

의미연결망 분석 방법을 활용한 국립자연휴양림 민원 특성 분석

Analyzing National Recreation Forest user complaints using Semantic Network analysis method

김경아¹, 이정희², 서경원², 지상훈²

¹서울과학기술대학교 데이터사이언스학과, ²국립산림과학원 산림휴먼서비스연구과

1. 연구 배경 및 목적

- 산림청에서는 전국 46개 국립자연휴양림을 조성·운영하고 있음. 그중 국립자연휴양림은 국민이 가장 선호하는 산림복지시설(이용의향 60.0%)로 숲나들e 예약 시스템을 통해 매년 약 400만명이 이용하고 있음.
- 2019년까지는 개별적인 자연휴양림 예약체계가 운영되었으나, 2019년 6월부터 단계적으로 전국 국/공/사립 자연휴양림이 숲나들e라는 통합 플랫폼으로 통합되어 운영되고 있음. 통합된 체계에서 전국 자연휴양림 이용 관련 정보가 수집 및 누적되고 있는 가운데, 자연휴양림 이용 현황을 진단하고 개선점을 도출하여 이용객 만족도를 제고하려는 노력이 필요함.
- 수집되고 있는 자연휴양림 이용 정보는 그 활용가치는 크나, 비정형 자료이기 때문에 적절한 전처리 과정과 분석 방법이 적용되어야 합리적인 시사점 도출이 가능함. 이에 따라 텍스트 분석에 활용되어온 의미연결망 분석 방법론을 적용해 체계적이고 과학적인 이용 현황 분석을 시도해보고자 함.
- 이를 위해 워드클라우드 분석과 CONCOR 분석 방법을 적용하였으며, 각각 1차, 2차 분석 과정으로 수행함

2. 연구 자료

- 자연휴양림 이용 관련 문의 및 민원 정보는 국민신문고, 숲나들e 1:1문의 게시글, 숲나들e 고객센터 이렇게 세 가지 경로로 수집이 가능함
- 본 연구에서는 숲나들e 통합 예약체계가 운영된 2019년 6월부터 2023년 11월까지 등록된 자연휴양림 이용 관련 민원 내역을 취득하여 분석함
- 국민신문고 건은 민원 홈페이지(<https://www.epeople.go.kr/index.jsp>)에서 수집되었음
- 숲나들e 1:1문의 게시글은 [숲나들e 홈페이지>참여마당>1대1문의]에서 수집되었음
- 숲나들e 고객센터 문의 정보는 고객센터(1588-3250)로 접수된 상담 내용을 정리함
- 수집된 데이터 현황은 아래 표1과 같음

표1. 자연휴양림 이용 관련 문의 및 민원 정보 수집 현황

연도\항목	숲나들e 1대1	숲나들e 고객센터	국민신문고	합계
2019	3,639	225	42	3,906
2020	4,009	284	31	4,324
2021	4,039	385	76	4,500
2022	3,638	355	48	4,041
2023	3,076	352	126	3,554
합계	18,401	1,601	323	20,325

(단위: 건)

- 수집된 비정형 데이터는 분석에 활용될 수 있도록 ①데이터의 통합, ②문장 분리, ③형태소 분석, ④데이터 정제 과정을 통해 전처리하였음
- 데이터 통합 및 문장 분리를 통해 가공 및 분석에 적합한 명사 형태로 1차 가공하였음
- 형태소 분석은 KoNLPy라이브러리의 Okt 형태소 분석기를 활용하였고, 문장 내 형태소를 정규화하고 변형된 형태가 아닌 원형의 형태소를 추출하였음
- 분석에 불필요한 휴양림 관련 단어와 불용어는 연구자의 주관으로 삭제하였음

3. 분석 방법

- 1차 워드클라우드 분석은 전처리가 완료된 명사 빈도를 대상으로 하여 단어 빈도가 높은 순서대로 상위 300개의 명사를 워드클라우드로 시각화하는 방식으로 진행함. 시각화에는 Python 언어를 활용, WordCloud 라이브러리를 사용하였음.
- 2차 CONCOR 분석은 데이터 내 명사형 단어들의 동시출현 빈도를 계산하였음. 이때 활용된 단어 목록으로는 빈도수 기준 상위 50개로 하였으며(이수상, 2012), 동시 출현빈도를 계산하여 단어동시출현행렬을 생성하였음.
- 이후에는 UCINET 프로그램을 활용, CONCOR분석을 통해 명사 간 연관성을 분석하였음
- UCINET은 사회연결망 분석을 위한 대표적인 프로그램으로, 주로 사회과학에서 사회측정조사 데이터를 분석하는 데 사용됨(Borgatti et al., 2014).
- CONCOR(CONvergence of iteration CORrealtion) 분석은 사회연결망 분석방법 중 하나로써, 전체 데이터 내에서 유사한 맥락과 연결구조패턴을 갖고 있는 단어들끼리 그룹화하는 기법을 말함(한국행정연구원, 2022)

4. 결과



그림1. 자연휴양림 이용객 민원 내용 워드클라우드 시각화

- 워드클라우드 시각화 결과, 예약/건의/요청/불만 등의 형태소 군집이 나타남. 출처 유형별로 차이가 나타나는데, 숲나들e 고객센터의 경우 불만/건의 등의 형태소가 주로 나타나다가 2022년부터 그 비중이 줄어들고 예약/요청 등의 단어로 전환된 것을 확인할 수 있음. 숲나들e 1대1 게시글 문의의 경우 예약이라는 형태소가 주요하게 나타나는 것으로 보아, 예약 관련 서비스 이용 상황에서 민원이 집중됨을 확인할 수 있음. 국민신문고의 경우, 요청/이용 등의 키워드가 나타나며, 서비스 개선 요청과 관련한 단어들을 확인할 수 있음

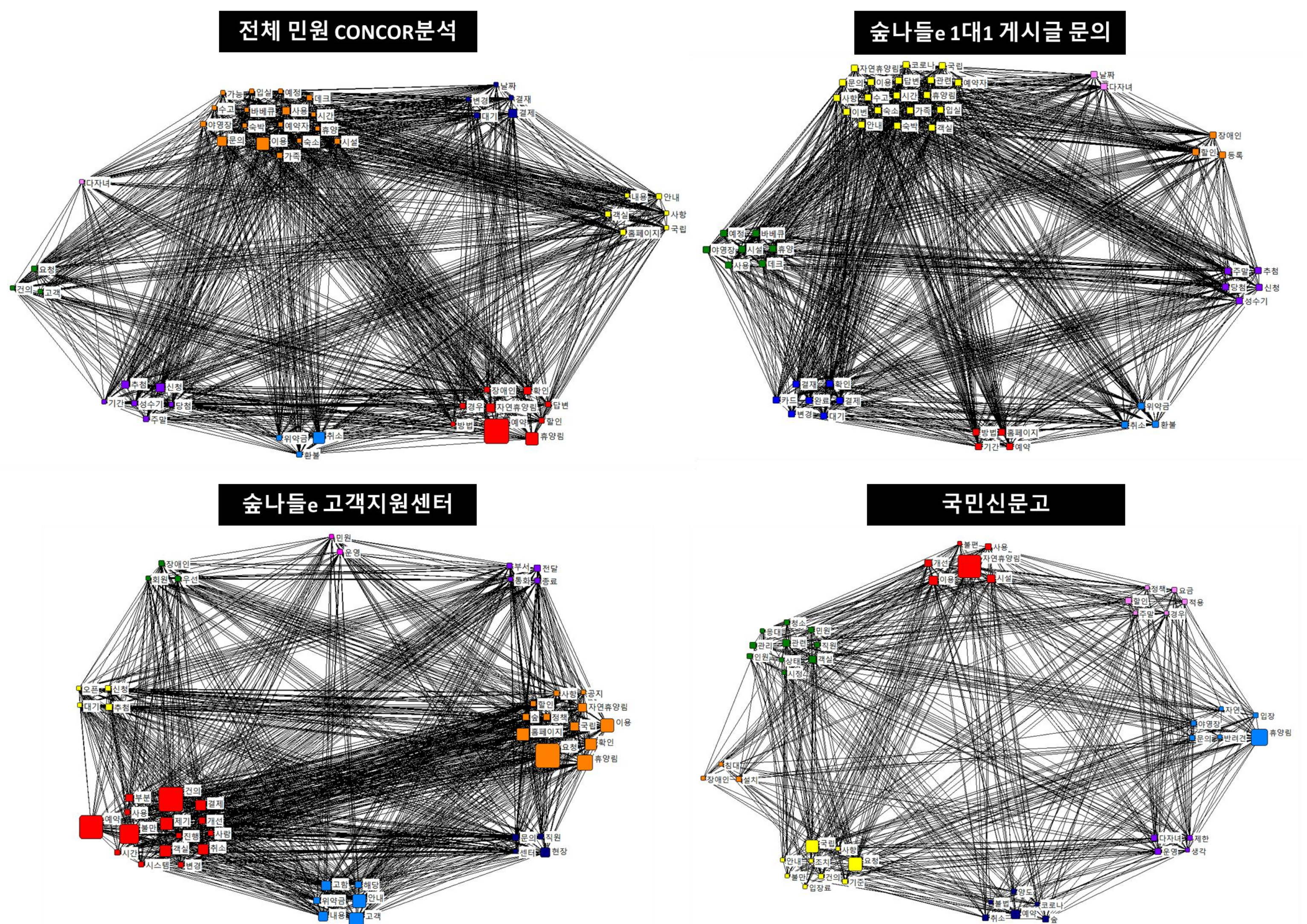


그림2. CONCOR 분석 시각화 결과

(좌상) 민원 전체 / (우상) 숲나들e 1대1 게시글 문의 / (좌하) 숲나들e 고객센터 / (우하) 국민신문고

- CONCOR 분석 결과, 각 출처 유형별로 주요 형태소들을 확인할 수 있고, 이와 동시에 출현 빈도가 높은 단어들을 확인할 수 있음. 워드클라우드 분석 결과 불만/건의의 관련 단어가 주요어로 나타났던 숲나들e 1대1 민원 유형의 경우, 객실/결제/취소/변경/시스템 등의 주변어와 함께 등장하는 경향이 있음을 확인할 수 있음.

5. 결론

- 분석결과, 민원유형은 크게 ①자연휴양림 내부 이용 시설 이용 문의, ②객실 예·결제 문의, ③장애인 및 다자녀 할인 문의 등 세가지로 구분됨
- 이와 관련하여 시설 이용에 대한 정확한 정보 전달 및 권역별 예약 일자 조정 등 시스템의 유연한 운영을 통한 대기시간 단축, 보호자 대리 예약 허용 등 현장 및 시스템 정비가 필요할 것으로 판단됨.
- 연도별로 나타나는 추이에 변화가 나타나 이용객들의 민원 형태의 변화를 확인할 수 있음.
- 본 연구에서 활용된 연구방법 및 데이터는 향후 머신러닝 기술을 적용한 감정 분석 등의 추가적인 분석방법과 결합을 시도해 볼 수 있을 것으로 기대됨. 더불어 시기별로 누적해 그 추이를 살펴봄으로써 시계열적인 이용 만족도 및 민원 유형 분석을 진행할 예정

참고 문헌

- 이수상, 2012, 네트워크 분석 방법론, 논형출판사
- 한국행정연구원, 2020, KIPA조사포럼
- Borgatti, S.P., Everett, M.G., and Freeman, L.C. 2008, Encyclopedia of Social Network Analysis and Mining. Springer, New York, NY. https://doi.org/10.1007/978-1-4614-6170-8_316

