

K-Modes 군집화 알고리즘을 활용한 인바운드 관광객 유형 및 특성 분석

최가연¹, 김도연², 홍수민³, 김현희¹

¹ 동덕여자대학교 정보통계학과

² 동덕여자대학교 국제경영학과 학부생

³ 동덕여자대학교 경영학과 학부생

chlrkdus26@gmail.com, kayadanu9@gmail.com, dbal00900@gmail.com, heekim@dongduk.ac.kr

Analysis of Inbound Tourist Types and Characteristics Using K-Modes Clustering

Gayeon Choi¹, Doyeon Kim², Soo min Hong³, Hyon Hee Kim¹

¹Dept. of Statistics and Information Science, Dongduk Women's University

²Dept. of International Business Administration, Dongduk Women's University

³Dept. of Business Administration, Dongduk Women's University

요 약

외국인 관광객은 증가하고 있지만 관광 수지는 감소하고 있다. 본 연구에서는 관광객을 군집화하여 관광객별 유형을 분석하고, 뉴스 기사 분석을 통해 한국 관광 관심사를 분석하여 인바운드 관광 산업에 도움이 되고자 하였다. 연구 결과 한국을 방문하는 외국인 관광객은 업무수행, 문화체험, 쇼핑 유형으로 나뉘고, 이들의 관심사는 한국 문화체험이라는 공통점을 지니고 있지만 상세한 주요 내용에 차이가 있는 특성을 확인할 수 있었다.

1. 서론

인바운드 관광이란 해외에서 국내로 들어오는 관광객들에 의해 이루어지는 관광 활동을 뜻한다.[1] 코로나 19 영향이 회복됨에 따라 관광수입이 점차 증가하고 있지만, 2023년 상반기 한국의 관광수지는 465,000만 달러 적자로 5년 만에 최대를 기록했다.[2] 이는 국민 해외여행객이 지출한 금액에 비해 인바운드 관광객의 지출이 훨씬 적다는 의미이다. 그렇기에 인바운드 관광객의 특성을 정확하게 파악하여 각 유형을 타겟으로 한 적절한 관광 정책을 실시한다면 관광수지 흑자 전환에 도움을 줄 것으로 보인다.

본 연구에서는 관광수지 회복을 위해 k-modes 군집 분석을 활용하여 인바운드 관광객의 유형 분석을 진행한다. 엘보우 기법을 통해 적절한 군집 개수를 3으로 설정하였으며, 세 가지 유형의 관광객 그룹으로 분류하였다.

군집 분석의 결과 업무수행 유형, 문화체험 유형, 쇼핑 유형으로 관광객 그룹을 나눌 수 있었다. 이를 토대로 진행한 추가적인 뉴스 자료 분석에서 통해 각 유형은 한국 방문 동기 및 관심사 측면에서 차이가 나타남을 알 수 있었다.

2. K-Modes 군집 분석

2.1 데이터 설명

외국인 관광객을 군집화하기 위해 MDIS의 외래관광객조사 2021년 데이터를 활용하였다. 이 데이터는 문화체육관광부에서 작성, 한국을 방문한 후 출국하는 외래관광객을 대상으로 한국 여행실태, 한국 내 소비실태 및 한국 여행 평가를 조사한 데이터이다.

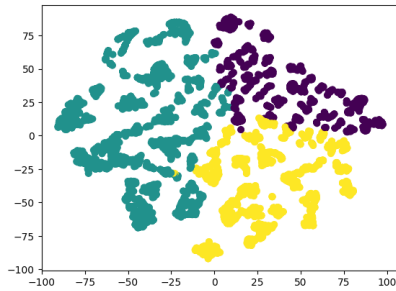
2.2 K-Modes

K-Modes는 범주형 데이터를 군집화하기 위해 개발된 비지도 학습 알고리즘으로 객체 간의 속성빈도 값의 비교를 통하여 군집 간의 거리를 정하는 방법이다. [3] K-Means와 유사하나 연속형 변수 대신 범주형 변수, 평균 대신 최빈값(mode)를 사용하는 것이 특징이다. 본 연구에서는 엘보우 기법을 사용하여 최적의 군집 개수를 3으로 설정하였다.

3. 분석 결과

3.1 군집 분석 결과

외래관광객 조사를 토대로 군집 분석을 진행하여 인바운드 관광객을 총 3개의 군집으로 나누었다.



(그림 1) 군집화 결과 t-SNE 차원축소 시각화 그래프

업무수행 유형은 남성, 서구 문화권(미국, 중국, 독일, 일본, 영국 등)이 많이 나타났다. 대부분이 방문목적은 사업 & 전문 활동이고 참여활동은 업무수행과 쇼핑으로 방문의도와 실제 활동이 일치함과 동시에 쇼핑 활동도 활발하다는 것을 알 수 있었다. 또한 세 유형 중 소비금액도 가장 크게 나타났다.

문화체험 유형은 다양한 지역을 방문하고 다양한 활동에서 비슷한 참여율을 보였다. 방문 지역은 주요 광역시이고 참여 활동은 자연경관, 전통문화체험, 유흥&오락, 뷰티&의료관광, 스포츠&레포츠 관람 및 참가 등에 참여하며 다른 유형에 비해 많은 활동을 하였다. 주요 방문 시기는 6 월, 7 월, 12 월로 학생들의 방학 기간에 해당하는 시기에 가장 높았으며 서구권 국가(미국, 프랑스 등)의 비율이 높았다. 또한 가장 높은 만족도를 보인 유형이다.

쇼핑 유형은 대다수가 서울 지역만 방문하고 주요 참여활동은 식도락, 쇼핑, 연수&교육&연구, 뷰티&의료로 분석됐다. 또한 중국&일본 국적의 10 대에서 30 대 사이 여성 관광객 비중이 높은 것이 특징으로 나타났다.

3.2 뉴스 단어 빈도수 분석 결과

유형별로 더 섬세한 유형 분류를 위해 뉴스 크롤링을 활용하여 단어 빈도수 분석을 진행하였다.



(그림 2) 문화체험 유형의 뉴스 단어 분석 결과 시각화

업무수행 유형은 지출 금액이 제일 높게 나타난 것이 특이하여 '외국인 관광객 출장'의 키워드로 분석을 진행하였다. 페스티벌, 내한공연, 콘서트, 한류, 공연의 단어들을 통해 사업 방문 외에 여가시간에 공연 관람, 페스티벌 참가를 하는 것으로 분석됐다.

문화체험 유형은 다양한 지역을 방문하는 것이 특이하여 인천, 부산, 대구, 광주, 문화, 외국 관광을 키워드로 분석을 진행하였다. 방문 동기는 찜질방, 한옥 호텔 등 한국에서만 체험 가능한 것에 관심을 가지

고 방문하는 것으로 분석됐다. 체험, 축제, 카지노, 콘서트, 단어를 통해 직접적인 체험활동에 관심이 있고, 찜질방, 전통, 편의점, 한옥호텔 단어들을 통해 한국에서 해당 서비스를 체험하는 것이 목적으로 보인다.

쇼핑 유형은 중국, 일본의 비율이 크게 나와 해당 국가, 내한 관광객을 키워드로 분석을 진행하였다. 방문 목적으로 친지&친구가 높게 나왔으나 일본 관광객에 대한 단어로는 공연, 내한, 고척돔이 주로 나오는 것을 통해 한국 방문 동기가 한류 공연 관람에 대한 기대가 있음을 발견하였다. 중국 관광객에 대한 단어는 체험, K가 주로 나와서 일본 관광객과 다르게 한국 관광 동기가 K(한국문화)를 직접 체험하는 것에 수요가 있음을 발견하였다. 따라서 쇼핑유형은 주로 한류 관련 소비를 하고 있음을 알 수 있었다.

4. 결론 및 활용 방안

본 연구에서는 한국데이터랩의 외래관광객조사를 활용하여 K-Modes 를 사용하여 외래 관광객의 유형을 분석하였다. 또한 유형별 특이점을 뉴스 크롤링을 통해 단어 빈도수 분석을 진행해 자세한 방문 목적을 알아보았다. 분석 결과 업무수행, 문화체험, 쇼핑 유형으로 나뉘었고, 이들의 방문 동기는 각각 업무수행 및 페스티벌 참여, 한국 고유문화 체험, 한류 공연 & K-문화 소비로 나타나는 것을 확인하였다. 공통적으로 한국 문화 체험에 관심을 가지고 있지만 세부적으로 유형별 수요 내용이 다르다는 것에 중점을 두어야 더 정확한 유형별 이해가 이뤄질 것이다. 현재 분석은 유형별 소비 이력, 정확한 방문 목적 동기를 찾아내는 것에 한계가 있어 향후 연구에서는 유형별 SNS 분석, 소비 내역 분석을 추가할 예정이다. 우리나라는 서울에 관광이 집중되어 있어 서울을 방문하고 나면 재방문의 매력도가 떨어진다. 따라서 본 연구 결과를 활용하여 유형에 따른 맞춤형 여행 상품 구성에 활용한다면 관광 만족도와 재방문 비율을 높일 수 있을 것으로 기대된다.

참고문헌

- [1] 한누리. "인바운드 관광 활성화를 위한 방한 외국인 관광지 선택요인 IPA 분석." 국내석사학위논문 중앙대학교 예술대학원, 2024. 서울, 17p
- [2] 김도훈. 방한 태국 관광객의 관광 지출 특성 및 결정 요인. 질서경제저널, 26(4), 29-47, 2023
- [3] 조진혁, 고송, 김대원. 클러스터 앙상블을 활용한 K-modes 알고리즘의 성능 향상 기법. 한국지능시스템학회 학술발표 논문집, Vol.19 No.1 145 - 148p, 2009.04
- [4] 이윤수, 테이퍼악떠라, 이종혁, 길준민 "키워드 군집화를 이용한 연구 논문 분류에 관한 연구" 정보처리학회 논문지 7.12 pp.477-484 (2018)