이터와 통신사 이동거리 데이터이다. 그림 1은 통합 데이터의 E-R 다이어그램을 보여준다. 선정된 주 데이터는 면적, 전기 사용량, 온도, 유동인구, 건물연식, 병원 접근성 총 6가지의 데이터이다.

추가로 그림 1에서처럼 x, y 좌표의 경우는 V-World를 이용하여 시각화 하기 위해 사용한다.



(그림 1) 주요 칼럼들 테이블

3.2 데이터 클리닝(Data Cleaning)

면적, 전기사용량 데이터에서 극단적인 값들이 발견되었다. 예를 들어 면적이 0으로 되어있거나 전기 사용량이 다른 값들에 비해 극단적으로 높은 경우 등이 발견되어 이러한 값들은 제거하는 작업을 수행하였다. 온도, 건물연식 데이터에서는 빈 속성 값이 발견되었다. 온도 데이터의 경우는 평균값으로 대체하였고 건물연식 데이터의 경우는 최근 건축물 인허가 년도로 대체 하였다.

3.3 데이터 통합(Data Integration)

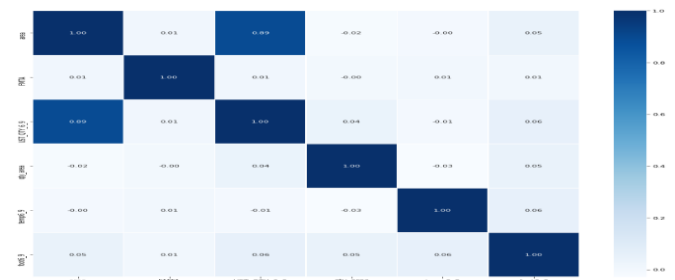
온도데이터의 경우 전주시 안에 총 16개의 온도측정소가 존재하고 간략한 위치는 그림 2에서 보여준다. 그리고 각 건물 별 가장 가까운 측정소의 온도를 부여하였다. 유동인구 데이터의 경우는 각 건물 별이 아닌 좌표 셀 별로 데이터가 존재하기 때문에 이 또한 가장 가까운 셀의 데이터를 부여하였다. 마지막으로 여름철 폭염에 대한 분석임으로 6~9월달의 데이터만 부여하였다.



(그림 2) 전주시 온도 측정장소

3.4 상관관계 분석(Correlation Analysis)

마지막으로 주 데이터의 상관관계를 분석하였다. 우선 각 칼럼 별 상관관계를 색상으로 구분을 준 그림 3과 같이 각 칼럼들간 전기사용량과 면적의 관계를 제외하고는 상관관계가 크게 없다는 것을 확인하였고, 따라서 전기사용량과 면적데이터를 면적당 전기 사용량으로 합성하였다.



(그림 3) 주요 칼럼들 상관관계 시각화

4. 예측 방법

폭염 취약계층의 데이터를 분석해본 결과 유동인구가 적고, 온도가 높을 때 면적당 전기사용량이 적은 경우가 많다는 결론을 내렸다. 추가로 병원 접근성이 낮고 건물 연식이 오래될 경우 소외계층일 확률이 높았다.

점수부여의 경우 면적당 전기사용량이 낮을수록 높은 점수를 부여하였으며 유동인구가 많을수록, 평균보다 기온이 높을수록 추가 점수를 부여하였다. 추가로 병원 접근성이 낮고 건물 연식이 오래되었을 경우 전체 점수에 큰 가산점을 주었다.

이 점수들을 기반으로 디지털 트윈 기법을 활용하여 추가점수와 가산점등의 비율을 조정해 가며 상위 30%의 건물들을 V-world를 이용해 시각화를 진행하였다. 그림 4는 V-world를 이용하여 시각화한 이미지를 보여준다.



참고문헌

- [1] 최병진, 정계명, 김남수, 이임학. (2017) 복사대류 온도 및 상대습도 가중치를 적용한 도시 미소공간 폭염지수 산출시스템. KR10-1841217
- [2] 김호용, 김지숙. (2020). Landsat 8 영상과 취약성 분석을 활용한 폭염재해 취약지역의 특성분석. 한국지리정보학회지, 23(1), 1-14.
- [3] 최예술, 김재원, 임엽. (2018). 서울시 폭염 취약지역의 공간적 패턴 및 적응능력 취약지역 분석. 국토계획, 53(7), 87-107.

(그림 4) 폭염취약계층 예측 건물

5. 결론

본 연구는 폭염취약계층예측에 영향을 주는 요소들을 공공데이터 안에서 선별하고 이에 가중치를 부여하여 분석한 후 폭염취약계층을 도출하였다.

본 연구에서는 전기사용량, 유동인구, 온도 데이터를 한 해의 데이터만 활용하였다는 한계가 있는데, 다수 년도의 데이터를 활용한다면 더욱 정확하고 의미 있는 결과가 도출될 것으로 보인다.

Acknowledgement

본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 대학 ICT 연구센터지원사업의 연구결과로 수행되었음 (IITP-2020-2018-0-01417).

또한, 2018 년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. NRF-2018R1D1A1B07048380)