

## 이상탐지 활용 전자집단민원 추정 방법론에 관한 탐색적 연구: 창원시 시민의 소리 사례를 중심으로\*

정 하 영\*\*

### 요약

최근 전자민원시스템에 집단민원을 제기하는 사례가 늘어나고 있으나 이에 대한 효율적인 관리시스템이 아직 마련되어 있지 않아 행정 업무량 증대와 사회적 갈등 양산 등의 부작용이 우려되고 있다. 이에 본 연구에서는 이상탐지와 코퍼스 언어학 기반의 내용분석을 활용한 전자 집단민원 추정 방법론을 제시하고자 하였다. 이를 위하여 1)집단민원의 개념에 대한 이론적 고찰과 2) 비모수적 비지도 학습에 기반 한 이상탐지를 활용한 전자 집단민원 추정과 3) n-gram 코사인 각도 거리를 활용한 민원의 내용 유사도 분석방법론을 제안하고 4) 창원시 시민의 소리에 대한 사례분석을 통하여 제시한 방법론의 유용성과 정책적 시사점, 향후 과제를 검토하였다.

주제어 : 참여 민주주의, 전자 집단민원, 이상탐지, n-gram 유사도, 코퍼스분석

## An Exploratory Study of Collective E-Petitions Estimation Methodology Using Anomaly Detection: Focusing on the Voice of Citizens of Changwon City\*

Jeong, Ha-Yeong\*\*

### Abstract

Recently, there have been increasing cases of collective petitions filed in the electronic petitions system. However, there is no efficient management system, raising concerns on side effects such as increased administrative workload and mass production of social conflicts. Aimed at suggesting a methodology for estimating electronic collective petitions using anomaly detection and corpus linguistics-based content analysis, this study conducted the followings: i) a theoretical review of the concept of collective petitions, ii) estimation of electronic collective petitions using anomaly detection based on nonparametric unsupervised learning, iii) a content similarity analysis on petitions using n-gram cosine angle distance, and iv) a case study on the Voice of Citizens of Changwon City, through which the utility of the proposed methodology, policy implications and future tasks were reviewed.

Keywords : participatory democracy, collective e-petitions, anomaly detection, n-gram similarity, corpus analysis

Received Nov 19, 2019; Revised Dec 6, 2019; Accepted Dec 16, 2019

\* This work was supported by Kyungnam University Foundation Grant, 2018(2018-0232)

\*\* Kyungnam University, LINC+ Assistant Professor(jhayeong@kyungnam.ac.kr)

## I. 서론

최근 전자민원시스템에 집단민원을 제기하는 사례가 늘어나고 있다. 전자민원시스템은 온라인을 기반으로 누구든 언제 어디서나 민원을 제기할 수 있도록 한 오픈 시스템으로, 복사해서 붙이기(Copy&Paste) 기능을 사용하여 불과 몇 초 사이에 동일한 내용의 민원을 반복적으로 제기할 수 있고, 다른 사람의 민원을 똑같이 복제할 수도 있다. 즉 여러 사람들이 동일한 시간에 동일한 공간에 모이지 않고서도 동일한 목소리를 낼 수 있는 손쉬운 집단민원 제기 수단으로 전자민원시스템이 사용될 수 있다는 것을 의미한다.

원칙적으로 전자민원시스템은 개인의 실명인증 후 담당부서를 통해 민원접수·검토 후 개개의 민원을 구분하여 독립적으로 처리한다. 단시간 내에 제기되는 방대한 집단민원에 대해 건건이 개별처리 하는 것은 비효율적이며 처리 결과의 신뢰성을 보장하기 어렵다. 전자집단민원 처리를 장기간 방치한다면 행정업무량은 무한급수로 늘어날 것이며, 처리결과에 대한 민원인들의 불만이 증폭되어 사회적 갈등을 야기할 우려가 크다. 효율적인 전자집단민원 관리시스템의 근간이 되는 전자집단민원 추정이 선행되어야 한다.

기존의 전자민원분석은 핵심어를 추출하거나 핵심어와 관련된 연관어 분석을 통해서, 일정 기간 내 제기된 민원내용 전체를 요약하는 포괄적 요약에 초점을 두기 때문에 개별민원과 집단민원을 구분하여 취급하지 않는다. 전자집단민원 추정을 위해서는 민원제기 패턴과 내용 유사도를 고려한 새로운 접근 방법이 필요하다.

본 연구에서는 이상탐지와 내용분석을 통한 전자집단민원 추정방법론을 제시하고자 한다. 이를 위하여 2절에서는 집단민원의 개념에 대한 이론적 고찰, 3절에서는 이상탐지와 내용분석 방법론 검토, 4절에서는 장

원시 시민의 소리를 대상으로 사례분석, 5절에서는 정책적 시사점과 향후 과제를 논한다.

## II. 집단민원 개념 정의 및 국내·외 연구고찰

### 1. 민원의 개념

일반적으로 민원(Petition)은 청원(Appeal), 정식요청(Formal Request), 민원인의 이해관계(Concern)나 불만 사항(Complaint)을 공공기관에 제기하는 것을 의미하며, 우리 법률에서는 시민이 행정기관에 대하여 처분 등 특정한 행위를 요구하는 것으로 민원의 대상을 공적 영역으로 규정하고 있지만, 현실은 사적 영역을 포괄하는 경우가 많다.

Oh(1984)는 ‘인허가’의 관련 여부에 따라서 ‘일반민원’과 ‘복합민원(Complex Petitions)<sup>1)</sup>’으로 구분하며 복합민원의 예로 인가·허가·면허·협의·확인 행위를 들고 있다. KCC(1985)는 이해관계의 관련 여부에 따라서 ‘일반민원’과 ‘특별민원(Specific Petitions)’으로 구분하며, 특별민원의 예로 진정·투서·청원·건의·호소문 등을 들고 있다. 특별민원은 복합민원의 불허처분과 인과관계를 가진다. 행정은 복합민원(이해관계의 인허가 요청)에 대하여 공적 권리·의무·법률적 차원의 ‘위법성(Illegality)’ 및 ‘공정성’(Impartiality)<sup>2)</sup> 원칙에 근거하여 불허 여부를 판정하게 되는데, 판정근거에 대한 민원인의 ‘인식의 차이(Cognitive Dissonance)’가 클 경우 행정처분에 불복(Disposition)하고, 본인의 이해관계를 실현하기 위하여 이의신청·손해배상청구·행정서비스의 개선요청 등을 하게 된다. 즉 복합민원의 불허처분에 대한 결과, 특별민원이 발생하게 되는 것이다.

Yi, et al.(2009)는 진정·고발·고소 등과의 관련 여부에 따라서 ‘일반민원’과 ‘형사민원(Criminal Petitions)’

1) 복합민원은 하나의 민원목적을 실현하기 위하여 관계 법령 등에 의하여 다수의 관계기관 또는 부서의 인가/허가/승인/추천/면허/협의/확인을 받아야 하는 민원(Oh, 1984)

2) 아주 공평하여 어느 쪽으로도 치우침이 없는 불편부당(不偏不黨)

〈표 1〉 민원의 정의와 유형  
 〈Table 1〉 Definition and Types of Petitions

Literature review	Division	
	General petitions	Interest-related petitions
Oh (1984)	A specific demand input	Complex petition: a petition that must be authorized, permitted, approved, recommended, licensed, consulted, and verified by a number of related agencies or departments in accordance with relevant laws, etc.
KCC (1985)	A general request such as proof, request for confirmation, application for registration or approval, and various requests	Special petition: a special request for the interests of the petitioners, such as calm, suggestions, and appeals
Yi, et al., (2009)	<ul style="list-style-type: none"> <li>• General inquiry type petitions</li> <li>• Format petitions: registrar, apply, report, etc.</li> </ul>	Criminal petitions: complaints, accusation, lawsuit, etc.
ACRC (2013)	General petitions	Chronic petitions: Same or similar petitions that are continuously or repeatedly raised, petitions involving illegal, unfair, or irrational requirements, complaints, accusations, and lawsuits, excessive private interests demand
ACT No.14839 (2017)	<ul style="list-style-type: none"> <li>• Court petitions</li> <li>• Query petitions</li> <li>• Consultation and Description petitions</li> </ul>	Repeated petitions: petitions repeated three or more times for the same issue(excluding court petitions) Duplicated petitions: petitions submitted to tow or more administrative agencies with the same contents, and transferred from another administrative agency. Grievance petitions: petitions about the violation, discomfort, or burden of rights resulting from the illegal, unfair, negative disposition of the administrative agency, and the irrationality of the administrative system.

으로 구분하는데, 형사민원 또한 복합민원의 불허처분과 특별민원(KCC, 1985)의 처분에 대한 불만의 결과, 손해배상의 수준을 넘어 사형, 징역, 금고, 자격상실·정지, 벌금 등 형사처벌 행위를 적극적으로 요구하는 것이다.

ACRC(2013)는 ‘일반민원’과 ‘고질민원(Chronic Petitions)’을 구분하는데, 고질민원은 특별민원, 복합민원, 형사민원의 ‘반복성(Repeatability)’, ‘부당성(Injustice)’, ‘만성소모성(Chronic Wastingness)’의

특징을 가진 민원으로 과도한 행정처분의 비용과 노력을 수반하여 행정서비스의 질을 떨어뜨리는 원인이 되는 민원을 뜻한다. 이러한 행정손실을 방지하고자 민원처리법<sup>3)</sup>에서는 반복민원(Repeated Petitions)<sup>4)</sup>과 중복민원(Duplicated Petitions)<sup>5)</sup>을 구분하여 관리하고 있다. 여기서 민원내용의 동일성은 내용적 유사성·관련성 및 종전 민원과 동일한 답변을 할 수밖에 없는 사안을 고려하여 판단한다. 또 민원처리법에서는 ‘일반민원’과 ‘고충민원(Grievance Petitions)’으로

3) 민원 처리에 관한 법률(법률 제14839호, 2017. 7. 26, 제23조)의 약칭.

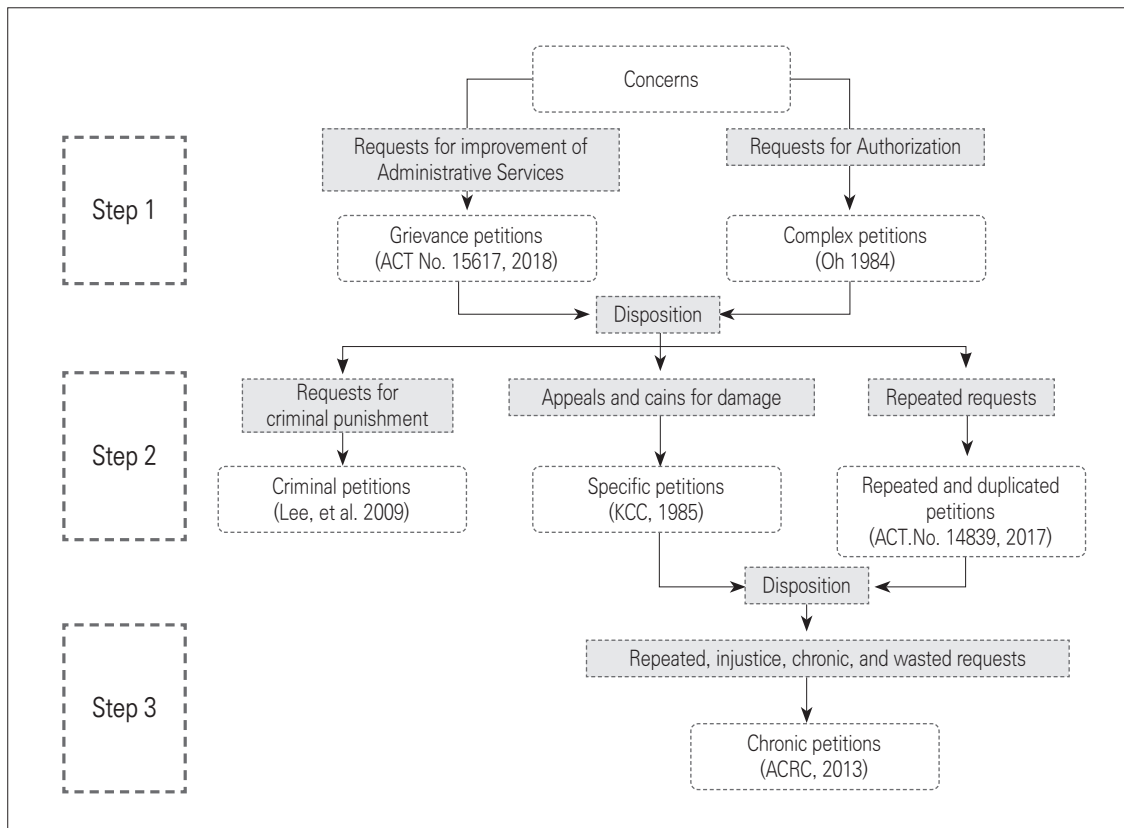
4) 반복민원은 한 민원인이 동일한 내용의 민원(법정민원 제외)을 3회 이상 반복 제기하는 민원

5) 중복민원은 한 민원인이 2개 이상의 행정기관에 제출한 동일한 내용의 민원을 다른 행정기관으로부터 이송받은 반복민원에 준용하는 민원

구분하는데, 고충민원<sup>6)</sup>은 행정서비스 개선에 대한 요청을 말한다.

위 내용을 정리하면 1단계에서 민원인이 본인의 이해관계를 달성하기 위해 행정서비스의 개선요청이나 인허가 요청을 하게 되는데, 전자가 고충민원, 후자가 복합민원에 해당한다. 만약 민원인이 행정 처분 결과에 대해 불복할 경우, 2단계에서 고소·고발 등 형사처벌이나 이의제기·손해배상청구를 신청 할 수 있으며,

이를 반복 요청하게 된다. 이것이 각각 형사민원, 특별민원, 반복·중복민원에 해당한다. 특별민원과 반복·중복민원의 경우 반복·부당·만성소모성의 정도가 매우 높을 경우 고질민원으로 분류한다. 이처럼 서로 다른 민원으로 분류되지만, 이들 사이에는 인과관계가 있는 것으로 이해되며, 본 연구에서는 이를 ‘이해관계민원’의 포괄적 개념으로 정의하고 <그림 1>과 같이 도식화하였다.



〈그림 1〉 민원 간 인과관계  
 〈Fig. 1〉 Causal Relations Between Petitions

6) 고충민원은 부패방지권익위법(법률 제15617호, 2018. 4. 17. 제2조 제5호)에서는 행정기관 등의 위법·부당하거나 소극적인 처분(사실행위 및 부작위를 포함한다) 및 불합리한 행정제도로 인하여 시민의 권리를 침해하거나 시민에게 불편 또는 부담을 주는 사항에 관한 민원(현역 장병 및 군 관련 의무부복자의 고충민원을 포함한다)이라 정의하며, 시민 신문고는 ① 법령·제도·절차 등 행정업무에 관한 질의 또는 상담형식을 통한 설명이나 해석의 요구, ② 정부시책이나 행정제도 및 운영의 개선에 관한 건의, ③ 행정기관의 위법·부당하거나 소극적인 처분 및 불합리한 행정제도로 시민의 권리를 침해하거나 시민에게 불편 또는 부담을 주는 사항의 해결요구라고 정의하고 있다(ACT No.15617, 2018).

## 2. 집단민원

집단민원은 ①민원인의 수, ②집단의 특성과 민원제기 형태, ③반복성 여부 등의 분류 기준에 따라서 다양하게 정의되고 있다.

첫째, 민원인의 수에 따라서는 단일민원과 집단민원(Collective Petitions)으로 구분한다. ACT No.14839(2017)에서는 집단민원을 ‘다수인관련민원(Multistakeholder-Associated Petitions)’과 ‘집중관리민원(Intensive Management Petitions)’ 두 가지 민원으로 분류하고 있다. 다수인관련민원은 ‘5세대(世帶) 이상의 공동이해와 관련되어 5명 이상이 연

명으로 제출하는 민원’이라 정의하며, 그중 20세대(世帶) 이상의 공동이해와 관련되어 20명 이상이 연명으로 제출하는 민원의 경우 ‘집중관리민원’으로 분류하여 시·도와 시·군·구의 감사실에서 해당 부서의 처리 상황을 확인 점검하고 있으며, 발생 및 처리 결과를 정기적으로 국민권익위원회에 보고하고 국민권익위원회에서는 법령과 제도의 개선방안 등을 강구하며 매년 집단민원 종합 관리대책(행정지침)을 마련하여 각급 지방 관서에 시달한다(Jung, 2006).

둘째, 집단의 특성과 제기 형태에 따라 집단민원, 무리민원, 조직민원으로 구분하기도 한다(WCPP, 2019). 공통점은 다수인의 요청사항에 ‘내용의 동질

〈표 2〉 민원인 수 기준 분류: 다수인관련민원과 집중관리민원

〈Table 2〉 Classification by Number of Petitioners:

Multistakeholder-associated Petitions vs. Intensive Management Petitions

Classification	Definition
Multistakeholder-associated petitions	Made up of five or more persons that is related to common interests of five or more households
Intensive management petitions <sup>7)</sup>	Made up of more than 20 persons that is related to common interests of more than 20 households

source: Act No.14839(2017)

〈표 3〉 집단의 특성과 제기 형태에 따른 집단민원의 구분

〈Table 3〉 Classification by Group Types and Petition Forms

Classification	Definition
Collective Petition	Includes a collection of signatures from a number of petitioners concerning a particular request or complaint
Group Petition	Made up of individual or group inputs from a number of petitioners concerning the same or substantially similar request or complaint
Association Petition	An individual or group request from an association or an individual mandated by that association, concerning a particular complaint or request of the association.

source: WCPP(2019)

7) 대통령비서실, 국무총리비서실, 감사원, 국가정보원(이하 “대통령비서실 등”이라 한다)에서 이송된 민원

성'이 있다는 것이다.

셋째, 반복성 여부에 따라 빈발민원과 다량민원으로 구분한다(PSIIF, 2005). 빈발민원의 정의에 사용된 용어 '계속해서'를 통해 집단민원은 일정한 시간적 간격을 두고 반복하여 나타나는 '주기성'을 가지는 것을 알 수 있다. 다만, 어느 정도의 주기를 '빈발'이라고 정의하는지에 대해서는 논의가 구체적이지 않다. 또한 다량민원의 정의에 사용된 용어 '다수'를 통해서, 집단민원이 '양적 성질'을 가지는 것을 알 수 있지만, 이 또한 기준이 명확하지 않다.

이들 정의에서 알 수 있듯이 집단민원은 네 가지 요소(다수인, 양, 주기성, 요구사항의 동질성)를 모두 가지고 있는 민원으로 해석할 수 있으며, 이 네 가지 요소를 어떻게 객관적으로 분석할 것인가가 본 연구의 핵심 관심사이다.

### 3. 전자민원

2001년 전자정부법<sup>8)</sup> 제정을 시작으로 전자정부 11대 로드맵 과제에 전자민원이 포함되면서 그해 12월 전자민원G4C(Government For Citizen)<sup>9)</sup>사업이 기획되고 2003년 참여정부가 약 9,300여 다수 부처 간 연계·통합 네트워크 구축과 온라인 민원서류 발급 시스템을 구축하면서 전자민원시스템이 출현하였다(EGOV, 2008). 그동안 오프라인에 의존했던 민원의 신청·처리 모두 온라인으로 해결할 수 있게 되었다.

〈표 5〉 전자민원서비스 현황을 살펴보면, 2003년 주민등록 등·초본 등 8종(신청 약 7,734천, 처리 약 57천)에서 2018년에는 1,182종(신청 약 68,359천, 처리 약 70,408천)으로 발급종수는 약 147.8배, 신청

〈표 4〉 반복성 기준 집단민원 구분: 빈발민원과 다량민원  
 〈Table 4〉 Classification by Repeatability: Repeated vs. Massive Petitions

Classification	Definition
Repeated Petitions	Repeatedly made up of a number of petitions on a certain issue
Massive Petitions	Multiple petitions on the same issue

source: PSIIF(2005)

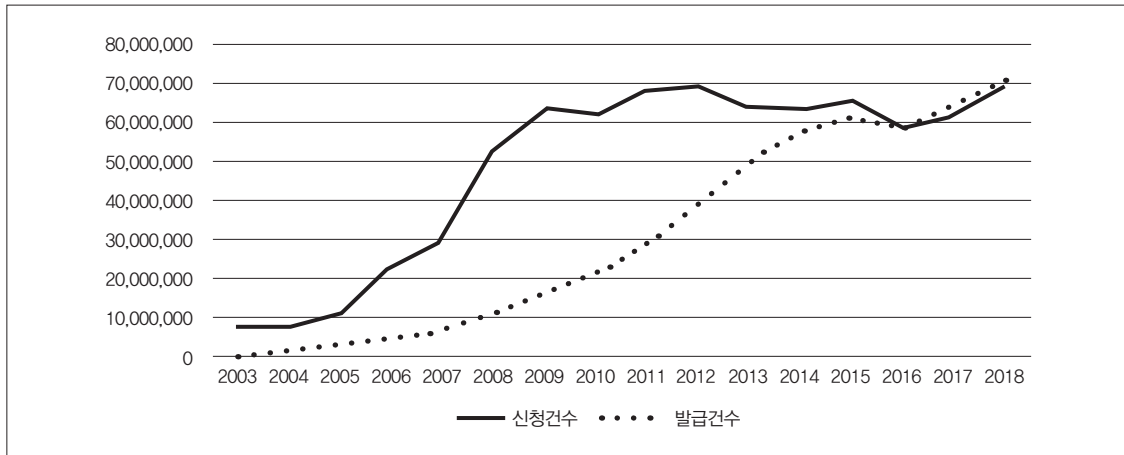
〈표 5〉 전자민원서비스 현황  
 〈Table 5〉 Status of E-petitions

	2003	2018	Increase(times)
Number of Petition Types	8	1,182	147.8
Number of Applications	7,734,925	68,359,627	8.8
Processed	57,826	70,408,837	1217.6

source: EGOV24(2019)

8) 전자정부구현을 위한 행정업무 등의 전자화 촉진에 관한 법률의 약칭

9) 전자민원G4C는 인터넷, 휴대전화, PDA, 케이블TV 등 다양한 매체를 이용하여 언제 어디서나 쉽고 편리하게 원하는 민원 행정서비스 및 기관별 행정정보를 이용할 수 있도록 정보화함으로써 국민과 더욱 가까워진 유비쿼터스 전자정부를 만들어 가는 사업이다(전자정부24, 2008).



source: EGOV24(2019)

〈그림 2〉 전자민원서비스 현황 시계열 변화(2003~2018)  
 〈Fig. 2〉 Time-series Change of E-petitions (2003~2018)

건수는 약 8.8배, 발급 건수는 약 1217.6배 증가했다. 인구 1인당 약 1.36건<sup>10)</sup>의 민원을 처리한 셈으로 비약적인 발전을 이루어 왔다.

하지만, 〈그림 4〉 전자민원서비스 현황의 시계열 변화를 살펴보면, 2016년 이전까지는 발급 건수가 신청 건수보다 적었다.

이는 제도적·재정적·기술적·민원인 관련 문제로 인해 전자민원 수요에 비해 공급 속도가 따라가지 못해 민원처리가 제때(Punctually) 이루어지지 못했기 때문으로, 그 결과 불허·반려·지연처리·종결처리·미처리 등 ‘미해결민원’<sup>11)</sup>이 많이 발생했을 것으로 예측할 수 있다.

민원24 전자민원은 행정기관이 정한 정규 민원서식을 사용하여 하나의 의사에 대해서 한 방향으로만 의사를 전달하는 시스템으로 민원인과 행정기관 사이의 상호 호혜적 의사소통(Reciprocal Communication)이 불가능한 구조로 되어 있다. 만약 미해결민원에 대

해 민원인이 궁금증을 문의하거나 해결방법을 상의하는 등의 즉각적인 후속 대응 절차는 아직 마련되어 있지 않다. 이러한 미비한 대응은 민원인의 불만을 가중시킬 수 있고 반복·중복민원 또는 잠재적 갈등요인을 야기할 우려가 높다. 미해결민원과 반복·중복민원과 상관관계를 분석하여 대응방안을 마련하는 것이 중요하다.

2003년 5월 12일 국민 참여 수석실이 국민과의 쌍방향 의사소통의 통로로서 인터넷에 ‘국민참여마당’을 개통하면서, 시민사회의 정책참여를 높이고 민주적 상호식 민원해결시스템 구축의 일환으로 ‘참여형 전자민원시스템’의 토대를 마련했다(Eun, 2006). 시민들은 자신의 이해관심과 요구사항을 자유로이 표현할 수 있고, 또한 다른 사람들의 관심과 요구사항을 접하면서, 사적인 생활불편에 대한 개선요구 뿐만 아니라 지역공동체의 가치를 이해하고 지역발전에 필요한 개선점을 제안할 수 있게 된 것이다. 또한, 공공정책결정의 계획

10) 2018년 (전국) 주민등록인구 51,826,059명을 기준으로 계산함.

11) 민원요청을 제때(punctually) 처리하지 못하고 불허·반려·지연처리·종결처리·미처리된 민원을 본 연구에서 미해결 민원이라 부르기로 한다.

〈표 6〉 형식형 민원과 참여형 민원  
 〈Table 6〉 Formal Petitions vs. Participatory Petitions

Classification	Definition	Example
Formal Petitions	Form-dependent petitions that apply by entering the required information on a form designated by the government.	Minwon24
Participatory Petitions	Private contextual petitions that apply freely to their concerns and needs in their own language system.	Sinmungo

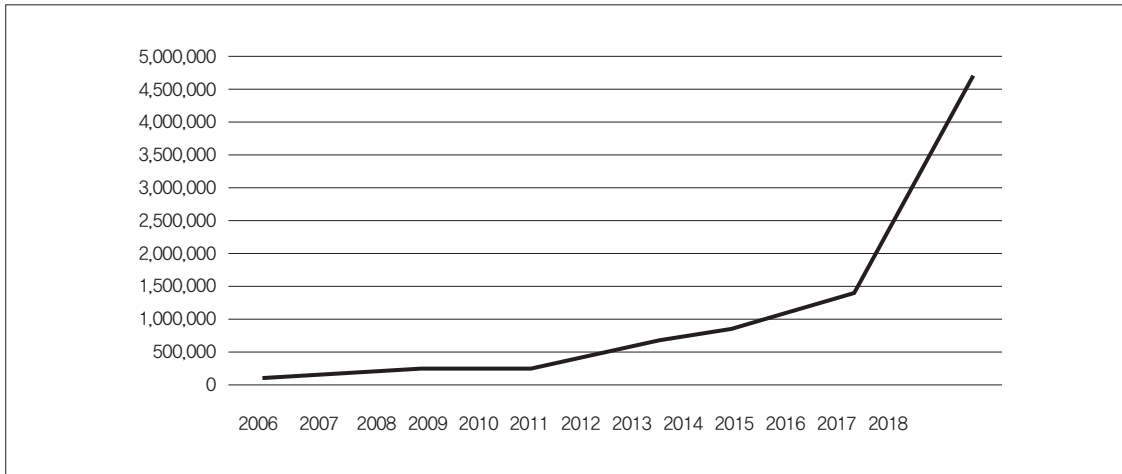
〈표 7〉 참여형 민원 현황(국민신문고)  
 〈Table 7〉 Status of Participatory Petitions(Sinmungo)

	2006	2018	Increase(times)
Number of applications	120,617	4,754,302	39.4

source: ACRC(2019b)

초기 단계부터 민원의 집계 결과를 의사결정의 근거자료로 명시화, 형식화함으로써 직접민주주의를 실현하는 전자적 시민참여를 통한 지역혁신의 기틀을 마련할 수 있다. 국내에서는 현재 ‘국민신문고’의 형태로 발전하여 운영되고 있다. 앞서 살펴본 민원24가 국가가 지

정한 서식에 맞게 요구되는 정보를 입력하여 신청하는 ‘서식 의존형 민원’인 것과 달리, 국민신문고는 시민 개개인 고유의 언어체계를 있는 그대로 표현하여 신청하는 ‘개인 문맥형 민원’으로, 본 연구에서는 이 둘을 구분하여 전자를 ‘형식형 전자민원’, 후자를 ‘참여형



source: ACRC(2019a)

〈그림 3〉 국민신문고 민원접수 추이(2006~2018)  
 〈Fig. 3〉 Sinmungo Application Trend (2006-2018)



전자민원'이라 부르기로 한다.

〈표 7〉은 대표적인 참여형 전자민원인 국민신문고의 민원접수 현황을 정리한 것이다. 2006년 연간 약 12만 건을 접수하였던 것이 2018년에는 약 475만 건 이상을 접수하여 약 39.4배 증가했다. 인구 100명당 9명<sup>12)</sup>이 국민신문고에 민원을 접수한 셈이다.

〈그림 5〉는 국민신문고의 민원접수 시계열 변화를 살펴본 것으로 2016년을 기점으로 급격히 증가하고 있는 것을 알 수 있다.

이러한 급격한 증가는 집단민원과 관련성이 매우 높을 것으로 예상된다. 국민신문고에 접수되는 민원 중 무엇을 집단민원으로 볼 것인가? 즉 참여형 전자민원을 활용하여 어떻게 집단민원을 정의할 것인가? 집단민원을 추정하는 새로운 방법론이 필요하다.

#### 4. 참여형 전자민원 내 집단민원 추정

최근, 코퍼스 언어학(Corpus linguistics)을 활용하여 참여형 전자민원의 내용을 분석하고 그 결과를 정책 의사결정에 반영할 수 있도록 지원하는 방안을 모색하려는 노력이 다양하게 시도되고 있다. 이는 전자민원이 단순히 지역구성원 개개인의 이해 관심의 집합체(Collection of Concerns)라는 인식을 넘어, 지역 문제<sup>13)</sup>를 이해하기 위한 근거자료로서 필히 검토해야 할 대상이며 행정서비스에 필요한 정보인프라(Information Infrastructure)라는 인식이 자리 잡기 시작했기 때문이다. 그러나 참여형 전자민원이 참여의 허울에 빠지지 않고 정책 의사결정에 잘 활용되기 위해서는 풀어야 할 숙제가 많다.

첫째는 전자민원분석에 대한 '정통성(Legitimacy)'

을 확보하는 것이 필요하다. 아직 전자민원분석의 체계나 방법론에 대한 명확한 원칙이 마련되어 있지 않은 상황에서 연구자들은 각자의 관점에 따라 다양한 각도에서 여러 가지 분석을 시도하고 있지만, 아직 초보적인 단계에 머물러 있으며, 이러한 분석 결과는 정당성과 합리성을 검증하기 어렵다. ACRC(2013)는 국민신문고를 대상으로 키워드 분석 및 성별, 지역 등 민원인의 특성을 고려한 크로스 분석을 통하여 주요 민원내용 요약 및 민원 경향을 분석하였다. 이는 거시적 관점에서의 분석은 가능하나 지역별 특성을 고려한 미시적 분석이 불가하며, 또 출현 빈도를 기준으로 한 키워드 중심의 분석 방법은 많이 언급된 단어를 중요하다고 판단하기 때문에 반복민원이나 다수인 관련 집단민원과 같은 빅보이스(Big Voice)가 두각 되고 소수의 잠재·소외민원<sup>14)</sup>이 분석대상에서 제외될 수밖에 없는 한계점을 내포하고 있다. 빅보이스 노이즈 제거 기준 마련을 통하여 민원 중요도의 적정성을 확보하는 것이 중요하다.

둘째는 참여형 전자민원의 역기능에 대한 철저한 예방관리 체계를 마련하는 것이다. 대표적인 역기능으로 여론몰이<sup>15)</sup>를 들 수 있다. 전자민원은 오픈 공간으로 이익집단이 관여하기 쉬우며, 이들은 정책 결정 과정에 영향력을 행사할 것을 목적으로 '복사해서 붙여넣기'를 통해서 조직적으로 단시간에 동일하거나 유사한 내용의 민원폭탄을 만들어 낼 수 있다. 집단 간 이해관심이 첨예하게 대립하는 사안의 경우는 상황이 더욱 심각해질 수 있다. 전자민원은 단편적 의견을 제기하는 일방적인 의사소통 방식으로 다른 민원인과의 상호호혜적인 납득의 과정(=토론의 과정)이 결여되어 있어, 집단 간의 간극을 좁힐 수 있는 장치가 전혀 마련

12) 2018년 (전국) 주민등록인구 51,826,059명을 기준으로 계산함.

13) 지역 문제는 하나 이상의 상호의존적인 잠재영역이 제 기능을 발휘하지 못하고 특정 영역의 발전에 불균형을 초래하는 시스템적인 현상을 말한다. 즉, 시민 생활 관련성이 존재하고, 불특정 다수의 시민에게 영향을 미치는 경우, 지역 문제로 볼 수 있다.

14) 소외민원은 제도권에서 논의되지 못한 행정 사각지대의 민원으로서 소외계층의 민원이 아니라 사회적으로 이슈화되지 않은 민원을 의미하며, 이를 그대로 둘 경우 장차 집단화할 가능성이 크므로 소외민원 발굴은 예방행정을 통해 행정비용을 절약하는 의미가 있다.

15) 여론몰이는 개인이나 집단이 개인의 사적인 목적이나 자기 집단의 이익을 위해 방송, 신문, 인터넷 등의 언론을 이용하는 행위로서, 공적인 이익에 부합하지 않는 경우를 말하며, 부정적인 의미를 담고 있다. 사실을 과장하거나, 왜곡하거나, 모순되는 말을 하여 여론을 조작하는 것을 일컬으며, 언론플레이(media brainwashing)라고 일컬어지기도 한다.

되어 있지 않다. 이러한 문제에 대해서 Yoon(1997)은 지금의 전자민원 시스템이 공적인 문제에 관해 속고 하고 다른 사람의 의견과 비교해 볼 수 있는 기회, 다시 말해서 사회를 위한 도덕적 책임감을 성찰할 기회를 봉쇄할 수 있다고 지적한다. 무방비로 노출된 참여형 전자민원은 '고충·잠재·소외민원의 발굴'이라는 고유의 역할을 제대로 수행하지 못하고, 오히려 '여론조작·갈등 고착화'라는 악순환의 늪'으로 전락하여 장기미해결 집단민원의 온상지가 되어 버릴 우려가 크다. 국민신문고와 각 지자체에서 운영되고 있는 온라인 시민 소통창구의 경우, 평소 시정에 관한 개선 또는 생활 불편사항 등 시민의 다양한 소리를 시정에 반영하기 위한 통로로 운영되고 있으나 '민원 처리에 관한 법률'의 적용을 받지 않기 때문에, 종결처리의 개념이 없고, 끊임없이 제기되는 반복성 민원에 대해서 개별 답변을 지속해야 하는 상황이다. 집단민원에 대한 신속한 대처는 지역공동체의 공익을 저해하고 정부의 신뢰성(Trustworthy)을 떨어뜨린다. 따라서 전자집단민원에 대한 별도의 관리시스템 마련이 시급한 실정이다.

이제 개인-공동체간의 이해관계의 대립과 다양한 사회 집단의 분화 현상이 어떠한 형태로 전자민원에 표출되고 있는지 과학적 분석이 시도되어야 할 시점이다. 본 연구자는 코퍼스 내용분석 방법을 활용하여 전

자민원에 내포되어 있는 시민들의 공적관심사(Public Concerns)를 객관적이고 자동적으로 집약하는 방법론을 개발해 왔다(Jeong, et al., 2007, 2008, 2010, 2017; Jeong, 2008; Lee, et al., 2017). 내용분석방법론의 핵심은 두 가지 문제로 요약할 수 있다. 첫 번째 문제는 문서에 등장하는 단어의 '중요도'를 어떻게 결정하여 핵심어(Keywords)로 정의할 것인가 하는 문제이며, 두 번째 문제는 핵심어와 관계되는 연관어(Co-Occurrences)를 어떤 기준으로 결정하여 문맥(Context)을 정의할 것인가 하는 문제이다. 집단민원을 이해하기 위해서는 '시계열 제기 패턴'에 기반을 둔 새로운 접근방법이 필요하다. 이에 본 연구에서는 이상탐지 방법론과 내용분석방법론을 활용한 전자집단민원 추정 방법론을 제시하고자 한다.

### III. 전자집단민원 추정 방법론

#### 1. 이상탐지 방법론

이상탐지(Anomaly Detection)는 기존 관측치와 비교하여 현저히 다른 메커니즘에 의해 생성되었다고 판단할 만한 관측치를 찾는 통계 방법론(Hawkins, 1980)으로 이상치 탐지(Outlier Detection)<sup>16)</sup>와 신

〈표 8〉 이상치의 구분  
 〈Table 8〉 Classification of Outliers

Classification	Definition	Example
Global Outliers (or point outliers)	Observations that deviate significantly from the entire group of observations, called anomaly	Anomaly detection
Contextual Outliers (or conditional outliers)	Obvious deviations associated with a particular context of an object	Data that may be normal or abnormal depending on the situation
Collective Outliers	Partial observation groups that deviate significantly from the entire group of observations	DDoS attack

Source: Edited by author referring to Modi&Oza(2016) and Aggarwal(2013)

16) 이상치 탐지(outlier detection)는 훈련데이터에 이상치(outlier)를 포함하는 경우로, 이상치 탐지 추정자(estimator)가 비정상적인 관측치들을 무시하여 훈련데이터가 가장 집중된 영역에 적합하도록 한다(scikit-learn, 2019).

〈표 9〉 이상치 탐지의 원리  
 〈Table 9〉 Principles of Outlier Detection

Classification	Principle	Approach	Criteria
Discordancy tests	After assuming the null hypothesis that the probability density function of a parameter follows a specific distribution, if the outlier does not belong to the distribution or does not correspond to a given significance level, the hypothesis is adopted to define it as an outlier.	Parametric	Statistical Probability Distribution Model
Proximity	Define outliers when the locality (or proximity) of the observations is sparse	Non-parametric	Clustering, Distance, Density
Labeling methods	Define outliers based on the location, spread, and skewness of univariate or multivariate data.		Supervised Semi-supervised Unsupervised

source: Edited by author referring to Aggarwal(2013)

상치 탐지(Novelty Detection)<sup>17)</sup>로 구분한다. 이상치 탐지(Outlier Detection)는 이상치가 의사결정에 영향을 미칠 수 없도록 제거하거나 중요도를 낮추는 것을 목적으로 하며, 대표적인 예로 잡음 제거(Oh, et al., 2018)가 있다. 신상치 탐지는 지금까지 발견되지 않았던 새로운 메커니즘을 찾아내고 정상의 범위에 포함시키는 것을 목적으로 하며, 대표적인 예로 신종 바이러스, 악성코드 발견(Richardson, 2009; Watson, et al., 2016)이 있다.

컴퓨터 시스템 침입 탐지(Min, et al., 2014), 보험금 청구 사기탐지, 유해 작업환경 이상탐지, 소비 패턴 이상 감지를 통한 신용카드 도난 탐지, 제조공정 불량 원인 감지, 편향의견검출(Yeon, et al., 2013), 치매 조기 진단(Oh, et al., 2018) 등 최근 소개되고 있는 빅데이터 기반의 이상탐지는 대부분은 글로벌 이상치를 대상으로 한 것이다. 한편 전자집단민원은 디도스(DDoS) 공격과 같은 집단적 이상치로 볼 수 있다.

이상탐지의 원리는 3가지로 구분할 수 있다. 첫째

불일치 테스트(Discordancy Tests)로 모수의 확률밀도 함수가 특정 분포를 따른다는 귀무가설을 가정 한 후 이상치가 분포에 속하지 않거나 주어진 유의 수준에 해당하지 않는 경우 가설을 채택하여 이상치로 정의하며, 이는 모수적 기법으로 통계적 확률분포 모형이 대표적인 예다.

둘째 근접성 기준(Proximity)으로 관측치의 지역성(또는 근접성)이 희박한 경우 이상치로 정의한다. 이는 비모수적 기법으로 거리기반·밀도기반 군집이 대표적인 예다. 셋째 라벨링 기법(Labeling Methods)으로 단변량·다변량 데이터(Univariate or Multivariate Data)의 위치(Location), 확산(Spread) 및 왜곡(Skewness)을 기준으로 이상치를 탐지하는 접근법이다. 이는 비모수적 기법으로 지도 학습, 준지도 학습, 비지도 학습으로 구분한다.

이상치 탐지의 구체적인 기법들의 예를 〈표 10〉로 요약하였다.

17) 신상치 탐지(novelty detection)는 훈련 데이터에 이상치(outlier)를 포함하지 않는 경우로, 새로운 관측치가 이상치 인지 아닌지 감지하는데 주안을 두며, 이러한 맥락에서 이상치를 새로움 (novelty)이라고 부른다(scikit-learn, 2019).

〈표 10〉 이상치 탐지 기법  
 〈Table 10〉 Approaches to Outlier Detection

Approach	Model	Algorithm	Criteria
Parametric	Normal Model	Grubbs	Average
		Cochran	Variance
	Regression model	Robust regression	Residual
		ARIMA	Residual
	Mixed parametric model	EM	Estimation of the distribution of latent variables
Non-parametric	supervised	KNN	Distance based clustering
		SVM	Binary Classification
		LOF	Density-based
		Decision Tree, Isolate Forest, Random Forest	Dimension Reduction of Independent Variables
		Naive Bayes	Binary and multiple classification
	unsupervised	Histogram	Density estimation
		Kernel Density Estimation	Density estimation
		Box Plot, Violin Plot <sup>18)</sup>	Interquartile Range Estimation
		Multi-Layer Perceptron	Dimension Reduction
		K-Means Clustering	Distance based clustering

## 2. 내용분석을 통한 전자집단민원 추정 방법론

### 1) 전자집단민원의 정의

본 연구에서는 전자집단민원을 두 가지의 관점에서 정의한다. 첫째 집단민원의 성격을 기준으로 ‘다수인 관련 민원으로, 내용에 유사성 또는 동질성이 있으며, 반복성과 주기성이 있는 민원’, 둘째 이상치의 성격을 기준으로 ‘1일 내 제기되는 민원 건수가 정상 수준을 넘으며 현저하게 다른 메커니즘으로 생성되는 민원’이다.

### 2) 전자집단민원의 추정 프로세스

전자민원은 시간 데이터와 문자 데이터를 포함하기 때문에 전자집단민원 추정 프로세스는 3단계로 진행된다.

1단계에서는 시간 데이터에 대해 시계열분석을 실시하여 정상수준을 확인하고 이상치를 탐지한다. 시간 데이터에서 이상치를 탐지하는 것은 1) 비정상적인 시점을 찾는 것과 2) 비정상적인 변화의 패턴을 찾는 것 두 가지로 나뉘는데(Oh, et al., 2018), 본 연구에서는 전자를 기준으로 이상치를 탐지하고, 후자는 다음 연구에서 발전시키고자 한다. 전자민원의 시간 데이

18) 박스 플롯에 커널 밀도 곡선을 추가하여 자료의 분포를 쉽게 파악할 수 있도록 한 그래프

〈표 11〉 전자집단민원의 정의와 추정 프로세스  
 (Table 11) Definition and Detection Process of Collective E-petitions

	Parts	Details	
Definition	Characteristics of Collective Petitions	<ul style="list-style-type: none"> <li>• Multiple stake holders involved.</li> <li>• High similarity or homogeneity of content.</li> <li>• Repeatability and periodicity are remarkable.</li> </ul>	
	Characteristics of Outliers	<ul style="list-style-type: none"> <li>• The number of petitions within one day is above the normal level and has a large volume.</li> <li>• Remarkably different generation mechanism.</li> </ul>	
Detection Process	Time data	Time series analysis	Augmented Dickey-Fuller for Stationarity check
		Outlier detection	Box plots
	Text data	n-gram vector generation	Korean morphological analyzing, MPH
		Similarity analysis	Lvenshtein distance, LD
Embedding	Visualization	Multi-dimensional scaling, MDS	

터는 이산형(Discrete)의 단변량 데이터(Univariate Data)로 특정한 확률모형을 가정하기 어렵고, 집단민원을 선형적으로 규정할 수 없으므로, 비모수적 비지도 학습에 기반을 둔 이상치 탐지방법이 필요하다. 이에 본 연구에서는 밀도추정에 기반을 둔 박스 플롯<sup>19)</sup>을 활용하여 이상치를 추정한다.

박스플롯의 분석 결과가 정상 범위 이내일 경우에는 개별민원으로 보고 기존의 민원처리 절차를 따르도록 하고, 범위를 벗어나 이상치로 판단되는 경우에는 집단민원으로 본다.

2단계에서는 문자 데이터에 대해 내용분석을 실시하여 문맥적 이상치<sup>20)</sup>를 탐지한다. 문자 데이터(Text

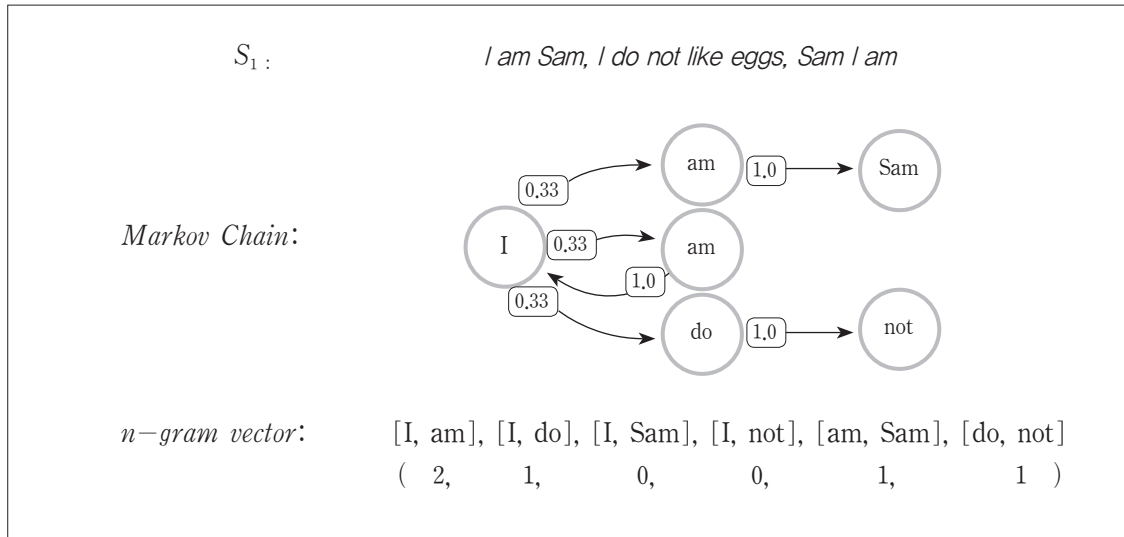
Data)의 이상탐지는 주로 문서 모음(Text Collection)에서 새로운 주제나 사건을 찾는 것을 목적으로 하는데(Oh, et al., 2018), 의미적 유사성(Semantic Similarity)을 기준으로 판단한다. 민원의 경우, 문맥을 잃지 않는 것이 중요하여 형태소 단위가 아닌 문자열 단위를 기준으로 유사성을 분석하게 되는데, 문자열의 의미적 유사성을 측정하는 다양한 방법<sup>21)</sup>이 제안되었다. 이 중 n-gram 모델은 마르코프 체인 알고리즘(Markov Chain Algorithm)<sup>22)</sup>을 적용한 확률론적 언어 모델(Probabilistic Language Model)로 기존의 문자열 비교 방식과는 달리 내부 문자열들의 순서가 다르더라도 맥락적 이해(Contextual

19) 박스 플롯(box plot)은 데이터의 분포와 이상치를 시각화한 그래프로 통계학자 존 투키(John Tukey, 1977)가 처음 제시했다. 의학(Laurikkala et al., 2000; Horn et al., 2001; Solberg & Lahti, 2005), 터빈 로터 자료(Guttormsson et al., 1999)의 이상치를 찾는 데 활용되었다. 박스 플롯은 사분위간 범위(Interquartile range, IQR)를 이용한 이상탐지 방법론이다. 사분위수(quartiles)란 최댓값과 최솟값을 기준으로 데이터의 분포를 25%, 50%, 75%의 사분위로 나눈 점을 말하며, 각각을 높은 사분위수(upper quartile, Q3), 중앙값(median, Q2), 낮은 사분위수(lower quartile, Q1)라 부른다. 즉, Q1은 최솟값과 중앙값 사이의 중간점이며, Q3은 최댓값과 중앙값 사이의 중간점이다. 사분위간 범위는 Q1과 Q3 사이의 데이터 중간 50%에 대한 범위로 수식으로는 이라 정의한다.

20) 문맥적 이상(contextual anomaly)은 시간 자료(Weigend et al., 1995; Salvador & Chan, 2003)와 공간 자료(Kou et al., 2006; Shekhar et al., 2001)에서 가장 흔하게 찾을 수 있다(Oh, et al., 2018).

21) Longest Common Substring (LCS), Damerau-Levenshtein거리, Jaro거리, Jaro-Winkler거리, Needleman-Wunsch거리, Smith-Waterman거리, n-gram, syntactic n-gram(Majumder, et al., 2016)

22) 마르코프 체인 속성은 미래 상태의 확률은 현재 상태에만 의존하며 그 이전의 이벤트에는 의존하지 않는 것을 말한다.



source: Edited by author referring to Yahalom(2008) and Guzman(2018)

〈그림 4〉 마르코프 체인 속성을 활용한 n-gram 벡터 생성  
 〈Fig. 4〉 N-gram Vector Generation Using Markov Chain Properties

Understanding)를 보정하여 유사성을 보다 정밀하게 측정할 수 있다는 장점이 있다(Kim & Son, 2018). 이 때, 연속된 문자열  $n$ 의 길이가 길수록 더 많은 문맥을 포함하여 정보가 명확해진다. 커뮤니케이션 이론, 자연 언어 처리, 생물학적 서열 분석, 데이터 압축, 철자 및 문법 교정, 음성 인식, 자동 번역, 자동 요약, 챗봇 등 광범위 하게 활용되고 있다.

본 연구에서는 내용분석은 다음 5단계를 따른다. 1) 전처리 단계로 특수문자, 숫자, 영문자와 같은 불용어를 제거한다. 2) 데이터에 존재하는 모든  $n$ -grams을 이용하면 차원이 매우 커져 계산이 어렵기 때문에 문자열  $S_1=(t_1 \cdots t_n)$ 에 대해 마르코프 체인 알고리즘을 적용하여 전후 단어  $t_{n-1}$ 와  $t_n$ 의 전이확률을 계산하여 이를 재조합하여 유의미한  $n$ -gram 벡터  $\vec{V}_1=(t_1, t_2, \dots, [t_{n-1}, t_n])$ 를 생성한다. 3) 한국어 형태소 분석을

통하여 명사 추출, 이때 단어빈도(Term-Frequency, TF)를 기준으로  $n$ -gram 벡터값  $\vec{V}_1^N=(t_1, \dots, t_k)$ 을 결정한다. 4) 코사인 각도 거리(Cosine Angle Distance)<sup>23)</sup>를 활용하여 유사도를 계산, 5) 다차원척도(Multi-Dimensional Scaling, MDS)를 활용하여 2차원 공간의 분포와 밀집의 정도를 시각화하여, 밀집 정도가 높은 경우 집단민원으로 판단한다.

‘주기성’에 대해서는 동일 키워드를 가진 민원 사이의 시간차를 계산하여 주기성이 높은 경우를 이상치(집단민원)으로 판단 할 수 있으나, 어느 정도의 시간차를 주기성이 있다고 판단할 것인지에 대한 명확한 규정이 아직 없어, 본 연구에서는 분석 대상에 포함하지 않는다. 또한 ‘다수인 관련성’은 개인정보 비식별 조치로 본 연구에서는 분석 대상에 포함하지 않는다.

23) 코사인 각도 거리는 두 벡터  $S_1=(t_1, \dots, t_n)$ 과  $S_2=(t_1, \dots, t_m)$ 간 각도의 코사인 값을 의미하며, 각도가 0°일 때, 즉 방향이 완전히 일치할 경우의 코사인 값은 1을 갖는다. 다차원의 유사도 측정에 유용하여 텍스트 유사도를 측정하는 데 많이 사용된다.

## IV. 사례분석: 창원시 전자민원

### 1. 창원시 시민의 소리

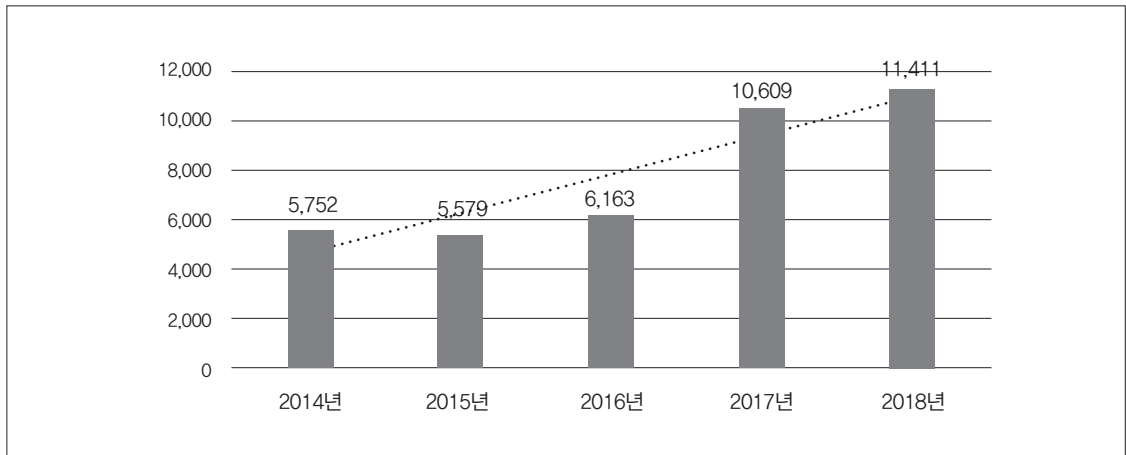
창원시 시민의 소리(Voice of Citizen, VOC)는 시민들이 시정 건의 사항과 시민불편사항 등에 대해서 게시할 수 있도록 한 참여형 온라인 창구로 정식 민원은 아니지만 7일 이내에 답변 처리를 하도록 규정

전자민원의 한 유형이라고 할 수 있다.

2004년 2,703건을 시작으로 2010년 통합창원 출범 이후 계속 증가, 2017년에는 1만 건을 넘겨, 2018년까지 15년간 약 2배 정도 늘었다.

### 2. 박스플롯을 활용한 시계열 이상치 탐지

본 분석에는 시간, 공개/비공개 상태, 민원내용의

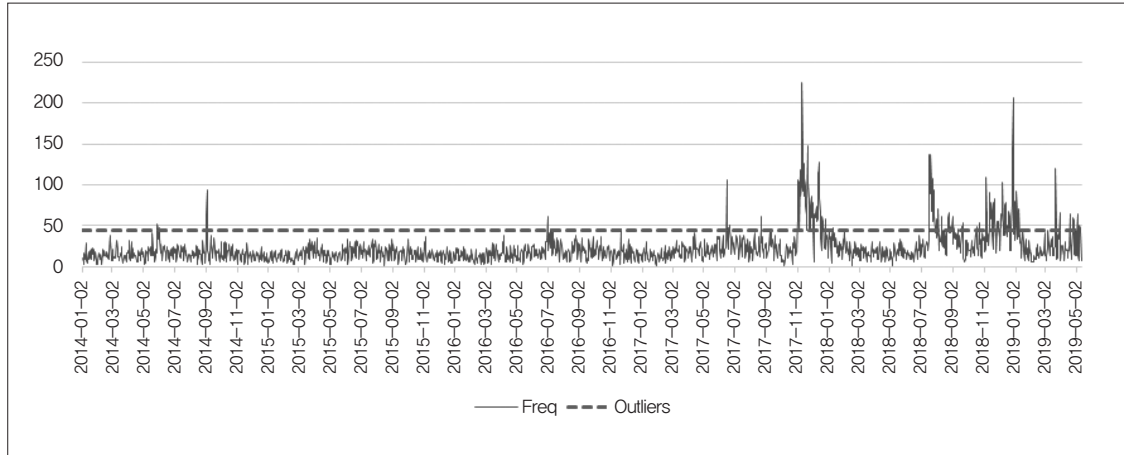


〈그림 5〉 창원시 시민의 소리 추이 (2014~2018년)  
 〈Fig. 5〉 Increasing VOC in Changwon City (2014 ~ 2018)

〈표 12〉 박스 플롯을 이용한 이상치 탐지  
 〈Table 12〉 Outlier Detection Using Box Plot

	Freq	Remarks
Q1	12	25% IQR
Q3	25	75% IQR
IQR	13	Q3-Q1
Upper Outliers	44.5	$Q3+(1.5 \times IQR)$

	Days	%	Number of Petitions	%
Total	1,951	-	42,880	-
Upper Outliers	153	7.8	10,937	25



〈그림 6〉 창원시 시민의 소리 전자민원 시계열 이상치(1일 단위)  
 〈Fig. 6〉 Time Series Outliers of Changwon VOC E-petitions (per day)

세 가지 변수가 사용되었으며, 수집된 데이터는 총 42,880건이다. 이 중 공개가 58%(24,917건)로 비공개 42%(17,963건)보다 상대적으로 조금 더 많은 비중을 차지한다. 1일 평균 21.98건, 최대 225건(2017년 11월 10일) 제기되었다(〈부록 1〉).

박스 플롯을 활용하여 시계열 이상치를 확인하였다. 그 결과 1일 민원제기 건수가 값 보다 큰 153일분 10,937건을 이상치로 판단하였다(〈표 12〉). 이는 전체 분석일 수의 7.8%, 전체 민원수의 25%로, 매우 높은 비율을 차지한다. 즉 참여형 전자민원이 집단민원 제기에 매우 취약한 구조를 가졌다는 것을 의미한다.

〈그림 6〉에서 보듯이, 이상치가 간헐적으로 나타나는 경우도 있지만, 주로 2017년 11월, 12월, 2018년 7월, 11월, 12월과 같이 일정 기간에 집중하여 연속적으로 나타나며 주 200건을 넘는 것을 확인할 수 있다. 이는 국민권익위에서 규정하는 집단민원의 확산 징후

24)에 해당하는 것으로, 집단민원을 조기에 파악하지 못하면 견잡을 수 없이 확산될 우려가 높다. 조기 예견이 매우 중요하며, 박스플롯의 시계열 이상치 탐지를 통해서 연속성 여부를 기준으로 집단민원의 가능성을 조기에 예견할 수 있음을 시사한다.

### 3. 내용분석을 통한 집단민원 추정

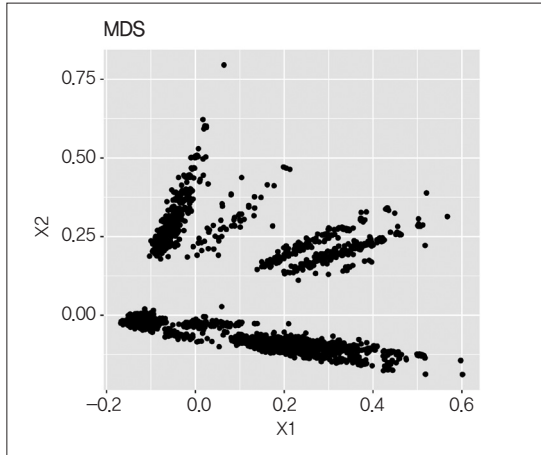
앞서 시계열 이상치로 판단 된 10,937건 중 내용분석이 곤란한 민원<sup>25)</sup> 6건과 비공개 민원 4,851건을 제외 한 총 6,080건에 대해 의미적 유사도를 분석하여 집단민원을 추정했다. 각 민원을 30-gram 벡터로 변환 한 후, 명사를 추출하였다. 그 결과 벡터의 차원 수는 14,769, 크기 6,080×6,080=36,966,400의 매트릭스를 만들었다(〈부록 2〉).

분석 결과 V5244, V5174, V5195와 같이 n-gram

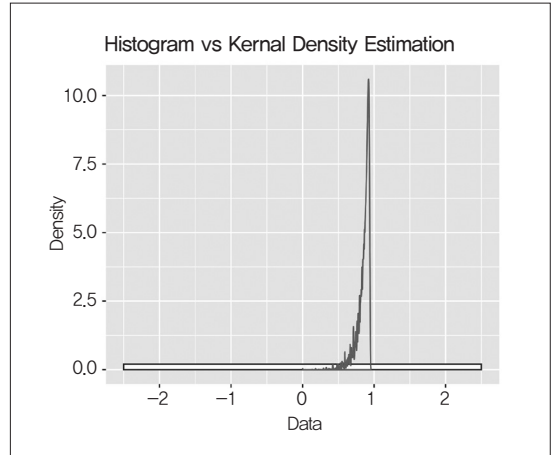
24) 국민권익위원회에서는 2014년 3월부터 「민원확산 조기경보」제도를 도입·운영하고 있는데, 2주 이상 지속하여 나타나며 주 100건을 초과하는 피해특이민원과 주 200건을 초과하는 갈등민원의 경우 '주의'를, 4주 이상 지속하여 나타나며 주 200건을 초과하는 피해특이민원과 주 400건을 초과하는 갈등민원의 경우 '심각'의 경보를 발령한다(Daegu, 2015).

25) 참여형 민원은 개인의 문맥에 따라 자유롭게 기술하기 때문에 내용 분석이 곤란한 유형이 존재한다. 주술구성이 포함되지 않은 민원, 내용 없이 공백으로만 작성된 민원, 특수문자만 포함된 민원 등이 해당한다.





〈그림 7〉 민원내용의 유사도  
〈Fig. 7〉 Semantic Similarity



〈그림 8〉 히스토그램과 밀도  
〈Fig. 8〉 Histogram and KDE

이 완전히 일치하는 민원이 총 124건 확인되었다. 이를 다시 주제별로 확인 해 본 결과 7가지의 집단민원이 추정되었으며, 추정된 민원이 실제 집단민원인지 확인하기 위하여 창원시 다량민원계시판<sup>26)</sup>과 비교 한 결과, 4가지 추정 집단민원이 실제 집단민원과 같았다(〈부록 3〉). 이는 대부분 인·허가와 관련된 복합민원이다. 창원시 다량민원계시판에서는 “공짜 관료 집단”과 같은 공무원에 대한 불평민원은 반복성과 주기성이 있음에도 불구하고 집단민원으로 처리하고 있지 않았다. 집단민원에 대한 합의와 행정적 대응책에 대한 기준마련이 시급하다.

다음으로 유사도가 1인 경우(=완전 불일치)를 제외한 3,056,939건에 대해 2차원 분포를 〈그림 7〉에 시각화하였다. 거리가 가까울수록 민원내용이 유사하다는 것을 의미하며, 거리가 멀수록 민원내용이 유사하지 않다는 것을 의미한다. 따라서 거리가 가까운 민원의 그룹을 집단민원으로 추정할 수 있다.

본 연구에서 추정된 집단민원은 n-gram이 완전히 일치하는 경우로, 그 외 내용의 동일성을 인정할 수 있

는 유사도의 기준(=거리의 크기)은 아직 명확히 분석되지 않았다. 향후 면밀한 분석을 통해서 유사도의 기준을 객관화하는 과제가 남아있다.

## V. 결론

전자집단민원 추정은 참여형 전자민원 시스템의 근간이 되는 매우 중요한 선결 과제이다. 본 연구에서는 선행연구를 검토하여 전자집단민원의 개념과 특징을 정의하고, 방대한 참여형 전자민원 데이터에서 전자집단민원을 객관적으로 추정하기 위하여 이상탐지와 내용분석기법을 응용한 방법론을 제안하였다. 그리고 이를 창원시 사례에 적용하여 7가지의 전자집단민원을 추정하였으며 실제 집단민원과 비교하여 4가지의 전자집단민원이 동일한 내용으로 확인되어 방법론의 유의미성을 검증하였다.

본 연구의 정책적 시사점은 다음 세 가지로 요약할 수 있다. 첫째, 참여형 전자민원의 경우 집단민원 계기에 매우 취약한 구조를 가지고 있기 때문에 집단민원

26) URL(<https://www.changwon.go.kr/mayor/mass/voice/list.do?mId=0306000000>)

을 조기에 예견하는 것이 중요하며, 집단민원의 조기 예견을 위해서 다부처간 민원 내용의 공유와 민원 간 인과관계를 분석할 수 있는 종합적이고 장기적인 전자 집단민원관리시스템이 필요하다. 둘째, 집단민원의 유형은 인·허가와 관련된 복합민원이 대부분으로, 선행 복합민원에 대한 처분과 집단민원과의 인과관계를 분석하여 집단민원 예견 후 대응 방향 검토를 위한 집단 민원 확산 징후가 있는 복합민원에 대한 종합대책지침 마련이 필요하다. 셋째, 행정에 대한 단순불평반복과 같이 집단민원으로 처리하고 있지 않아 전자집단민원에 대한 행정적 정의에 대한 합의와 동시에 이를 수행하기 위한 행정조직의 역할 분담과 소관부처 간 원활한 연계를 통한 전자집단민원 처리의 효율성 제고를 위한 제도개선이 필요하다.

아울러 참여형 전자민원의 단점으로 개인 문맥에 대한 의존도가 매우 높아 문장의 정확성이 낮은 경우가 많이 관찰된다. 이러한 정확성이 낮은 민원은 기간 내에 처리하기가 매우 곤란하므로, 이에 대한 처리기준 마련이 시급하다. 그리고 요구사항의 정확성을 제고하기 위하여 육하원칙에 따라 내용을 기술할 수 있도록 사용자 인터페이스를 개선하고 '문장 자동완성', '유사 문장 검색' 등 보조적 기능을 제공하여 민원작성이 용이하고 정확하게 이루어질 수 있도록 하는 노력이 필요하다. 또한 창원시 외 지역에서 제기되는 민원에 대해서는 관할 이송 등의 처리기준을 마련하는 것이 필요하다. 향후 과제는 추정 결과에 대한 신뢰성 보장을 위하여 전문가에 의한 크로스체크 등 다각도의 검증 방법을 개발하는 것이다.

본 연구는 전자민원 시스템 내 집단민원의 부작용을 줄이는데 기여할 수 있을 것으로 기대된다. 특히, 키워드 확인과 유사 민원탐색, 각 집단민원의 특성에 따른 관리지침 마련, 전자집단민원과 미해결민원과의 상관관계 분석을 통한 미해결민원에 대한 처리방안 마련 등 실무적으로 기여할 수 있다. 또한 추정된 전자집단 민원을 둘러싼 이해관계자들의 대립 구도와 다양한 사회 집단의 분화 현상을 객관적이고 체계적으로 설명할

수 있으며, 다양한 분야에 확대 적용하는 등 학술적 기여가 높을 것으로 기대된다.

## ■ References

- ACRC (2013). *A Study on Reasonable Management of Public Sector Chronic Petitions by Reviewing Overseas Cases*. Anti-Corruption & Civil Rights Commission.
- {국민권익위원회 (2013). 고질민원 해외사례 고찰을 통한 공공부문고질민원에 대한 합리적 관리방안 연구, 국민권익위원회.}
- ACRC (2019a). [Two years of government launch] ③ National participation and communication policy culture settled in .... Reinforcing the resolution of national discomfort based on "Civil Petitions Big Data", Anti-Corruption & Civil Rights Commission(May 23, 2019. released).
- {국민권익위원회 (2019a). ③ 국민 참여·소통 정책문화 자리 잡아... '민원 빅데이터' 기반 국민불편 해소 강화, 국민권익위원회 (2019.5.23.배포).}
- ACRC (2019b). *Sinmungo Petition Data: Q1 2019*. Anti-Corruption & Civil Rights Commission.
- {국민권익위원회 (2019b). 국민신문고 민원 통계자료 (2019년 1분기).}
- ACT No. 15617 (2018). *Act on the Prevention of Corruption and the Establishment and Management of the Anti-Corruption and Civil Rights Commission*.
- ACT No. 14839 (2017). *Civil Petitions Treatment Act*. {민원처리법 (2017). (법률 제14839호, 2017. 7. 26, 제2조 제6항).}
- Aggarwal, C. C. (2013). "Proximity-Based Outlier Detection, Outlier Analysis." in *Outlier Analysis*, 101-133, Berlin, Heidelberg: Springer.
- Daegu (2015). *A Study on the Development Direction of Big Data and Cloud-based Civil Service: Focusing on Civil Service Proposal Integrated Management System*. Daegu Metropolitan City.

- {대구광역시 (2015). <빅데이터와 클라우드기반의 민원행정 서비스 발전방향에 대한 연구: 민원제안 통합관리 시스템 구축을 중심으로>. 대구광역시.}
- EGOV24 (2008). E-Petitions G4C Introduction, Ministry of the Interior and Safety.  
{전자정부24 (2008). 전자민원G4C 소개자료, 행정안전부.}
- EGOV24 (2019). E-Petition Service Status. Ministry of the Interior and Safety.  
{전자정부24 (2019). 전자민원 서비스 현황, 행정안전부}
- Eun, J. (2006). "The Roh Administration's Strategies of Government Innovation: Continuity and Discontinuity." *Korean Journal of Public Administration*, 44(3), 207-247.
- {은재호 (2006). 참여정부 정부혁신의 평가와 과제: 연속성과 비연속성을 중심으로. <행정논총>, 44권 3호, 207-247.}
- Guzman, R. (2018). "Sampling: Markov Chains.", <https://relguzman.blogspot.com/2018/04/sampling-markov-chains.html>. (Retrieved on November 1, 2019).
- Hawkins, D. (1980). Identification of Outliers, Chapman and Hall.
- Jeong, H., Hatori, T. & Kobayashi, K. (2007). Discourse Analysis of Public Debates: a Corpus-Based Approach. Proceedings of the IEEE Systems, Man, and Cybernetics Conference, Montréal, Canada, 7-10 October: 1782-1793.
- Jeong, H. (2008). *A Study on Public Deliberation in Provision of Social Overhead Capital*, Kyoto University (in Japanese).
- Jeong, H., Kobayashi, K., Hatori, K. & Shiramatsu, S. (2010). "Facet Decomposition and Discourse Analysis of Public Debate." *Journal of Japan Society of Civil Engineers, Ser. F (Construction and Management)*, 66(1): 45-56 (in Japanese).
- Jeong, H, Lee, T. & Hong, S. (2017). "A Copus Analysis of Electronic Petitions For Improving the Responsiveness of Public Services: Focusing on Busan Petition." *The Korean Journal of Local Government Studies*, 21(1), 423-436.
- {정하영·이태현·홍순구 (2017). 공공서비스의 수요대응성 향상을 위한 전자민원 코퍼스분석: 부산시 사례를 중심으로. <지방정부연구>, 제21권 1호, 423-436.}
- Jeong, H., Shiramatsu, S., Kobayashi, K. & Hatori, T. (2008). "Discourse Analysis of Public Debates Using Corpus Linguistic Methodologies." *Journal of Computers*, 3(8): 58-68
- Jung, H. (2006). "A Study on Efficient Solution for Collective Petitions- Centering on Chungcheongnam-do Province." *The Korea Local Administration Review*, 20(4), 83-108.
- {정하용 (2006). 집단민원의 효율적 해결방안 연구. <지방행정연구>, 20권 4호, 83-108.}
- KCC (1985). *Petitoin's Management Guidelines, The Korea Coal Corporation (1992 revised)*.
- {대한석탄공사 (1985). <민원업무관리지침(1992년 개정)>.}
- Kim, K. & Son, Y. (2018). "Change Acceptable In-Depth Searching in LOD Cloud for Efficient Knowledge Expansion." *The Journal of Intelligence and Information Systems*, 24(2), 171-193.
- {김광민·손용락 (2018). 효과적인 지식확장을 위한 LOD 클라우드에서의 변화수용적 심층검색. <지능정보연구>, 24권 2호, 171-193.}
- Lee, T., Hong, S. & Jeong, H. (2017). "A Time Series Analysis on the Daily Media Corpus of Public Project in Local Government: Focusing on the Gijang Seawater Desalination Plant Project", *The Korean Journal of Local Government Studies*, 20(4), 1-23.
- {이태현·홍순구·정하영·홍혜승 (2017). 지방정부 공공 사업에 관한 일간지 미디어자료의 시계열 코퍼스 분석: 기장 해수담수화플랜트사업을 중심으로. <지방정부연구>. 20권 4호, 1-23.}
- Majumder, G., Pakray, P., Gelbukh, A. & Pinto, D. (2016). "Semantic Textual Similarity Methods, Tools, and Applications: A Survey." *Research Advances and Applications of Evolutionary Computation*, 20(4), 647-665.
- Min, H., Park, J., Lee, D. & Kim, I. (2014). "Outlier Detection Method for Mobile Banking with User Input Pattern and E-finance Transaction

- Pattern." *Journal of Internet Computing and Services(IJCS)*, 15(1), 157-170.
- {민희연·박진형·이동훈·김인석 (2014). 사용자 입력 패턴 및 전자 금융 거래 패턴을 이용한 모바일 뱅킹 이상치 탐지 방법. <인터넷정보학회논문지>, 제15권 1호, 157-170.}
- Modi, K. & Oza, B. (2016). "Outlier Analysis Approaches in Data Mining." *IJIRT*, 3(7), 6-12.
- Oh, M., Park, A., Kim, Y. & Jin, J. (2018). *A Study on anomaly detection based on Machine Learning*. Sejong: Korea Institute for Health and Social Affairs.
- {오미애·박아연·김용대·진재현 (2018). <기계학습(Machine Learning) 기반 이상 탐지(Anomaly Detection)기법 연구: 보건사회 분야를 중심으로>, 세종: 한국보건사회연구원.}
- Oh, S. (1984). "A Study on Civil Petition Administration." *Korean Journal of Public Administration*, 22(1), 1001-1023.
- {오석홍 (1984). 민원 행정에 관한 연구. <한국행정학보>, 제22권 1호, 1001-1023.
- PSIIF (2005), Citizen Participatory Petition System Improvement Manual, Petitions System Improvement and Innovation Forum supported by The Korea Institute of Public Administration.
- {민원제도개선혁신포럼 (2005). 시민참여형 민원·제도개선 매뉴얼, 한국행정연구원 지원.}
- Rahman, M.S. & Brishty, F. P. (2018). "Anomaly detection of river data for disaster prevention." *The International Journal of Engineering and Science (IJES)*, 7(9.1), 59-68.
- Richardson, C. (2009). Virus detection with machine learning. University of Bristol.
- Scikit-learn (2019). Novelty and Outlier Detection (2019-07-07 accessed).
- Tukey, J.W. (1977). *Exploratory data analysis*, Addison-Wesley.
- Watson, M. R., Shirazi, N., Marnierides, A. K., Mauthe, A. & Hutchison, D. (2016). "Malware Detection in Cloud Computing Infrastructures." *IEEE Transactions on Dependable and Secure Computing*, 13(2), 192-205.
- WCPP (2019). How to submit a petition, Western cape provincial parliament.
- Wilcox, R. R. (1996). *Statistics for the social sciences*. Academic Press.
- Yahalom, S. (2008). "URI Anomaly Detection Using Similarity Metrics." Master's Thesis, School of Computer Science, Tel-Aviv University.
- Yeon, J., Shim, J. & Lee, S. (2013). "Outlier Detection Techniques for Biased Opinion Discovery." *The Journal of Society for e-Business Studies*, 18(4), 315-326.
- {연종흠·심준호·이상구 (2013). 편향된 의견 문서 검출을 위한 이상치 탐지 기법. <한국전자거래학회지>, 제18권 제4호, 315-326.}
- Yi, S., Rhew, S. & Lee, K. (2009). "A Study on the Construction of the System Integration Management System for Civil Affairs." *Journal of the Korea Society of Computer and Information*, 14(8), 107-109.
- {이성기·류성열·이규원 (2009). 민원업무통합관리시스템의 구축방안 연구. <한국컴퓨터정보학회 논문집>, 14권 8호, 107-109.}
- Yoon, Y. (1997). *Information Society and Teledemocracy: Issues and Prospects*. in Lee, H. & Choi, J. (ed.) *Utopia of communication? : Informatization and Korean Life*, 23-40. Kyunggi: Nanam Publishing.
- {윤영철 (1997). "정보사회와 텔레데모크라시: 쟁점과 전망." 이현조, 최정호(편). *커뮤니케이션의 유토피아? : 정보화와 한국인의 삶*, 23-40, 경기: 나남출판.}

[Appendix]

〈부록 1〉 창원시 시민의 소리 데이터 셋의 개요  
 〈Appendix 1〉 Outlines of Changwon VOC Dataset

Division		Details
Crawling	Period	From 2014-01-02 to 2019-05-10
	# of days	1,951
Variables	Time	yyyy-mm-dd, hh:mi:ss
	Status	Open or Close
	Content	Text
Dataset Size	variables × # of petitions	3 × 42,880
[1] "2014-01-02, 16:47:32", "공개", "살기좋은 창원시를 만들기 위해 매일 고생많으신 공무원 여러분의 올 한해 행복이 가득하길 바랍니다....(중략)..아이들을 많이 낳고 살수 있도록 창원시가 다시한번 다자녀 가족에게 힘을 주세요." ∴ [42880] "2019-05-10, 08:16:24", "비공개,"		
Descriptive Statistics	# of E-petitions per day	min:1, mean: 21.98, max: 226
	# of open and close	open: 24,917, close: 17,963

〈부록 2〉 N-gram 의미적 유사성  
 〈Appendix 2〉 N-gram Similarity

ID	col	row	similarity	30-grams
1	V2613	V898	0.95	V2163: 이런거 유치하고싫음 당신들양심이나 좀지키면서 도로에 불법 주정차 좀하지 말고 그리고 이관거 유치 해달라하소 불법주차나 좀 하지 마세요 죄다 유니스타 분양 받아가 자기들 아파트값 올리려고 혈안이데가 스타필드 유치 하라고 하지요 아파트값 V898: 천국, 지하주차장은 물난리,곰팡이, 누수 외부에 는 크랙, 난간 부착 부실, 울타리 절단등새 아파트가 현아파트로 둔갑되어 있구나하자보수는 하고 있는건지 알수도없고시간이 갈수록 늘어나는건 입주민 한숨뿐 눈가리고 아웅 하는 덕산건설에 기가막혀 헛 웃음뿐언제쯤 하자보수 마무리되어 V2618: 방면 창원 테미널방면 소담시장길 농협에 서 부터 소담금호온천까지 도계동 한마음 아파트부터 의창구청 까지 봉곡동 히엔리병원부터 지귀상가 삼거리 까지 반림동 현대아파트 주변 반송시장 주변 반송 주민센터주변 봉곡동상가입구 봉림중학교 봉림휴먼시아입구삼거리 주변 등등 이무슨 도로가
...	...	...	...	...
3056938	V5244	V5174	-2.22e-16	V5244, V5174, V5195: 기원 합니다 기원 합니다 스타필드 유치를 기원 합니다 기원 합니다 스타필드 유치를 기원 합니다 창원 스타필드 유치를 기원 합니다 기원 합니다 기원 합니다 기원 합니다
3056939	V5244	V5195	-2.22e-16	

〈부록 3〉 추정 집단민원과 실제 집단민원 비교  
 〈Appendix 3〉 Estimated Petitions vs. Actual Petitions

Match	Estimated Collective E-petitions	Actual Collective E-petitions
○	유니시티 중앙공원 화장실(7)	유니시티 아파트 3단지 앞 중동 중앙공원 화장실 철거 (220)
○	스타필드 유치(67)	스타필드 창원 관련하여 (6958)
○	북면 변전소 건설 반대(2)	북면지역 송전선로 건설 반대 (333)
○	아내 아파트 누수 준공 승인 문제(14)	감계 덕산아내2차 아파트 준공 관련 (1245)
×	진해구 식당 영업 어려움(2)	진해구 자은동 협성아이존빌 아파트 사전점검 관련 민원 (182)
×	대로 불편(16)	진해구 청안동 부영3차 아파트 분양전환 승인요청 (212)
×	공짜 관료 집단(16)	제덕 서희스타힐스 지역주택조합 관련 (29)
×	-	창원NC파크 마산구장 (148)