

국토정책 Brief

국토연구원에서 수행한 주요 연구과제의 핵심 내용과 정책제안 등을 압축해 국민께 알려드리고자 하는 발간물입니다.

장요한 부연구위원
손재선 부연구위원
최혜림 연구원

2024. 12. 2.
No. 992



발행처 국토연구원
발행인 심교언
www.krihs.re.kr

이 브리프는 나무를 베지
않고 만든 생분해성 펄프
용지를 사용하였습니다.

A.I.와 민원 빅데이터를 활용한 민원지도의 민생현안 모니터링 방안

주요 내용

- ① 민원불편신고 등 각종 민감정보가 취합되고 기록되는 국민신문고 민원 빅데이터는 전국적으로 분포한 민원인의 다양한 관심사와 세부적인 내용이 담긴 데이터로, 민생현안 모니터링 도구로 그 효용가치가 높음
- ② 국민권익위원회에서 운영 중인 ‘한눈에 보는 민원 빅데이터’의 민원지도는 개인정보 보호의 문제, 데이터의 정확성과 신뢰성, 사용자 접근성과 활용성 등의 한계로 민원 빅데이터의 잠재력을 충분히 활용하지 못하고 있음
- ③ 단순 민원 빅데이터 내에서 그 활용방안을 발굴하는 것을 넘어 다양한 이종(異種)데이터 융·복합을 통한 동적인(dynamic) 민원지도를 시범 개발하고, 이를 통한 정책적 활용방안 필요
- ④ 딥러닝 등 A.I. 방법론을 활용한 시범분석을 통해 언제-어디에서-얼마나-어떻게 민원이 발생할지에 대한 예측 모델을 제시함으로써 선제적 정책지원 도구의 발전방향을 제시

정책방안

- ① 민원데이터 공개 확대 및 체감형 분석방법론 개발: 지자체 및 부처 협의를 통해 민원 빅데이터 공개를 확대하고, 시민 주도 자정환경 및 정책소통 기반 마련 필요
- ② A.I. 기술을 응용한 방법론 고도화로 정보의 효용성 제고: 민원지도 정보의 편향성을 방지하기 위해, A.I.와 이종데이터 융합을 통한 지역맞춤형 이슈 키워드 발굴 및 방법론 고도화
- ③ 공간정보 기반 데이터 융·복합을 통한 민원지도 효용성 및 정보 공개 확대: 공공관심 민원 분야에서는 격자 기반 공간 단위 일원화와 비식별화로 통계생산과 정보 공개 확대 및 활용도 제고
- ④ 국민참여형 정책소통 도구 발전: 민원 빅데이터 분석방법론 개발을 통해 악성 반복 민원에 대한 자동정제 기능을 구축하고, 정보 왜곡 최소화 및 행정인력의 효율적 활용과 시스템 고도화 필요

01. 민생현안 모니터링과 민원 빅데이터

민생현안이란 일반 국민의 생활에 직접적으로 영향을 미치는 사회·경제·문화 등 국민의 관심을 의미

일반 국민의 일상생활 및 생계와 직결된 단어인 '민생'은 각종 정책과 법 모두에서 중심이 되는 단어인 만큼 그 중요성은 강조해도 지나치지 않음

- 국토의 각종 계획, 운영, 인프라 구축 등 정책적 방향은 반드시 국민을 위한 것이어야 하며, 계획이나 추진 중인 정책의 실효성 역시 국민을 위한 것이어야 함

기존 민생현안 해결은 하향식 발굴 및 추진방식이 주를 이루어, 국민체감형 민생현안 해결과는 괴리가 발생

- 최근 사회·경제적 변화로 인해 다양한 민생현안이 지속해서 발생하고 있으며, 이에 대한 정부의 신속하고 효과적인 대응이 요구됨
- 국민체감형 민생현안 해결을 위해 뉴스기사, SNS, 인터넷정보 등 다양한 노력이 시도되고 있으나, 지역별 대표성 문제와 막대한 데이터 구매 및 수집비용 등은 한계로 지적

민원 빅데이터와 뉴스기사는 민생의 다양한 관심사가 상세하게 기록된 현장감 있는 데이터

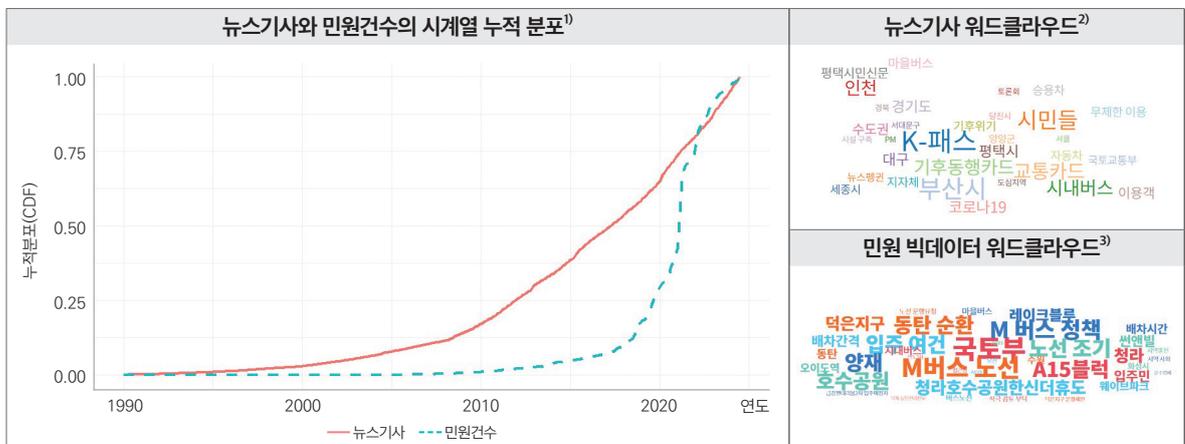
뉴스기사와 민원 빅데이터 두 가지 모두 전국에 걸친 고른 공간적 대표성과 다양한 관심사가 기록된 데이터로, 민생현안 모니터링을 위한 정책 모니터링 도구의 효용성이 높으나, 세부적인 지역 특성 고려 측면에서는 민원 빅데이터가 유리

- 뉴스기사는 정형화된 구성과 공신력이 높은 정보로서 범국민·범국토 차원의 공공현안 관련 내용이 주를 이루지만, 그 내용이 다소 개괄적이고 거시적인 특징이 있음
- 민원 빅데이터는 전국에 걸친 고른 공간적 대표성과 개인 차원의 세밀하고 다양한 관심사가 기록된 데이터로, 민생현안과 직접적인 연관성이 가장 높다고 할 수 있으나, 그 내용은 다소 비정형화되어 있음

'대중교통' 키워드를 예시로 민원 빅데이터와 뉴스기사의 시계열 및 워드클라우드를 비교하여 살펴보면, 민생현안 모니터링의 관점에서 두 데이터의 특징이 두드러지게 확인

- 시계열 노출 빈도를 누적분포함수(Cumulative Distribution Function: CDF)로 나타내면, 뉴스기사는 시간의 흐름에 따라 완만한 누적 증가를 보이지만, 민원 빅데이터는 특정 시기에 따라 증감의 변화가 다양한 패턴으로 확인
- 워드클라우드에서는 같은 조건하에서도 다소 추상적이고 거시적인 뉴스기사의 키워드임에 반해, 민원 빅데이터는 더 구체적이고 생활권 단위의 미시적 키워드가 다수 확인되어 지역별 민생현안 파악에 더욱 효과적

그림 1 '대중교통' 검색 키워드에 따른 뉴스기사와 민원 빅데이터의 시계열 분포 및 워드클라우드 비교 예시



주: 1) 뉴스기사는 시계열 변화에 따라 완만하게 누적함에 수렴하는 추이와 달리, 민원은 상대적으로 가파른 추이로 누적함에 수렴하는 형태를 나타냄.

2), 3) 2014년 1월 1일부터 2024년 7월 31일까지 집계된 '대중교통' 키워드에 따른 키워드 워드클라우드 결과임.

자료: 1) '대중교통' 키워드에 대해 국민권익위원회 '민원 빅데이터 분석정보 API 2022'와 '빅카인즈' 뉴스 기사를 중심으로 저자 재구성.

2) 빅카인즈 <https://www.bigkinds.or.kr> (2024년 8월 8일 검색).

3) 한눈에 보는 민원 빅데이터. <https://bigdata.epeople.go.kr> (2024년 8월 8일 검색).

현재 민원 빅데이터는 특정 지역과 분야에 집중되어, 지역맞춤형 정보 전달에는 한계

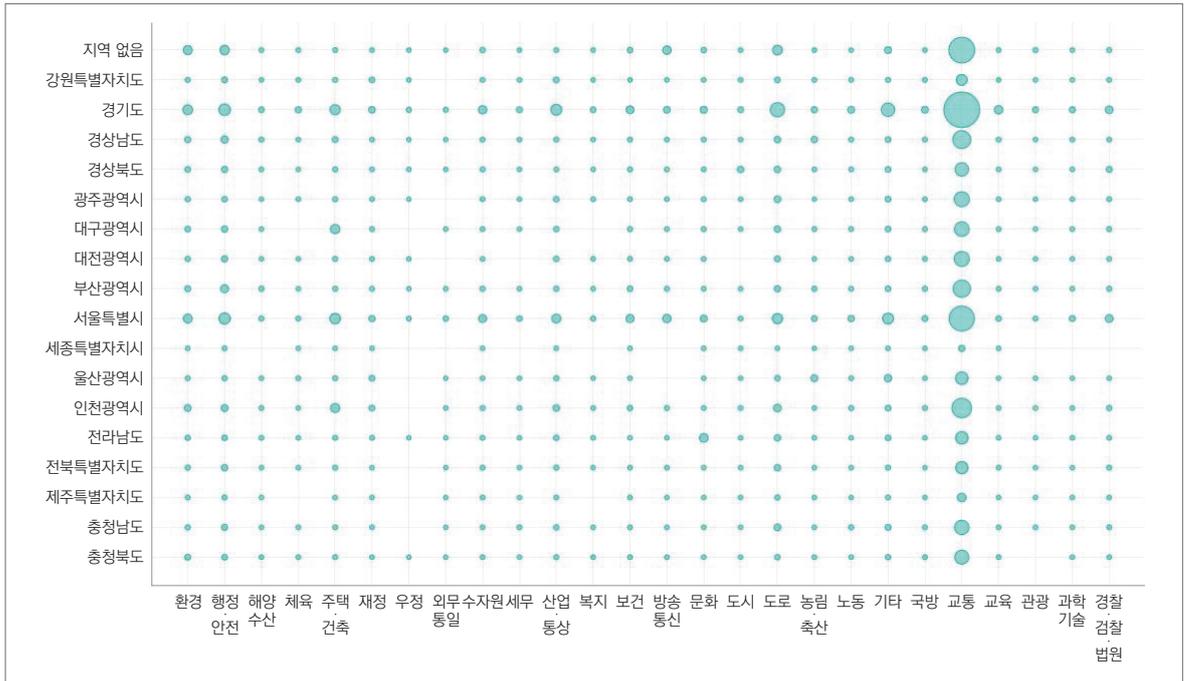
민원으로 축적된 민원 빅데이터는, 다수의 불규칙성과 특정 관심 분야에 집중될 수 있다는 점에서 여과 없이 활용할 경우 정보 왜곡 발생에 유의할 필요가 있음

- ‘교통’ 분야에 집중된 민원, 인구밀도가 높은 수도권에서의 민원이 대다수를 차지하여 정보의 왜곡이 우려됨
- 다양한 접수방식을 통해 전국에 걸쳐 지속적으로 수집되고 축적되는 민원정보를 활용하여, 국민과 정부가 소통하고 이슈 공론화를 통한 문제 해결방식의 시도가 필요

정보의 편중성, 공간정보의 불투명성 등 민생현안 모니터링을 위한 다양한 민원 빅데이터의 활용 노력이 필요

- 현재 운영 중인 민원지도의 공간적 범위는 시군구 단위에 국한되고, 제공되는 정보도 주요 민원 발생지역과 특정 키워드에 집중되어 정보 전달력을 저해

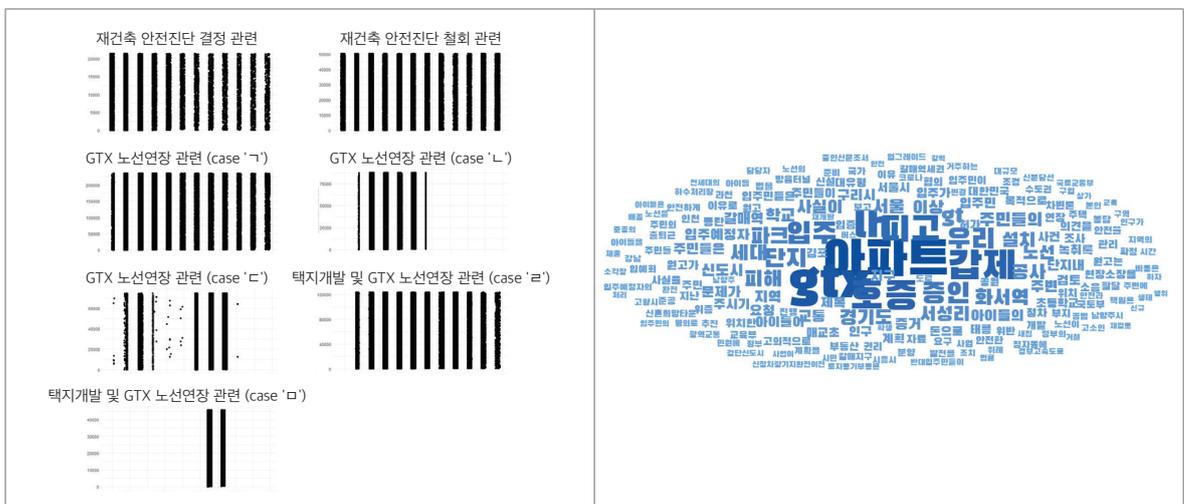
그림 3 민원 빅데이터의 분야-지역 간 버블차트



주: 가로축(x축)은 민원 빅데이터에서 정의하는 각 분야를 의미하고, 세로축(y축)은 전국 시도 단위 지역구분을 의미. 각 축이 교차하는 원의 크기가 클수록 관련 민원이 상대적으로 많음을 의미.

자료: 한눈에 보는 민원 빅데이터. <https://bigdata.epeople.go.kr> (2024년 8월 9일 검색).

그림 4 2021년 기준 최빈 반복 민원(7천여 건/연)의 시계열 빈도 분포(좌)와 민원내용 워드클라우드 예시(우)



주: 좌) 최빈 반복 민원 기준 7건에 대한 시계열 빈도 분포의 시각화; 우) 특정 인물, 지역, 비속어 등은 삭제 후 시각화함.

자료: 2023년 국토연구원-국민권익위원회 간 업무협약을 통해 한시적으로 제공받은 (개인정보 비식별화 처리된) 민원 빅데이터를 활용하여 저자 작성.

03. 민원지도를 통한 민생현안 모니터링 도구 시범구현

민원 빅데이터를 활용한 민생현안 모니터링을 효과적으로 추진하기 위한 민원지도를 시범 개발

2021년부터 2023년까지의 약 2,600만 건의 민원 빅데이터를 활용하여, 이종데이터 융·복합을 통한 민생현안 모니터링 시범분석 및 정책수요 발굴 시범분석 등을 제시

- ‘우리 지역 생활안전 중점 관리지역’(가제)에서는 민원 빅데이터로부터 불법주정차 민원 및 그 위치정보를 추출하고, 500m 격자(grid) 단위 취합 및 상위 20% 상습 불법주정차 민원과 소방·응급 시설 접근성 격자를 중첩하는 방식으로 안전 관심 필요지역을 도출
- ‘n-순위 키워드’(가제)에서는 지역별로 공통으로 상위권에 머무르고 있거나 보편적으로 익숙한 키워드 등을 제외하고, 차순위에 있는 키워드를 중심으로 지역별 현안 키워드를 도출
- ‘우리 지역 생활안전 중점 예상지역’(가제)에서는 포트홀, 불법현수막, 노상 적재물 등 국민의 안전과 직결된 정보를 AI. 모형이 학습하여 언제-어디에서-얼마나-어떻게 이슈가 발생할 것인가를 예측하는 모형을 제시

① ‘우리 지역 생활안전 중점 관리지역’(가제)을 통한 응급·의료 시설 접근성 관심지역 제시

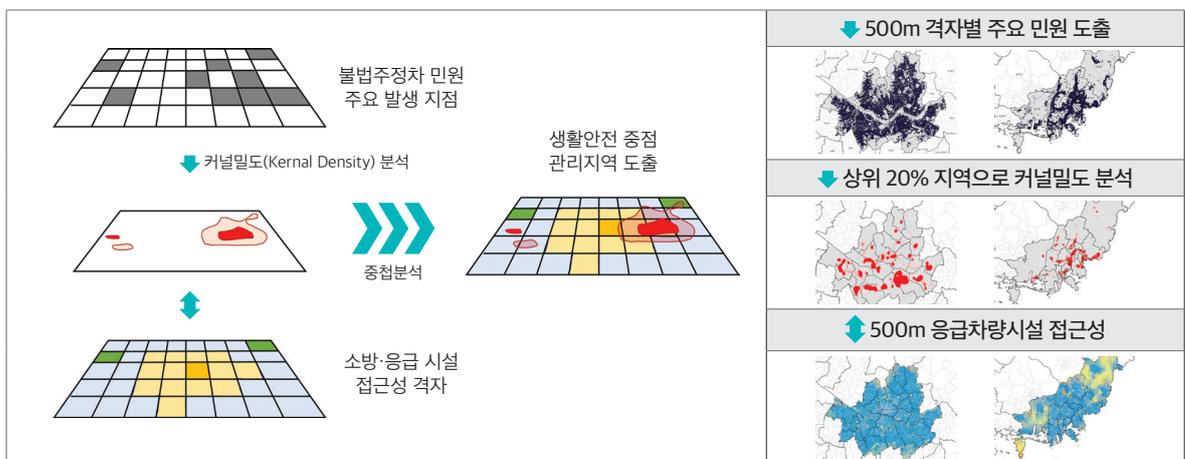
불법주정차와 관련한 이슈는 특정 지역의 문제가 아닌 사회 전반의 문제로, 불법주정차로 인한 주차공간 부족 문제뿐만 아니라, 소방 및 응급이송 차량 등 긴급차량의 이동에도 심각한 영향을 줄 수 있음

- 실제로 2015년 1월 의정부시 의정부동 도시형생활주택 화재와 2017년 2월 제천시 하소동 스포츠센터 화재 모두, 건축물 구조 문제와 함께 불법주차 차량으로 인한 소방차 진입 지연이 화재 진압 지연의 주요 원인으로 지목됨

본 연구에서는, 격자 단위로 집계한 불법주정차 민원 주요 지역과 ‘국토정보플랫폼’에서 제공하고 있는 소방시설과 응급시설의 격자별 접근성 정보를 융·복합하여, 불법주정차로 인한 생명권과 안전권의 위험정도를 분석

- 500m 격자 단위로 주요 불법주정차 민원지역을 추출하고, 격자별 집계값을 기준으로 상위 20% 격자로부터 커널밀도(Kernel density)¹⁾ 분석을 통해 주요 지역들을 도출한 뒤, 500m 격자 단위별 응급차량 접근성 격자를 중첩
- 시범분석을 수행한 서울과 부산 두 도시 모두 응급차량의 접근이 우수한 곳에서 중첩격자가 대다수 산출되었으나 부분적으로 접근성이 취약한 격자에서도 중첩격자가 산출되어, 해당 위치의 추가적인 관심이 필요
- 단순한 데이터 융·복합 설계를 통해 안전 사각지대 발굴에 효과적인 활용 가능성을 확인

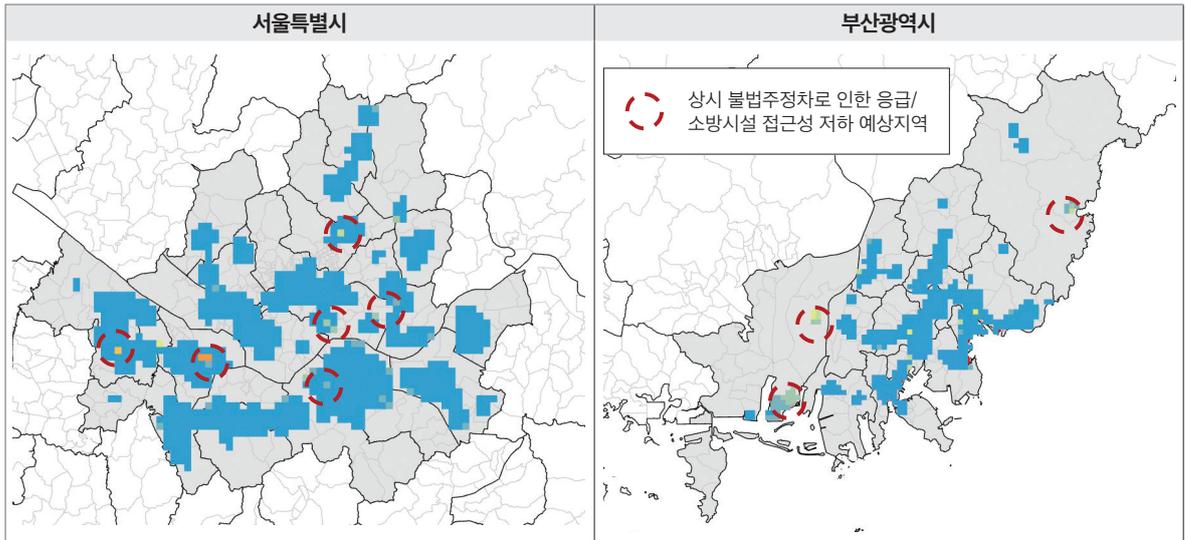
그림 5 ‘우리 지역 생활안전 중점 관리지역’(가제) 분석 개념도(좌)와 서울시범사례(우)와 부산지역 시범사례(우)



자료: (우) ‘국토정보플랫폼 국토정보맵’의 ‘응급의료시설 250m 격자’와 2023년 국토연구원-국민권익위원회 간 업무협약을 통해 한시적으로 제공받은 (개인정보 비식별화 처리된) 민원 빅데이터를 격자 변환 후 중첩하여 저자 작성.

1) 표본 간 밀도 분포를 바탕으로 전체 모수를 추정하는 방법론 중 하나임(https://en.wikipedia.org/wiki/Kernel_density_estimation, 2024년 11월 9일 검색).

그림 6 '우리 지역 생활안전 중점 관리지역'(가제) 서울 및 부산지역 시범분석 결과



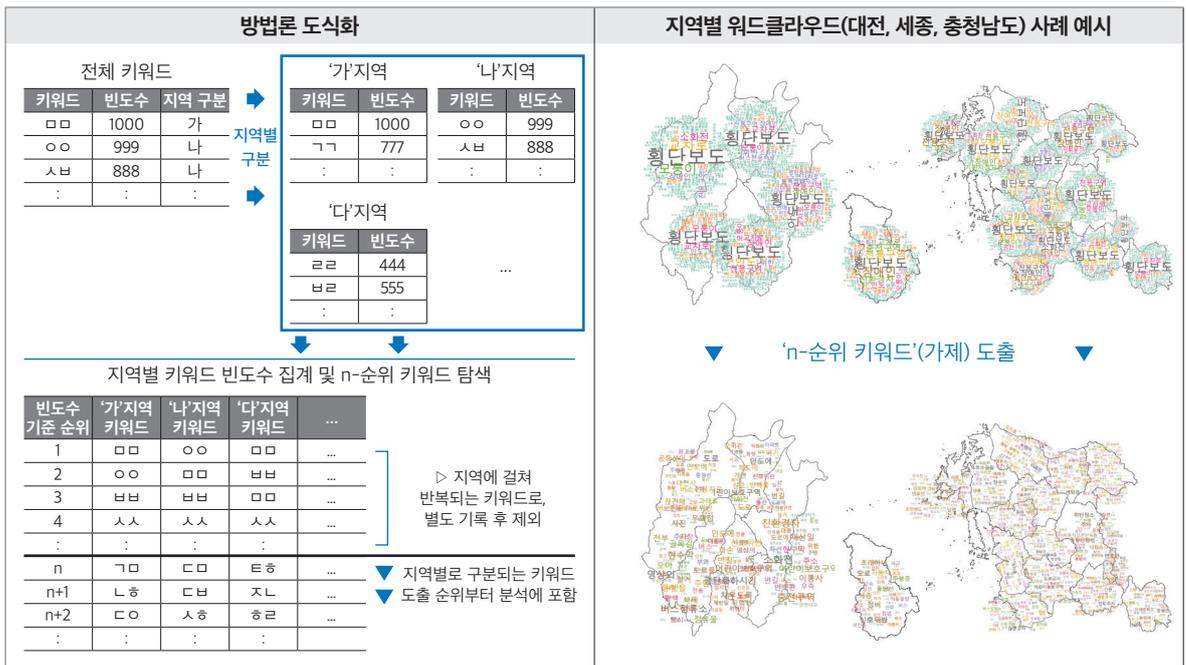
주: 푸른색 > 노란색 > 붉은색 순으로 접근성이 낮아짐(부산 해안선은 커널 보간법으로 붉은색이 도출).
 자료: '국토정보플랫폼 국토정보맵의 '응급의료시설 250m 격자'와 2023년 국토연구원-국민권익위원회 간 업무협약을 통해 한시적으로 제공받은 (개인정보 비식별화 처리된) 민원 빅데이터를 격자 변환 후 중첩하여 저자 작성.

② 'n-순위 키워드'(가제)를 통한 지역별 현안 도출

빈도수의 사소한 차이는 있으나, 지역에 걸쳐 동일하게 반복되는 키워드가 특정 순위까지 집중되는 민원 빅데이터의 특징을 활용한 전개방법

- 현재 운영 중인 민원지도의 워드클라우드에서는 수도권 및 교통 등 특정 지역과 분야의 민원이 지배적이어서 워드클라우드의 가독성이 상실되고 지역에 따른 민생현안 파악이 어려움
- 'n-순위 키워드'(가제)는 지역별 반복 키워드 제거를 통해 지역별 현안을 부각하는 방식
- 빈도수를 기준으로 키워드를 내림차순으로 정렬하고, 지역별로 다르게 취합되는 키워드의 n번째 순위부터 지역별 주요 키워드를 추출하는 방식

그림 7 'n-순위 키워드'(가제) 방법론 도식화 및 지역별 워드클라우드 사례 예시



주: 빈번하게 확인되는 단어일수록 진하고 큰 글씨로 확인
 자료: (우) 2023년 국토연구원-국민권익위원회 간 업무협약을 통해 한시적으로 제공받은 (개인정보 비식별화 처리된) 민원 빅데이터를 활용하여 dplyr, reshape2, data.table, showtext, ggwordcloud, wordcloud2, wordcloud, string, textclean, tidytext, KoNLP, tm, sf 등의 R오픈소스 라이브러리를 바탕으로 저자 작성.

③ '우리 지역 생활안전 중점 예상지역'(가제)

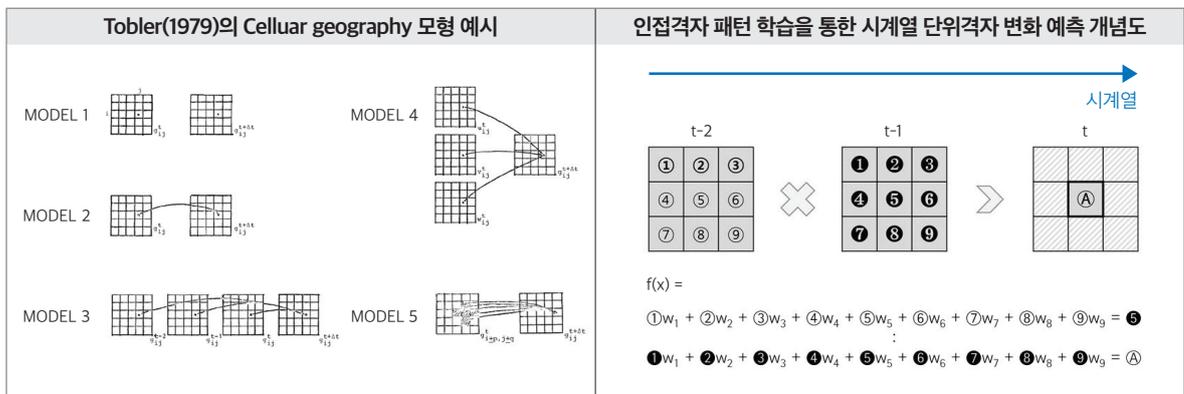
지역별 개인정보가 민감하지 않은 포트홀, 노상 적재물, (불법)현수막 등을 1km 격자 단위로 시계열 취합하고, 이를 A.I.에 학습시켜 언제-어디에서-얼마나-어떻게 이슈가 발생할 것인가를 예측하는 모형을 제시

- 시계열 변화는 시간과 공간의 독립적인 특성만으로는 이루어지기 쉽지 않으며, 시공간적인 인과관계 특성에 따라서 인접 공간 영향에 지배적인 변화가 유리한 것으로 설명
- '우리 지역 생활안전 중점 예상지역'(가제) 분석은, t-2 시기의 인접격자 패턴을 바탕으로 t-1 시기의 중심격자 패턴을 도출해 내고, 마찬가지로 t-1의 패턴을 바탕으로 t 시기의 격자 패턴을 도출해 내는 방식

본 연구의 A.I. 모형은 딥러닝 모형 중 격자와 시계열에 적합한 CNN과 LSTM을 혼합한 모형을 구현하여 활용

- 합성곱신경망(Convolutional Neural Network: CNN, 이하 CNN)과 장-단기-기억(Long-short-term memory: LSTM, 이하 LSTM) 모델을 혼합한 형태는 격자별 민원 발생량을 매트릭스로 간주하여 CNN으로 학습한 뒤, 이를 시계열 예측에 우수한 LSTM에 전달하여 시간과 공간의 특성 모두를 학습할 수 있도록 구현

그림 8 '우리 지역 생활안전 중점 예상지역'(가제) A.I. 학습 개념도

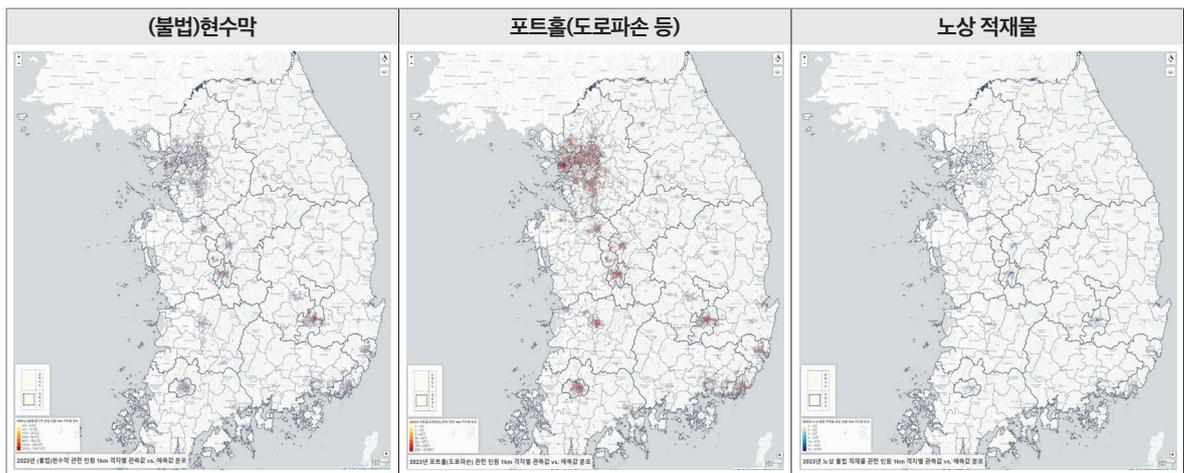


자료: (좌) Tobler 1979, 4-5.

'우리 지역 생활안전 중점 예상지역'(가제)은 격자별로 민감정보가 포함되지 않은 항목을 중심으로 가까운 미래에 발생할 것으로 예상되는 이슈를 시공간 통합방식을 통해 사전 예측 모형으로 제시

- 지역의 거주민뿐만 아니라, 대응의 1차 책임이 있는 행정 분야에도 유익한 예보가 될 수 있을 것으로 기대

그림 9 '우리 지역 생활안전 중점 예상지역'(가제) A.I. 학습 개념도



주: 1km 격자별로 연도별 취합하여 누적 합계가 높을수록 진하게 구성됨.

자료: 2023년 국토연구원-국민권익위원회 간 업무협약을 통해 한시적으로 제공받은 (개인정보 비식별화 처리된) 민원 빅데이터를 활용하여 plotly, sf, dplyr, reshape2, leaflet, leaflet.minicharts 등의 R오픈소스 라이브러리를 바탕으로 저자 작성.

04. 민생현안 모니터링 도구의 민원지도 개선방안

민원 빅데이터의 단순 집계 및 시각화를 넘어, 이종데이터 융·복합 등 지역맞춤형 민생현안 모니터링 제공 필요

‘한눈에 보는 민원 빅데이터’의 민원지도는 행정, 교통, 도시 등 다양한 분야에서 수집되는 다양한 민원을 시도 및 시군구 단위의 공간적 집계를 통해 지도화 서비스를 제공하고 있어, 직관적인 정책적 소통도구로 그 역할을 하고 있음

- 지역별로 어떠한 민원이 주를 이루는지, 어떠한 현안이 있는지 등을 가까운 시계열 비교를 통해 직관적으로 확인할 수 있다는 점은 정책실무자뿐만 아니라 일반 시민에게도 직관적인 정보를 제공할 수 있다는 강점
- 민원지도에서 제공하고 있는 정책소통방식은 다양한 정책현안에 대해서 대중의 의견을 끊임없이 수렴하고, 의견 반영을 위해 노력하고자 하는 정책적인 기여도와 그 효과가 매우 높은 도구라 할 수 있음

그럼에도, 민원을 바탕으로 현재 운영 중인 민원지도는 민원정보가 가지고 있는 빅데이터의 잠재력을 충분히 활용하지 못하고 있으며, 개인정보 보호와 데이터 공개의 균형, 데이터의 정확성과 신뢰성 확보 등의 문제 해결이 요구됨

- 민원의 다수 집계지역인 수도권의 목소리와 ‘교통’ 등 특정 분야에 집중된 민원 내용은, 정보의 편중성과 자칫 특정 지역의 목소리가 전국의 불편으로 간주되는 편향적 정보를 제공할 우려가 있음
- 사회문제 해결의 노력을 민원 자체에서만 발굴하기보다는 데이터의 시공간 융·복합을 통한 대리 지표 발굴이나 선제적 접근방법을 통한 해결이 필요

민원 빅데이터 공개 확대와 A.I. 분석 고도화를 통해 정보 효용성 제고 및 시민참여형 정책소통도구 발전 추진

(민원 빅데이터 공개 확대) 지자체와 부처별 협의를 통해 민원 빅데이터의 시공간적 공개 범위를 확대하고, 민원지도 내용의 선별적 공론화를 통한 시민 주도적 자정환경 조성 및 정책소통도구로 활용될 수 있는 기반 마련이 필요

(A.I. 기반 지역맞춤형 키워드 발굴과 민원지도 고도화로 정보 효용성 제고) ‘교통’ 등 특정 주제나 수도권 중심의 특정 지역에서 제기되는 민원이 전국의 민원으로 오인될 수 있는 현재의 민원지도 정보 왜곡 및 편향 현상을 방지하기 위해 A.I. 기술 기반의 지역맞춤형 키워드 발굴 방법론 및 민원지도 시각화 방법론 고도화를 통한 정보의 효용성 제고에 노력

(공간 단위 비식별화를 통한 통계 및 정보 활용도 제고) 포털, 노상 적재물 등 안전과 같이 개인정보 노출 우려가 적고 공공의 관심이 필수적으로 요구되는 민원 분야에 대해서는 민원 빅데이터의 격자 등 공간 단위 일원화 및 비식별화 취합을 통해 통계생산과 지도 기반의 정보공개 확대, 이종데이터 융·복합을 통한 활용도 제고에 이바지할 필요

(민원 자동정제 기능 구축으로 정보 왜곡 최소화 및 행정 효율 개선) 민원 빅데이터의 분석방법론 정립을 통해 악성 반복 민원에 대한 자동정제 기능을 구축하여, 민원지도에 표출되는 정보의 왜곡을 최소화하고 악성 민원 대응 등 불필요한 행정에 요구되는 인력을 효율적으로 관리할 수 있는 제도적 개선 및 시스템 고도화

참고문헌 Tobler, Waldo R. 1979. *Philosophy in geography in Cellular geography*: 379-386.

- **장요한** 국토연구원 국토인프라-공간정보연구본부 부연구위원(ycanns@krihs.re.kr, 044-960-0406)
- **손재선** 국토연구원 국토인프라-공간정보연구본부 부연구위원(jsson@krihs.re.kr, 044-960-0418)
- **최혜림** 국토연구원 국토인프라-공간정보연구본부 연구원(chr1142@krihs.re.kr, 044-960-0186)

※ 이 브리핑은 “장요한, 손재선, 최혜림, 2024. 민생현안 모니터링을 위한 민원지도 개선방안 연구. 세종: 국토연구원” 보고서를 요약 정리한 것임.

※ 이 브리핑은 연구자 개인의 의견으로서, 정부나 국토연구원의 공식적인 견해와 다를 수 있음.