

텍스트마이닝 기법을 활용한 지자체 콜센터 민원 분석

홍한솔, 유명식

승실대학교

Analysis of Civil Complaints at Local Government Call Centers Using Text Mining

HanSol Hong, Myungsik Yoo

Soongsil Univ.

요 약

정부 및 공공기관에서 행정서비스 확산으로 인해 시민들은 다양한 매체를 통하여 민원을 접수할 수 있게 되었다. 대표적으로 지방자치단체 콜센터는 전화 및 문자 상담을 기반으로 하여 시민들에게 더 편하고 쉽게 불만과 궁금증을 해결해주고 있다. 민원은 매년 증가하고 있으므로 해당 데이터들을 활용하여 정책 개발에 도움이 되고자 한다. 최근 2020년부터 2022년까지 총 3년간 수집되어있는 자료를 수집하여 키워드별로 분류하고 지역 데이터를 조합하여 분석하려고 한다. 분석 결과는 빈도수와 키워드만을 가지고 찾을 수 없었던 지역별 연관성 분석을 이용하여 특정 지역에 어떤 요구사항이 많은지 더욱 정확하게 파악할 수 있다. 본 연구를 통하여 지역별 특정 문의를 파악하여 정책 개발에 적극적으로 활용될 것으로 판단된다.

키워드 : 공공기관, 콜센터, 민원 분석, 텍스트 마이닝

1. 연구 배경 및 목적

전국의 많은 정부 및 공공기관에서 행정서비스 향상을 위해 다양한 방법을 시도하고 있다. 대표적으로 지방자치단체 콜센터를 구축함으로써 시민 편의 증진과 시민 중심의 민원 서비스 소통창구를 마련하고 있다. 공공서비스 제공에 대한 주민 의식이 향상되고 정보기술이 발전함에 따라 민원 제기 수단이 다원화되면서 신속하고 정확한 민원 해결을 위한 상담 서비스를 제공하는 수단으로서 콜센터의 역할이 강조되고 있다. [1]

최근 빅데이터의 관심과 더불어 민원 데이터의 저장량이 매년 증가하고 있는 상황에서 민원 분석에 대한 관심이 높아지고 있으며 민원 처리의 관리체계 및 시스템에 대한 변화가 시작되었다. 민원을 효율적으로 분석하고자 현재까지 워드 클라우드 방식[2], 네트워크 관계도 방식[3] 등 다양한 연구들이 제시됐지만 콜센터의 경우 예전 방식의 건수, 백분율 통계보고서 정도로만 분석해야 하는 실정이다. 해당 방식은 콜센터에서 연도별, 월별, 일별로 센터에서 지정해놓은 상담유형(대/중/소)에 맞게 분류하여 상담유형별 통계, 처리유형별 통계 등 지정해놓은 유형으로 들어온 데이터의 건수를 통계로 사용하고 있다. 상담사가 직접 유형을 선택하여 저장하는 방식이기 때문에 잘못된 유형으로 저장되어 있거나 2가지 이상의 문의가 들어오면 한 가지 유형으로만 분류된다는 문제가 발생하고, 지정해놓은 유형만으로 분류하여 통계 데이터를 추출하기 때문에 문의 내용에 대한 정확한 분석 자료로 활용되기엔 미흡한 부분이 있다.

위의 문제점을 해결하기 위해, 본 연구에서는 최근 3년간 콜센터를 통하여 제기된 민원 자료를 수집하여 ‘문의 내용’에 대한 데이터를 텍스트 마이닝 기법을 통해 소셜 네트워크 관계도와 민원인이 거주하는 행정 구역 데이터를 조합하여 민원 발생 지역 분포 특성과 유형의 연관성을 연구하고자 한다.

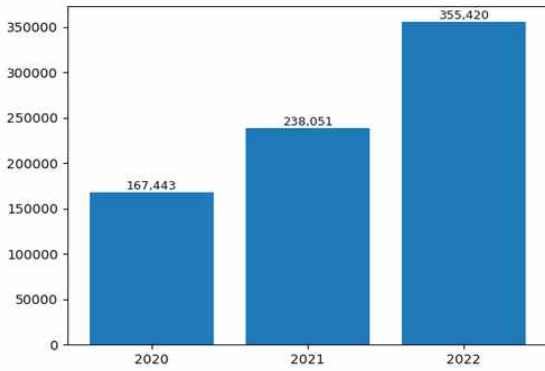
2. 연구 방법

본 연구에서는 콜센터로 접수된 2020년부터 2022년까지의 자료를 수집

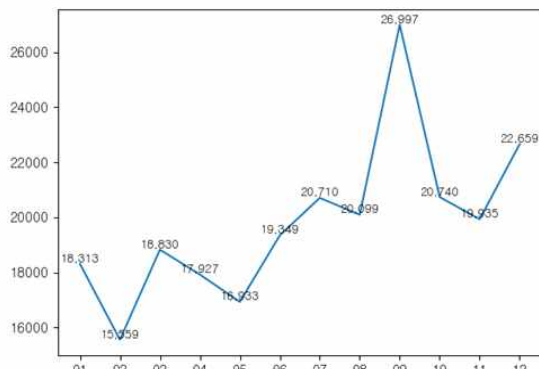
하여 민원 내용의 데이터만 추출하여 텍스트 마이닝을 진행하였다. 텍스트 마이닝이란 텍스트 문서 속에서 의미 있는 정보를 추출하기 위한 기법으로 비정형 데이터를 정형화한 후, 자연어 처리 기술에 기반하여 내용을 분석하는 것을 말한다. [4] 텍스트 방식으로 입력되기 때문에 맞춤법에 문제가 있을 수도 있으므로 네이버 맞춤법 검사기를 이용한 파이썬용 한글 맞춤법 검사 라이브러리인 py-hanspell을 이용하여 텍스트의 정확도를 높였다. py-hanspell의 경우 맞춤법 검사 성공 여부, 오류 개수, 요청 시간, 교정 결과를 비교하여 값을 반환해 주고 있다. 데이터 전처리 및 형태소 분석은 Python의 한글 자연어 처리 패키지인 Konlpy 중 mecab을 사용하여 명사만을 추출하였고, 민원 내용과 관련 없는 키워드들을 넣을 수 있는 불용어 사전과 데이터 분석 특성에 맞는 사용자 사전을 추가하여 의미를 갖는 단어가 빠지지 않도록 전처리 과정을 진행하였다.

2.1 데이터 시각화

2020년부터 2022년까지의 데이터는 760,914건으로 [그림1]을 보면 2020년엔 167,443건 이었으나 2021년에는 238,051건, 2022년에는 355,420건으로 콜센터의 민원 발생량이 해를 거듭할수록 증가한다는 것을 확인할 수 있다. 2021년에 발생한 민원 데이터를 월별로 시각화한 [그림2]와 해당 월에 키워드를 빈도수 별로 나타낸 [그림3]을 보면 9월의 경우 26,997건으로 한참 COVID-19로 인하여 재난 지원금 신청 및 백신 접종에 대한 문의가 많은 것으로 판단되었다. 반대로 가장 적은 2월의 경우 15,559건으로 재산세 납부 및 여권 발급에 관한 내용이 많았다.



[그림1] 연도별 통계 현황



[그림2] 2021년 월별 통계 현황



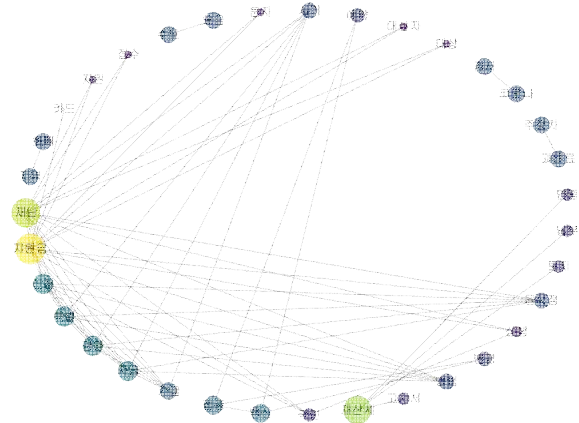
[그림3] 월별 키워드 시각화

추출된 키워드를 통해 서로 다른 키워드들의 집합이 얼마나 연관성이 있는지 파악하기 위해서 Apriori 알고리즘을 사용하였다. 지지도 = $P(A \cap B)$ 의 공식을 가지고 있으며, 그림4와 같이 지지도가 0.01 이상인 데이터만을 추출하였다.

antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(소상공인) (지원)	0.052925	0.056351	0.044568	0.842105	9.752122	0.039996	5.786444
1	(지원) (소상공인)	0.066351	0.052925	0.044568	0.516129	9.752122	0.039996	1.957289
2	(코로나) (지원)	0.047354	0.037604	0.037604	0.794118	21.117647	0.035824	4.674493
3	(지원) (코로나)	0.037604	0.047354	0.037604	1.000000	21.117647	0.035824	inf
4	(소상공인, 코로나) (지원)	0.034819	0.037604	0.034819	1.000000	26.592593	0.033510	inf
...
879	(코로나, 지원) (방역)	0.037604	0.005571	0.005571	0.148148	26.592593	0.005362	1.167373
880	(방역, 지원) (코로나)	0.005571	0.047354	0.005571	1.000000	21.117647	0.005307	inf
881	(지원, 방역) (코로나)	0.047354	0.005571	0.005571	0.117647	21.117647	0.005307	1.127019
882	(방역) (코로나, 지원)	0.005571	0.037604	0.005571	1.000000	26.592593	0.005362	inf
883	(지원) (방역, 코로나)	0.037604	0.005571	0.005571	0.148148	26.592593	0.005362	1.167373

[그림4] 키워드 연관성 분석 결과

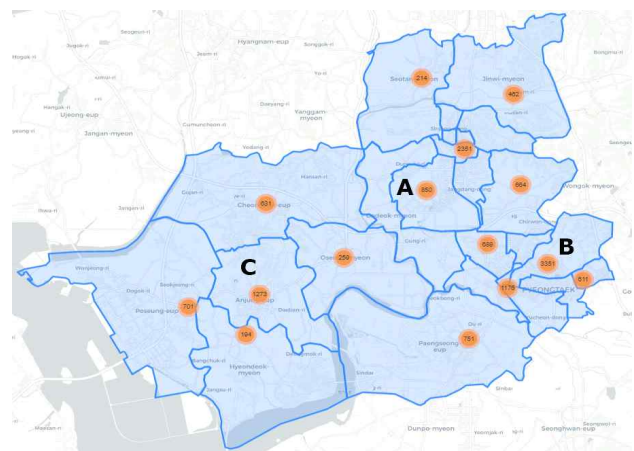
민원 내용의 분석을 위해 키워드 간 연결 관계를 분석하여 각 민원 키워드별로 시각화하였다. [그림5]의 키워드 간 관계를 살펴보면 재난, 지원금, 재산세가 민원의 중심으로 나타났으며 재난-지원금, 지역-화폐, 재산세-고지서 등의 연결 관계가 가장 많은 것으로 나타난 것으로 보아 재난 지원금 신청 및 지역 화폐 사용 문의와 같은 민원이 가장 많았던 것으로 판단된다.



[그림5] 키워드 간 네트워크 관계도

위도와 경도 및 지역코드 등이 포함된 GeoJson 데이터를 활용하여 folium 라이브러리를 사용해 지역별 민원 분포도를 확인하기 위하여 [그림6]과 같이 지도상에 민원 건수를 표현하여 어느 지역에서 민원이 많이 발생했는지를 파악할 수 있도록 하였다. 지역별 키워드 연관성을 분석하여 다음과 같은 내용을 파악할 수 있다.

A 지역의 경우 불법 주정차 신고 및 단속 문의가 많이 들어온 것으로 확인된다. 해당 지역은 반도체 현장이 있는 곳으로 현장 주변 곳곳에 불법 주차와 출퇴근 시간대에 겹치는 차들로 인하여 시민들의 불편이 많은 것으로 확인된다. B 지역의 경우 자동차세 연납 관련하여 문의가 가장 많았다. 고지서가 오지 않아서 기존에 신청한 사항이 정상적으로 처리되었는지의 관한 문의와 신규 신청 관련 문의가 많은 것으로 보인다. C 지역은 불법 쓰레기 무단 투기 관련 문의가 많았는데, 해당 지역은 쓰레기 배출 구역이 명확하게 정해져 있지 않은 곳이 많아 이와 같은 민원이 자주 발생하는 것으로 확인된다. 이처럼 행정 구역별로 각기 다르게 나타나는 민원들 확인할 수 있다.



[그림6] 행정구역별 민원 분포도

3. 연구 결과 및 기대 효과

본 연구는 지방자치단체 콜센터에서 활용하는 통계 자료에서 민원 분석에 대한 심층적인 분석의 어려움을 해결하고자 2020년부터 2022년까지 제기된 민원 데이터를 이용하여 텍스트 분석을 하였다. 연도별/월별 빈도 분석과 연관성을 파악하는 네트워크 분석을 사용하여 데이터 시각화를 진행하였고, 기존 반복되는 단어를 가지고 분석하는 것에서 끝내는 것이 아닌 행정 구역별로 민원을 분류하여 지역과 민원 키워드 간의 연관성을 분석하였다. 매년 민원 발생량이 증가하고 있고 지역별로 특정 문의가 지속적으로 제기되고 있는 불편함이 있기에 민원 데이터를 분석하여 확인하고자 하였다. 분석을 통하여 도출된 대표적인 주요 민원은 '불법 주정차'와 '쓰레기 무단 투기'이다. 발생하는 주요 민원들을 토대로 반복 제기되는 민원을 처리하는 방법에 대하여 정책을 추진하여 민원을 최소화하려는 노력이 필요하다고 생각된다.

참 고 문 헌

- [1] Kayeon Noh, Seunghee Shon, Bongju Jeong. (2012). A Study on Diffusion of Public Call Centers in Korea. IE interfaces, 25(3), 327-337.
- [2] Y. B. Park, J. M. Lee, D. H. Kang, S. D. Eun, J. Y. Park. (2022). Civil Complaint Data Analysis Using Text Mining : Focusing on the Visually Impaired. Journal of Rehabilitation Welfare Engineering & Assistive Technology, 16(2), 57-70.
- [3] Kim HyunJong, Lee TaiHun, Ryu SeungEui, Kim NaRang. (2018). A Study on Text Mining Methods to Analyze Civil Complaints - Structured Association Analysis -. Journal of the Korea Industrial Information Systems Research, 23(3), 13-24.
- [4] Mary Wu. (2017). Overview of Big Data Technology and Study of Diagnosis for Activation. Theology and Ministry, 48(), 351-371.