

집단민원의 감성분석을 이용한 공간빅데이터 시각화 방안*

주용진¹*

A Study on the Visualization of Geospatial Big Data using Sentiment Analysis of Collective Civil Complaints*

Yong-Jin JOO¹*

요 약

전통적으로 공공 서비스에 대한 만족도 요인을 측정하기 위해 설문조사나 인터뷰 연구가 주를 이뤄 왔다. 민원의 단순 빈도를 떠나 민원에 내포된 감정의 경중까지 고려되지 않아 민원인이 체감하는 민원의 시급성, 고충의 심각 정도를 판단하기 어렵다. 이에 본 연구의 목적은 헤도노미터 단어별 행복도 점수를 활용해 집단민원이 내포하는 부정적 감성수치를 산정하는 방안을 제시하였다. 국민권익위원회의 2021년 지역별 상위 민원 토픽과 연관키워드 데이터를 대상으로 헤도노미터를 적용하여 민원의 주제별 부정적 감성수치를 산출하고, 지역별로 분포를 가시화하였다. 본 연구결과로 도출된 부정적 감성수치를 이용해 민원에 내포된 감정의 경중을 고려하여 민원인이 체감하는 민원의 시급성, 고충의 심각 정도를 판단하는데 도움이 될 수 있을 것으로 기대된다.

주요어 : 집단민원, 감성분석, 부정적 감성수치, 빅데이터

ABSTRACT

Traditionally, surveys or interview studies have been used to measure satisfaction factors for public services. This method focuses on the simple frequency of civil complaints and does not consider the aggravation of emotions implied in civil complaints. As a result, it is difficult to judge the urgency of civil complaints and the severity of grievances experienced by civil petitioners. This study aims to calculate the negative emotional value of collective complaints by using the happiness score for each word on the Hedonometer. The Anti-Corruption and Civil Rights Commission applied a Hedonometer to the top civil complaint topics and related keyword data by region in

2022년 12월 11일 접수 Received on December 11, 2022 / 2023년 01월 09일 수정 Revised on January 09, 2023 / 2023년 01월 12일 심사완료 Accepted on January 12, 2023

* 이 논문은 2022년도 인하공업전문대학 학술연구사업 지원에 의하여 연구되었음

1 인하공업전문대학 공간정보빅데이터과 교수 Professor, Dept. of Geospatial Big Data, Inha Technical College

※ Corresponding Author E-mail: jjy@inhac.ac.kr

2021 to calculate negative sentiment values by subject of civil complaints, and visualize the distribution by region. Using the negative emotional values derived from the results of this study, the severity of emotions contained in civil complaints can be considered. It is also expected to be helpful in determining the urgency of civil complaints and the severity of grievances experienced by civil petitioners.

KEYWORDS : *Collective Complaints, Sentiment Analysis, Blue Score, Big Data*

서론

1. 연구 배경 및 목적

코로나19 팬데믹 이후 비대면 수요가 증가하고 일상에서도 재택근무, 비대면 수업 등과 같이 디지털 중심사회로의 전환이 가속화되고 있다. 전통적으로 공공 서비스에 대한 만족도 요인을 측정하기 위해 설문조사나 인터뷰 연구가 주를 이뤄 왔다. 최근 몇 년간 국민신문고 및 지자체의 개별 온라인 게시판을 통해 접수된 민원 빅데이터를 이용해 민원 추이, 기관별 민원 현황 등 민원정보의 모니터링과 사회적 이슈 분석에 활용되고 있다. 현행 민원 분석에 있어 키워드 도출과 상위키워드 분석 등은 민원이 제기된 발생빈도를 의미한다. 하지만, 민원은 공공 재화나 서비스에 대한 일반 시민이 느끼는 심리적·감성적 스트레스라고 볼 수 있다. 하지만 민원의 단순 빈도를 떠나 민원에 내포된 감정의 경중까지 고려되지 않아 민원인이 체감하는 민원의 시급성, 고충의 심각 정도를 판단하기 어렵다. 디지털 시대에 표정과 같이 사용자의 만족도 등 감정을 읽어낼 수 있는 가장 직관적인 방법 중 하나가 텍스트에 대한 감성분석이다. 본 연구는 민원이 고충과 불만족이 내포된 개선 요청을 목적으로 작성되는 글임에 착안한다. 또한, 민원 분석에 있어 단순 빈도를 떠나 전체 민원 내용에서 해당 특정 키워드가 갖는 감성적 가중치를 측정하여 활용한다. 이에 본 연구의 목적은 헤도노미터 단어별 행복도 점수를 활용해 집단민원이 내포하는 부정적 감성수치(Blue Score)를 산정하는 방안을 제시하는 것이다. 본 연구에서 부정적 감성수치는 Blue 단어가 우울,

가라앉은 기분, 불안 등 부정적인 감성 색채를 가지고 있으며 Score(수치)와 조합된 감성점수를 의미한다. 이를 위해 우선, 단어별 행복도 점수를 도입한 국내의 감성분석 사례를 조사한다. 둘째, 국민권익위원회(<https://bigdata.epeople.go.kr/>)의 2021년 지역별(3개 시·도) 상위 민원 토픽과 연관키워드의 크롤링을 통해 집단민원 데이터를 수집한다. 셋째, 토픽별 연관 키워드의 형태소를 분리하고 헤도노미터 단어 사전에 매칭을 위한 데이터 전처리를 수행한다. 파이썬 Konlpy 라이브러리를 이용하여 집단민원 토픽과 연관키워드의 형태소를 분리하여 헤도노미터에 매칭시켜 점수를 부여한다. 이때 헤도노미터에 매칭되는 단어가 없어 점수 배점이 불가능한 경우 한글 유의어 또는 영단어로 매칭이 되도록 단어 사전을 추가 보완한다. 주요 민원 토픽과 연관 키워드를 대상으로 부정적 감성수치의 분포를 100점 만점으로 산출하여 정규화한다. 분리된 형태소의 헤도노미터를 적용하여 민원 토픽별 부정적 감성수치를 산출한다. 마지막으로 민원 토픽별 부정적 감성수치를 지역별로 가시화하여 민원의 지역별 분포와 발생 현황을 분석한다. mapbox 파이썬 지도라이브러리를 이용해 민원 토픽별 부정적 감성수치를 지역별 단계구분도로 가시화하여 민원의 지역별 분포 차이와 시사점을 도출한다.

2. 선행연구의 비판적 고찰

불만은 특정 상황에서 현실과 기대 사이의 부정적인 불일치를 표현하기 위해 인간과 컴퓨터를 매개로 한 의사소통에서 정기적으로 사용되는 기본 언어 행위로 소셜 미디어에서 불만을 자동으로 식별하는 것은 조직이나 브랜드가 고

객 경험을 개선하거나 불만 처리 및 응답을 위한 대화 시스템을 개발하는 데 가장 중요하다(Daniel *et al.*, 2019). 민원 분석 관련 대표적인 국외 연구사례로 Zhou and Ganesan(2016)는 Yelp 리뷰를 연구하는 전산 방법에 대한 불만으로 정의된 개념을 제시했다. Mikolov *et al.*(2018)은 소비자 금융 보호국에 제출된 금융 불만이 적시에 답변을 받을 수 있는지를 분석했다. 또한 Yang *et al.*(2019)는 고객 지원 대화를 연구하고 이러한 불만이 정부 기관으로 확대되거나 소셜 미디어에 공개될 것인지 예측했다. 국내 민원 빅데이터와 관련된 연구로 Hong(2018)은 환경분야의 민원 데이터의 텍스트마이닝을 통한 분석 방안을 제시하고 Hwang(2019)은 소음에 관한 민원을 감성적 측정 방법을 도입하였다. 최근 데이터 시각화의 가장 큰 잠재력은 실시간으로 상호작용할 수 있는 동적인 그래픽 도구의 발전에 의한 데이터 시각화라고 할 수 있다(Friendly, 2006). 데이터 시각화(data visualization)는 데이터를 시각적으로 표현하는 과학으로, 정보의 속성 혹은 변수 등을 포함하여 개략적인 형태의 추상화된 정보(Friendly, 2009)로 정의되고 있다. Kim(2018)은 공간 빅데이터 동태적 시각화 모형을 제시하고 Jeon and Shin(2018)은 행위자 기반 공간 빅데이터 융합 시각화 연구를 수행했다. 헤도노미터는 트위터 데이터를 이용하여 사용자들의 행복도를 측정하는 도구이다(Jin *et al.*, 2020). 2011년 미국 버몬트(Vermont) 대학 연구팀에서는 트위터 데이터를 이용해 사용자들의 행복도를 측정할 수 있는 ‘헤도노미터’라는 분석틀을 개발하였다. 헤도노미터 분석 틀은 매일 5,000만 개의 트윗 데이터를 추출하여 기 산정된 단어 매칭을 통해 행복 점수를 산정하게 되며 구글북스(Google Books) 데이터, 1987-2007년까지의 뉴욕타임스(Newyork Times) 기사 등 가장 빈번하게 사용되는 상위 5,000개의 단어들을 조합하여 최종적으로 10,222개의 단어를 추출한다. 이 단어들은 클라우드 소싱 방식을 통해 인터넷 사용자들이 직접 개별 단어에 대한 행복 점수를 매긴 뒤 통계 처리하여 최

종 점수를 산정하게 된다. 모든 단어의 점수는 행복도가 가장 낮은 1점부터 가장 높은 9점 사이에 존재하게 되며 점수가 낮을수록 행복도가 낮음을 뜻한다. Korea Environment Institute(2014)는 빅데이터를 활용한 환경 분야 정책수요 분석 방안을 제시했다. 이는 소셜 미디어 사용자들이 하루에도 수십 번씩 자신의 생각이나 감성이 담긴 글을 온라인에 게시하고 있다는 점을 이용하여 국민들이 느끼는 행복의 정도를 정량화하여 향후 환경 분야에서 응용할 수 있는 방안을 제안하였다. 헤도노미터의 사례는 시민의 행복이라는 사회적 혹은 철학적 요인을 트위터, 구글북스, 뉴욕타임스 등에서 추출한 빅데이터를 활용하여 분석하였다 본 연구에서 민원은 ‘사회 현안문제’로 공공에게 개선을 요구하는 다양한 정보를 담고 있어 사회의 수으로 인지되어, 정책수립 또는 연구자료 등으로 빈번히 활용되는 활용가치가 높은 데이터라는 점에 주목한다. 또한 국내외 연구사례를 통해 각 민원이 가지고 있는 토픽의 감성적 수치화 방안과 위치 정보의 시각화에 대한 연구가 미비함에 착안한다. 즉, 민원의 접수량으로 사안의 심각성을 평가하던 기존 개념에서 탈피하여, BlueScore 라는 부정적 감성분석을 통해 민원 고충의 심각성과 처리 시급성을 직관적으로 파악 가능한 정보의 새로운 해석 방법을 고안한다. 또한 민원 빅데이터를 공간상에 시각화하여 민원이라는 사회 현안의 공간적 패턴을 지역적으로 분석할 수 있는 방안을 제시하고자 한다.

헤도노미터기반 민원 데이터의 감성분석

1. 민원 데이터 수집 및 전처리

1) 데이터 스크래핑

국민권익위원회 「한 눈에 보는 민원빅데이터」 웹페이지이 지역별 키워드 TOP 5까지 스크래핑하였다. 데이터 정렬과 저장 후 파이썬 판다스를 이용해 민원내용을 데이터프레임으로 변환한다. 감성분석을 위하여 연관키워드는 Python

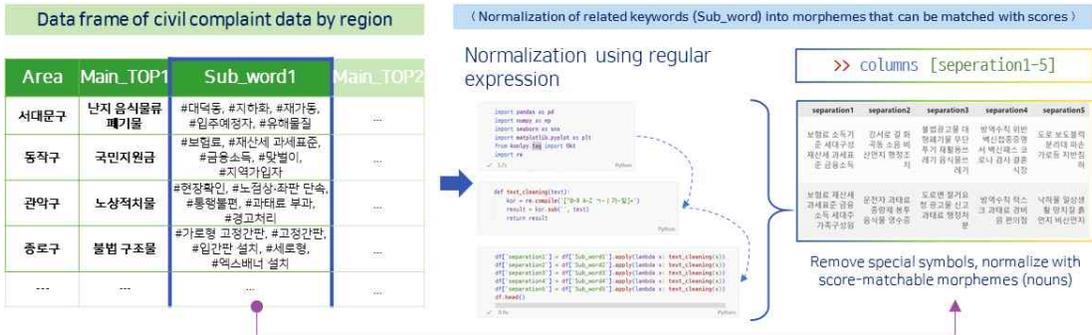


FIGURE 1. Data preprocessing: Morphological separation

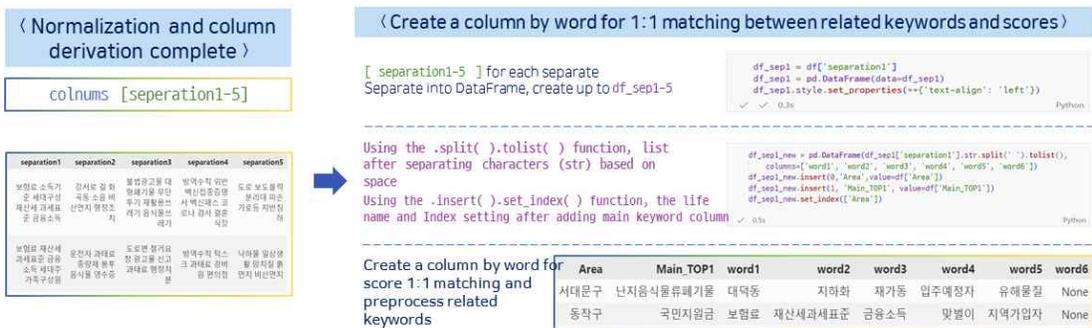


FIGURE 2. All related keyword pre-processing: Separation of columns by keyword

Konlpy로 형태소 분리·불용어 처리 후, 헤도노미터에 있는 키워드와 매치시킬 수 있도록 별도의 컬럼을 생성하여, 요구하는 행위가 되는 키워드 어미의 ‘명사’ (예: 반대, 단속, 폐쇄..., 요청 등은 불용어 처리)만을 남겨, 헤도노미터에 있는 키워드와 매치시킬 수 있도록 별도의 컬럼을 생성하고 값을 입력한다.

2) 형태소 분리

점수를 매치시킬 연관키워드 항목(Sub_word) 내 형태소 정규화, 명사화 등 데이터 전처리를 실시하고 전처리된 내용들을 별도의 컬럼으로 생성한다.

정규화 되어 키워드가 묻쳐있는 컬럼들의 경우 각각의 키워드별 점수 매칭을 위하여 키워드 별로 컬럼을 분리시켜 모든 연관키워드의 전처리를 한다.

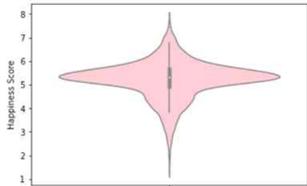
2. 헤도노미터 기반 블루 스코어 부여

1) 블루 스코어 산정 방안

헤도노미터는 일상생활 속에서 가볍게 나누는 SNS인 트위터 어구 기반으로 만들어진 데이터이다. 전처리된 컬럼들을 헤도노미터 키워드와 매치시켜 점수를 부여하기 전, 헤도노미터의 점수 체계를 블루 스코어에 맞게 재편성하는 과정이 필요하다. 즉 헤도노미터는 긍정성을 기반으로 한 단어 감성 점수를 부여하는 반면 본 연구에서의 블루 스코어는 부정성을 강조한 감성점수이기 때문이다. 헤도노미터는 점수는 max=7.84, min=1.34, mean=5.23 형태로 분포해 있으며, 아래 그림1과 같이 주요 감정지수가 모두 5-6점대에 분포하여 max=7.84임을 감안할 때, 긍정적인 어구가 많은 것으로 분석 하였다.

Blue Score를 산출하기 위해서는 아래의 산식을 작성 및 적용하여 점수를 도출한다. max=

TABLE 1. Hedonometer score distribution configuration

main_sensibility	min_Score	mean	max_Score	score distribution
positivity (HappyScore)	1.34	5.23	7.84	

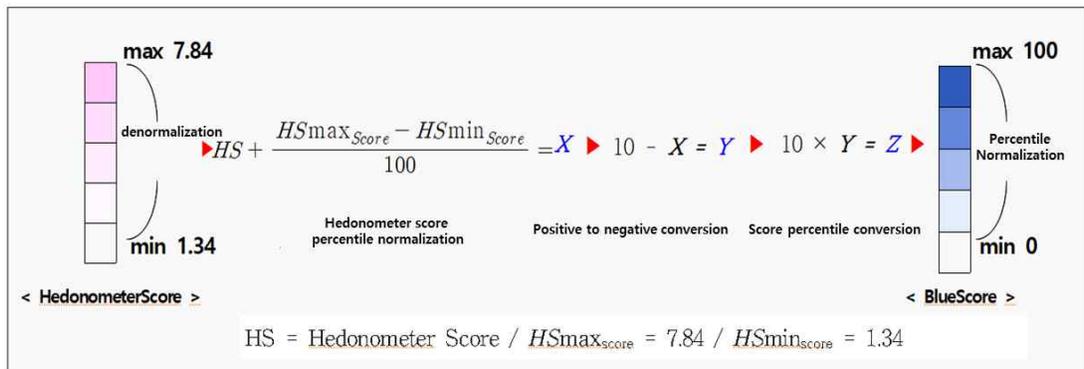


FIGURE 3. Hedonometer-based Blue Score score normalization process

7.84이 100점 만점, min=1.34이 0점이 되도록 백분화를 실시하고 헤도노미터에서 낮게 배점한 부정적 감정 어구들이 높은 점수가 되도록 한다. 또한, 기존 1의 자리에서 끝나는 최대값은 100점이 최대가 되도록 정규화 실시한다.

- 2) 사용자 정의 사전 구축
본 연구 내 사용자 정의 사전은 Blue Score

산정 단어용 단어·점수목록을 의미한다. 헤도노미터는 일상생활 속에서 가볍게 나누는 SNS인 트위터 어구 기반으로 만들어진 데이터이기 때문에, 민원처럼 사무적인 어휘들이 자주 오가는 문서에서 기존의 데이터로는 매치되지 않을 수 있다. 동일 단어 매칭이 되지 않아 점수 배점이 불가능한 경우 유의어와 영단어를 이용하여 보완하였다. 즉, 단어를 네이버 국어사전 유

TABLE 2. Word example of hadonometer happiness scores before normalization

word	Hedonometer	word	Hedonometer	word	Hedonometer	word	Hedonometer
반대	4.04	신설	6.12	철저	6.22	쓰레기	3.14
단속	5.47	설치	5.44	점검	4.89	폐해	3.68
신고	4.22	추가	5.78	폐해	4.65	점검	4.96
불법	3.22	방역	6.22	요구	4.98	추진	6
폐쇄	2.04	수칙	5.74	철회	4.14	방음	3.74
위반	3.48	불편	3.66	신고	4.22	하수	2.16
이전	5.1	검사	4.7	위반	3.48	섭취	4.98
착공	6	코로나	1.6	중단	4.32	금지	3.72
축구	4.86	문의	5.28	반입	5.26	직결	5.38
연장	5.7	횡포	2.06	섭취	6.07	재건축	5.34

TABLE 3. Exapmle of normalized result of Blue Score

word	Hedonometer	percentile deviation (X)	10 - X	normalization
반대	4.04	4.105	5.895	58.95
단속	5.47	5.535	4.465	44.65
신고	4.22	4.285	5.715	57.15
불법	3.22	3.285	6.715	67.15
폐쇄	2.04	2.105	7.895	78.95
위반	3.48	3.545	6.455	64.55
이전	5.1	5.165	4.835	48.35
착공	6	6.065	3.935	39.35
축구	4.86	4.925	5.075	50.75

의어를 헤도노미터에 검색하여 유의어의 배점을 적용한다. 같은 의미의 유의어가 2개 이상 헤도노미터에 있는 경우, 해당 단어들을 모두 취합 후 평균값을 적용하였다. 또한, 한글유의어로 매

칭되는 유의어가 없을 때, 해당 단어와 동일 행에 위치한 영단어를 헤도노미터 eng_ver에 검색하여 값을 지정하였다.

표 3은 헤도노미터 보완용 사용자정의 감성

TABLE 4. Ranking change comparison (Blue Score left = not applied, right = applied)

Anti-Corruption and Civil Rights Commission No. 1 keyword				Civil complaint ranking based on Blue Score			
Region (1st place)	Civil petition keyword	Blue Score		rank	Civil petition keyword	Blue Score	region
		score	rank				
Seodaemun-gu	난지 음식물류폐기물	73.73	1	1	난지 음식물류폐기물	73.73	Seodaemun-gu
Jongno-gu	불법 구조물	67.15	4	2	노점상 단속	69.67	Jongno-gu
Gangbuk-gu	쓰레기 무단투기	65.98	5	3	마스크 미착용	67.61	Jongno-gu
Gwanak-gu	노상적치물	64.85	9	4	불법 구조물	67.15	Jongno-gu
Gangnam-gu	의료법 위반	63.17	15	5	쓰레기 무단투기	65.98	Gangbuk-gu
Gangdong-gu	아파트 재건축	56.5	54	6	수거 요청	65.9	Yongsan-gu
Seocho-gu	국민지원금	45.51	97	7	쓰레기 무단투기	64.98	Yongsan-gu
Songpa-gu	국민지원금	45.51	97	8	방역	64.91	Gangnam-gu
Jung-gu	국민지원금 이의신청	44.85	101	9	노상적치물	64.85	Gwanak-gu
Seongbuk-gu	국민지원금 이의신청	44.77	102	10	쓰레기 무단투기	64.6	Dobong-gu
Seongdong-gu	국민지원금 이의신청	44.65	103	11	공사장 소음	64.09	Gangnam-gu
Dongdaemun-gu	국민지원금 이의신청	44.55	104	12	마스크 미착용	63.83	Seongbuk-gu
Dobong-gu	국민지원금 이의신청	44.49	105	13	쓰레기 무단투기	63.65	Mapo-gu
Nowon-gu	국민지원금 이의신청	44.31	106	14	쓰레기 무단투기	63.35	Jung-gu
Dongjak-gu	국민지원금	44.21	107	15	의료법 위반	63.17	Gangnam-gu
Jungnang-gu	국민지원금 이의신청	44.21	107	16	은평광역자원순환센터 결사반대	62.22	Eunpyeong-gu
Geumcheon-gu	국민지원금	44.19	109	17	공사현장	62.08	Gangseo-gu
Yeongdeungpo-gu	국민지원금	43.55	112	18	쓰레기 무단투기	62.07	Gwanak-gu
Eunpyeong-gu	국민지원금 이의신청	43.53	113	19	수거	62.05	Gangseo-gu
Guro-gu	국민지원금	43.51	114	20	보수정비 점검	61.81	Dongdaemun-gu
Yongsan-gu	국민지원금 이의신청	43.23	115	21	공사장 소음	61.74	Gangdong-gu
Gwangjin-gu	국민지원금 이의신청	42.75	116	22	쓰레기 무단투기	61.55	Seongbuk-gu
Yangcheon-gu	국민지원금 이의신청	42.69	117	23	쓰레기 무단투기	61.52	Gangnam-gu
Mapo-gu	국민지원금 이의신청	41.67	121	24	견인지역	61.05	Guro-gu
Gangseo-gu	국민지원금	41.51	122	25	쓰레기 무단투기	61.03	Gwangjin-gu

사전에 적용 및 Blue Score의 정규화 결과값을 확인 할 수 있다.

3) 연관키워드의 Blue Score 적용 결과

앞서 도출한 정규화된 해도노미터를 기반으로 한 Blue Score를 컬럼에 매치 및 합산 후, 컬럼 내 단어의 수만큼로 나눠 평균값으로 해당 민원키워드별 감성점수를 아래 표와 같이 도출할 수 있다. 25개 자치구 중 19개 자치구의 민원키워드 1위는 국민지원금 관련 키워드였으나 Blue score 적용결과, 총 7가지의 1위 민원키워드 중 국민지원금 관련 민원이 가장 낮은 순위로 나타났다. 오히려 ‘난지음식물류폐기물’, ‘불법구조물’, ‘쓰레기 무단투기’ 등 생활 불편과 직접 관련된 사항이 높은 우선순위를 보였다.

민원 감성분석의 공간시각화

1. 민원 감성수치의 결과분석

민원 키워드와 Blue Score의 상관관계를 직관적으로 이해하고자 국민권익위 2021 서울시 자치구별 민원키워드 TOP1~5(125개)를 민원이 주로 처리되는 법령 및 처리부서 명칭 등을 참고하여 민원을 표 5와 같이 분류하였다.

1) 감각·감성 자극 키워드

후각, 청각 등을 자극하는 키워드에 수치가 높게 나타났다. 특히 자치구 25개 구 중 17개 구에서 TOP5 이내에 쓰레기 무단투기와 관련된 키워드가 들어가 있으며, 이는 개인이 가진 쓰레기와 관련된 악취를 연상시켜 부정적 감정을

을 유발한 것이라고 유추할 수 있으며, 비슷한 예로 공사장의 소음이 청각을 자극한다는 면에서 상기 예와 같은 맥락으로 설명할 수 있다. 또한 마스크 미착용도 질병감염에 대한 불안정서를 자극하는 요소로서 상위권에 많이 포진하였다.

2) 주거지 인근 생활 밀접형 키워드

불법주정차 관련 민원과, 도보 등 공공시설물의 보수요청 민원, 난지 음식물류폐기물 시설 관련 민원, 유일하게 키워드 중에서 결사반대라는 키워드가 등장한 은평구 광역자원순환센터 결사반대 민원과 같이 개인의 생활 환경 내, 주거지·직장 등 장기체류하는 공간 주변에서 발생하는 민원일수록, 물리적 상호작용이 일어날 수 있는 거리로 좁혀질 수록 Blue Score의 수치가 높은 것을 알 수 있다. 특히, 쓰레기와 폐기물 처리시설은 악취로 인해 감각적인 공해로 인지함과 동시에 주거지 가치 하락 등 정서적 불쾌감을 동시에 유발할 수 있어, 이번 분석 중 어떤 키워드들 보다도 높은 Blue Score를 보여주고 있어 민원의 고충의 강도를 언급량이 아닌 감각·감성적 접근을 통하여 접근하는 방식이 유효하다고 가정해볼 수 있는 단서를 제공한다.

3) 언급량과 고충의 상관관계

기존의 25개 자치구 중 19개 구(76%)의 1위 민원은 국민지원금과 이에 대한 이의신청 민원이다. 그러나 Blue Score 기준 국민지원금의 순위는 120위 밑을 웃돌고 있어 자원순환센터, 난지 음식물류폐기물 처리시설보다 불쾌감을 덜

TABLE 5. 국민권익위 2021 서울시 자치구 민원키워드 대분류표

rank	Classification of complaints	ratio	major complaints
1	living environment	28.0%	Litter dumping, construction noise
2	health care	21.6%	not wearing a mask
3	self-government	18.4%	Objection to National Subsidies
4	urban landscape maintenance	14.4%	Illegal banners, items left on the street, repair requests for public facilities
5	city traffic	9.6%	illegal parking
6	architectural housing	7.2%	New apartment construction quality
7	park green space	0.8%	Long-term unused green space management request

주는 키워드로, 부정적 감성수치 및 고충의 정도가 상대적으로 낮다고 볼 수 있다. 즉, 언급량과 부정적 감성수치가 정비례관계가 아님을 알 수 있다. 따라서, 언급량이 많아 빅데이터 분석상 상위순위권으로 진입한다는 것이 반드시 하위순위의 민원에 비해 고충이 중하다고 할 수 없다는 것을 의미한다. 언급량은 관심의 척도로서 이해할 때 유의미하다고 볼 수 있다.

2. 감성분석의 지도가시화

Blue Score와 같이 텍스트의 감성을 분석하고 수치화하여 지도 위에 점수의 편차에 따라 원형크기·색상의 진하기에 차이를 두어, 특정 주제의 민원이 많이 몰리는 공간적분포를 확인할 수 있으며, 더 나아가 발생지 주변의 공간적 특성을 확인하여 민원예상 발생지도 유추가 가능하다. 행정동 기반의 지도 시각화를 위해 대한민국 행정동 경계 json 파일에서 서울시 지역만 추출하여 저장한다. 속성값을 결합하여 폴리곤 형태의 geojson 파일을 생성한다. 최종 geojson 파일은 mapbox 지도 라이브러리를 사용하여 서울시 민원의 웹 기반 지도 가시화를 한다. 자치구별 Blue Score 1위 민원키워드 단계구분도를 제작하였으며 아래 그림은 주거지 인근 생활 밀접형 키워드 중 은평구 광역자원순

환센터 결사반대에 대한 매시업 표시 결과를 나타낸다.

map box의 ChoroplethViz() 함수를 이용해 자치구별 Blue Score 1위 민원키워드 단계구분도를 작성하였다. 자치구별 가장 Blue Score (부정적감성수치)가 높은 민원을 순위별로 조희가 가능(구별 1위~5위)하다. 공간에 민원정보를 표시, 민원 군집위치 확인 및 해당 민원문의 텍스트마이닝 및 Blue Score 적용을 통해 고충의 경중도(=Blue Score) point 진하기, 크기로 직관적 가시화가 가능하다. 또한 클러스터의 크기 및 색상의 진하기로 부정적 감성수치가 높은 공간을 찾을 수 있으며, 민원 대분류 및 소분류별 색상을 달리할 경우, 보다 직관적인 가시화가 가능하다.

결론

본 연구에서는 헤도노미터 단어별 행복도 점수를 활용해 집단민원이 내포하는 부정적 감성수치를 산정하는 방안을 제시하였다. 국민권익위원회의 2021년 지역별 상위 민원 토픽과 연관키워드 데이터를 대상으로 헤도노미터를 적용하여 민원 주제별 부정적 감성수치를 산출하고, 지역별로 분포를 가시화하였다. 본 연구결과로

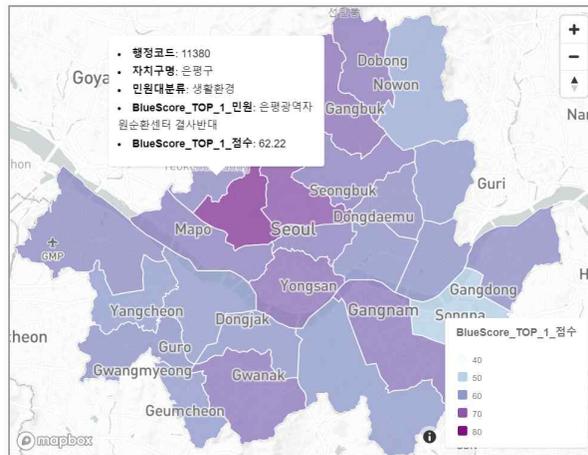


FIGURE 4. Choropleth map of civil petition keyword by district in Blue Score 1st place



FIGURE 5. Clustering map of civil petition keyword by district in Blue Score 1st place

도출된 부정적 감성수치를 이용해 민원에 내포된 감정의 경중을 고려하여 민원인이 체감하는 민원의 시급성, 고충의 심각 정도를 판단하는데 도움이 될 수 있을 것으로 기대된다. 향후, 위치 정보와 결합된 민원 데이터가 수집되어 관리된다면 민원인이 현실과 동일한 위치에 직접 민원 내용을 작성할 수 있어 ‘근처’, ‘옆에’, ‘우측에서 10m 정도’와 같은 텍스트보다 명확히 민원 접수가 가능할 것이다. 민원처리기관 입장에서 민원에 대한 명확한 위치표시와 더불어 민원 내용에 대한 텍스트마이닝 및 감성분석이 적용되므로, 전반적인 고충의 경중 이해와 문제의 시급성을 검토하는데 유용하게 사용될 수 있을 것이다. 결론적으로 본 연구는 민원의 접수량으로 사안의 심각성을 평가하던 기존 개념에서 탈피하여, BlueScore 라는 부정적 감성분석을 통해 민원 고충의 심각성과 처리 시급성을 직관적으로 파악 가능한 정보의 새로운 해석 방법이다. 또한 민원 빅데이터를 공간상에 시각화하여

민원이라는 사회 현안의 공간적 패턴을 지역적으로 분석할 수 있는 새로운 접근 방안이라는 데에 의의가 있다. 장기간 데이터가 쌓이게 되면 통계적으로 관내 민원의 발생 분포의 분석이 가능하고, 감성별 공간분석을 통해 감성의 정도와 환경·범죄·상권 등의 카테고리를 결합해 통계적 예측이 가능할 것이다. **KAGIS**

REFERENCES

- Daniel, P.P., G. Mihaela and A. Nikolaos, 2019, Automatically identifying complaints in social media, Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, July 28 – August 2, pp.5008-5019.
- Friendly, M. 2006. A brief history of data visualization. In: C. Chen et al.(ed.). Handbook

- of Data Visualization. Springer Handbooks Computational Statistics. Springer, Berlin, Heidelberg, pp.15-56.
- Friendly, M. 2009. Milestones in the history of thematic cartography, statistical graphics, and data visualization. http://www.math.usu.edu/~symanzik/teaching/2009_stat6560/Downloads/Friendly_milestone.pdf
- Hong, S.E. 2018. Sentimental & pattern analysis of environment complaint by big data mining. Inha University. Korea. 89pp (홍성은, 빅데이터 마이닝에 의한 환경민원의 감성분석 및 공간패턴분석. 2018. 인하대학교 석사학위논문, 89쪽).
- Hwang, S.W. 2019. A study on the influence of happiness score on the movement of residence in the noise complaints: Focused on the complaint board in 25 districts of Seoul. Yonsei University. Korea. 75pp (황성욱, 소음 민원에 나타난 행복 점수가 거주지 이동에 미치는 영향 연구 - 서울시 25개 자치구 새울민원게시판 중심으로. 2019. 연세대학교 석사학위논문, 75쪽).
- Jeon, S. E and D. B. Shin, 2018, A study on the agent based infection prediction model using space big data -focusing on MERS-CoV incident in Seoul. Journal of the Korean Association of Geographic Information Studies 21(2):94-106 (전상은, 신동빈, 2018, 공간 빅데이터를 활용한 행위자 기반 전염병 확산 예측 모형 구축에 관한 연구 -서울특별시 메르스 사태를 중심으로, 한국지리정보학회지 21(2):94-106).
- Jin, H. I., G.Y. Ock and M.s. Lee, 2020, A study on a methodology to measure happiness index with social big data: Focusing on Google Trends. Korean Journal of Korea society of innovation 15(5):215-237 (진형익, 옥가운, 이미숙. 2020, 소셜 빅데이터를 이용한 행복도 측정 방법론 개발 연구:구글 트렌드를 중심으로, 한국혁신학회지 15(5):215-237).
- Kim, D.H. and D. Kim, 2018, Development and application of dynamic visualization model for spatial big data. Journal of the Korean Association of Geographic Information Studies 21(1):57-70 (김동한, 김다윗, 2018, 공간 빅데이터를 위한 동태적 시각화 모형의 개발과 적용, 한국지리정보학회지 21(1):57-70).
- Mikolov, T., W. T. Yih, and G. Zweig. 2018. Deconfounded lexicon induction for interpretable social science. In Proceedings of the 2018 Annual Conference of the North American Chapter of the Association for Computational, pp 1615-1625.
- Yang, W., L. Tan, C. Lu, A. Cui, H. Li, X. Chen, K. Xiong, M. Wang, M. Li, J. Pei, and J. Lin. 2019. Detecting customer complaint escalation with recurrent neural networks and manually-engineered features. In Proceedings of the 2019 Annual Conference of the North American Chapter of the Association for Computational Linguistics (Industry Track), NAACL, pp. 56-63.
- Zhou, G. and K. Ganesan. 2016. Linguistic understanding of complaints and praises in user reviews. In Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA), NAACL, pp. 109-114. [KAGIS](https://www.aclweb.org/anthology/W16-0114)