

의미 연결망 분석과 MCLP를 통한 천안시 전통시장 활성화 방안 제안

CUAI 7기 DA 3팀

김민하(응용통계학과), 김예은(경영학과), 박아현(응용통계학과), 이해원(수학과)

[요약]

천안시 전통시장 활성화를 위해 상황식 접근 방식을 활용했다. 크롤링 데이터를 통해 워드클라우드와 LDA 기법으로 문제를 재정의하고, 의미 연결망 분석을 통해 카페 입점을 제안했다. DBSCAN과 MCLP 모델을 활용하여 주차장 입지를 선정하였고, 이를 통해 청년층과 중장년층을 위한 전통시장 활성화 방안을 제시했다.

주차에 대한 부정적인 의견이 많다고 결론지을 수 있었다.



[그림 1] 구글 맵 리뷰 워드 클라우드

1. 서론

데이터 분석으로 우리 사회가 가진 문제를 실질적으로 해결하는 것을 큰 목표로 하였다. 이에 따라 '2024 천안시 데이터분석 아이디어 경진대회'에 참여하여 지역사회에 분석 결과를 효과적으로 제안할 예정이다.

2022년 기준 천안시 전통시장 및 상점가 보고서에 따르면, 전통시장에서 40세 이상 고객이 평균적으로 85%를 차지하고 있다. 이는 천안의 전통시장이 관광지로서의 역할을 하지 못하고 있음을 시사한다. 특히, 20대 방문 비율이 저조한 상황에서 전통시장의 활성화 방안을 모색하고자 한다.

이러한 문제를 해결하기 위해, '상황식 접근 방식(Bottom Up Approach)'을 활용하여 문제를 재정의하고 효과적인 해결방안을 탐색한다. 크롤링 데이터로 워드클라우드와 LDA 기법을 적용해 문제를 재정의하였으며, 이후 의미 연결망 분석과 DBSCAN 알고리즘, MCLP 모델을 활용하여 더욱 심도 있는 분석을 진행했다. 이러한 방법들을 통해 전통시장을 활성화하는 방안을 구체적으로 제시할 것이며, 지역 경제에 기여할 다양한 방안을 제시할 것이다.

2. 본론

1) 1차 데이터 수집 및 문제 정의

천안시 전통시장의 보완점을 정의하기 위해 5개의 시장(중앙시장, 역전시장, 성정시장, 성환이화시장, 병천시장)의 2022~2024년 리뷰를 5개의 사이트(구글 맵, 카카오 맵, 네이버 블로그, 유튜브, 구글 뉴스)에서 크롤링했다. konlpy의 okt 라이브러리를 이용한 어간 추출, 불용어 제거 등의 텍스트 전처리를 마친 후, 명사와 형용사만 추출하여 분석을 진행했다.

워드 클라우드와 LDA를 통해 크롤링 데이터에서 문제를 정의하였다. 워드 클라우드는 단어의 등장 빈도를 글자 크기로 나타낸 이미지이며, 특히 구글 맵 데이터로 워드 클라우드를 생성했을 때 '가격'과 '주차' 키워드의 언급 빈도가 높은 것을 확인할 수 있었다. 이에 '주차' 키워드가 언급된 리뷰만 분석해보았고,

LDA는 문서의 집합에서 핵심 토픽을 찾는 확률 모델이다. 각 토픽 단어들의 일관성을 나타내는 Coherence score를 이용하여 데이터프레임별 최적의 토픽 개수를 도출했고, 이 토픽 개수로 LDA 토픽 모델링 및 시각화를 진행했다. 천안시 5개 시장 리뷰와 MZ세대에게 인기 있는 시장 리뷰로 각각 LDA를 진행한 결과, MZ세대에게 인기 있는 전통시장의 특징으로 언급된 토픽(팝업스토어+로컬+지역/축제+관광+행사+골목+거리)이 천안시장에서는 언급되지 않는 것을 확인할 수 있었다.

2) 타겟층 구분

앞서 정의한 2가지 문제를 해결하기 위해 전통시장의 타겟층을 청년층과 중장년층 둘로 구분하였다.

청년층은 주로 교통수단을 이용해 천안에 방문하는 관광객들로 정의했다. 의미 연결망 분석을 통해 MZ세대에게 인기 있는 전통시장의 특징 중 천안시장에 없는 키워드를 파악하고, 천안시장 공실에 해당 키워드를 활용한 공간 신설을 제안하고자 한다.

중장년층은 주로 승용차를 이용해 천안시장에 방문하는 천안시민 및 충청도민으로 정의했다. 시장 주변 주차장 및 불법주정차 위치 시각화로 주차장이 부족한 시장을 파악한 후 DBSCAN과 MCLP를 통해 신설 주차장의 입지를 제안하고자 한다.

규모가 작은 5일장 형태인 병천시장은 관광객 대상으로 적합하지 않고, 주차공간을 신설할 필요성이 없다고 판단하여 앞으로의 분석에서 제외하기로 결정했다. 또한, 천안시 2024 주요 업무 계획에 '성환이화시장 주차장 조성'이 포함되어 있어 중장년층을 위한 분석에서는 해당 시장을 제외하였다.

3) 청년층을 위한 시장 공실에 카페 입점 제안 가. 의미 연결망 분석

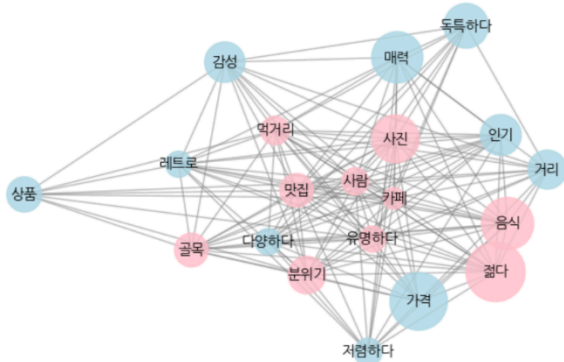
'MZ 전통시장', '힙한 시장', '시장 핫플' 검색어로 네이버 블로그를 추가 크롤링한 후 의미 연결망 분석을 진행했다. 분석은 크게 2가지로 구분된다. (1) MZ세대에게 인기 있는 전통시장의 특징과 (2) 많이 언급된 카페의 특징이다. 후자로는 스타벅스 경동1960점, 아베베베이커리, 어니언, 일호상회, 업스탠딩을 선정했다.

의미 연결망 분석(Semantic Network Analysis)은

문서 내 단어들 사이의 관계를 네트워크로 시각화하는 분석 기법이다. 문서에서 빈도수 상위 60개 단어 중 아이겐벡터 중심성(Eigenvector Centrality) 상위 20개의 단어로 그래프를 생성 및 시각화했다. 특정 노드(단어)와 연결된 노드의 중심성이 높을수록 아이겐벡터 중심성이 높아지는 특징이 있다.

MZ세대에게 인기 있는 전통시장의 특징으로는 카페, 먹거리, 사진, 분위기, 감성, 경험, 새롭다, 지역, 문화 등의 단어들이 추출되었다. 이 단어들을 5개의 카페에 적용한 결과 공통적으로 추출되는 단어를 바탕으로 천안시장에 필요한 특징을 정리해보았다.

〈MZ 키워드를 어니언에 적용 - TOP 20〉



[그림 2] 의미 연결망 시각화 예시(어니언)

- 감성, 분위기, 매력, 사진, 독특하다, 새롭다
→ 고유한 감성을 가진 카페에 방문하여 사진 촬영
- 로컬, 지역 → 지역 특산물을 이용한 메뉴
- 골목, 거리, 지역 → 골목/거리/지역 특색 반영
- 팝업스토어, 체험, 경험 → 음식 외의 즐길 거리

나. 카페 입점 제한

천안시가 중시하는 역사(유관순)와 빵(천안빵빵데이), 그리고 천안시의 특산물(흥타령쌀, 신고배, 거봉포도, 수신크로, 광덕호두 등)을 활용한 카페 입점을 제안하고자 한다.

첫째로, 인근에 카페가 없는 성환이화시장과 성정시장 공실에 쌀 베이커리 카페 입점을 제안한다. 흥타령쌀을 이용해 쌀호두과자, 쌀케이크, 배빵, 포도빵 등의 쌀 디저트를 판매하고, 천안빵빵데이의 주요 명소로 자리잡는 것을 목표로 한다. 과일 특산물을 이용한 착즙주스나 흥타령쌀 식혜 또한 판매가 가능할 것이다.

둘째로, 역전시장과 중앙시장 공실에 역사 테마 카페 입점을 제안한다. 역전시장은 6.25 전쟁 이후 피난민들이 모여 형성된 시장이며, 중앙시장은 100년 이상의 역사를 자랑하는 시장이다. 천안시의 역사 관련 관광 명소(독립기념관, 유관순열사기념관 등)를 소개하고, 광복 및 6.25 전쟁 관련 체험 거리를 제공하는 것을 목표로 한다.

4) 중장년층을 위한 신설 주차장 입지 제안

가. 분석 방법론 및 분석 흐름

천안시 전통시장의 신설 주차장 입지 선정에 위하여 천안시 불법주정차 단속현황 데이터를 대상으로 DBSCAN 알고리즘과 MCLP(Maximal Covering Location Problem) 모델을 적용하였다. 이때, 도보 10분 거리에 위치하고 있는 중앙시장과 역전시장은 하나의 구역으로 보고 분석을 진행하였다.

DBSCAN(Density-Based Spatial Clustering of

Applications with Noise) 알고리즘은 밀도 기반으로 서로 가까운 데이터 포인트를 같은 클러스터에 할당하는 군집분석 방법이다. 불법주정차 적발 위치 데이터에 대하여 DBSCAN을 수행하고 각 클러스터에 속한 데이터 포인트의 평균을 주차장 입지 후보지로 선정하였다.

MCLP 모델은 시설물의 개수가 제한되어 있을 때, 주어진 제한 조건 하에서 지역 내 수요를 최대한 많이 충족할 수 있는 입지를 선정하는 기법이다. MCLP 모델은 다음과 같다.

$$\text{Subject to Max } \sum_{i \in I} w_i y_i \quad (1)$$

$$\text{such that} \\ \sum_{i \in N_i} y_i \geq x_i \quad (2)$$

$$\sum_{i \in I} y_i = P \quad (3)$$

$$x_i \in \{0, 1\} \quad (4)$$

$$y_i \in \{0, 1\} \quad (5)$$

Inputs:

I	수요지점 세트
J	주차장 입지 후보지 세트
S	최대 커버리지 범위
P	설치할 주차장의 수
N_i	수요지점 i 로부터 커버리지 반경 내에 있는 후보지 집합
w_i	수요지점 i 에서의 주차장 수요 (불법주정차 적발 건수)

Decision variable:

$$x_i = \begin{cases} 1 & \text{수요지점 } i \text{가 적어도 하나의 주차장에 의해 커버되는 경우} \\ 0 & \text{그렇지 않은 경우} \end{cases}$$

$$y_j = \begin{cases} 1 & \text{주차장이 후보지 } j \text{에 입지하는 경우} \\ 0 & \text{그렇지 않은 경우} \end{cases}$$

목적함수 (1)은 신설되는 주차장에 의해 커버되는 주차장 수요의 합을 최대화한다. 이때, 주차장 수요 w_i 는 고유한 불법주정차 적발 위치에서의 총 적발 건수로 가정하였다. 제약조건 (2)는 N_i 에 속한 후보지 중 적어도 한 곳에 주차시설이 입지하면 수요지점 i 는 커버될 수 있다. 제약조건 (3)은 최종 설치할 주차장의 수에 대한 것으로, 여기서는 각 전통시장 당 2곳의 주차 시설을 입지한다고 설정하였다. 최대 커버리지 범위는 선행연구에 근거하여 중앙, 역전시장의 경우는 500m, 성정시장의 경우는 그 규모를 고려하여 입지 후보지와 수요지점 간 거리 평균인 200m로 설정하였다.

나. 신설 주차장의 최적 입지 선정 결과

1. 중앙시장과 역전시장

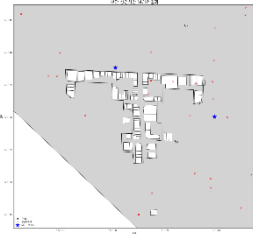
DBSCAN 수행 결과, 총 93곳의 수요지점 중 31곳의 입지 후보지가 선정되었다. MCLP 모델을 적용한 결과는 [그림 3]과 같다.

2. 성정시장

DBSCAN 수행 결과, 총 26곳의 수요지점 중 16곳의 입지 후보지를 선정하였다. MCLP 모델을 적용한 결과는 [그림 4]와 같다.



[그림 3]
중앙시장, 역전시장



[그림 4]
성정시장

3. 결 론

본 제안에서는 천안시 전통시장의 활성화 방안을 제시하고자 하였다. 크롤링을 통해 수집한 데이터를 활용하여 워드클라우드와 LDA 분석을 진행한 결과, 다음 두 가지 문제점을 발견하였다.

1. 청년층이 최근 전통시장에서 선호하는 요소가 천안시의 전통시장에는 부족하다.
2. 시장 주변에 중장년층을 위한 주차 공간이 부족하다.

청년층을 위해 ‘의미 연결망 분석’을 사용하여 MZ세대가 선호하는 전통시장의 특징과, 많이 언급된 카페에 대한 특징을 도출하였다. 이를 바탕으로, 천안시가 중시하는 명물과 역사를 조합하여 시장의 공실에 ‘쌀 베이커리 카페’와 ‘역사 테마 카페’ 등의 입점을 제안했다.

중장년층을 위한 신설 주차장 입지 선정에서는 DBSCAN 알고리즘과 MCLP 모델을 적용하여 신설 주차장 입지를 선정하였다. DBSCAN 분석 결과, 중앙시장 및 역전시장에서 총 31곳, 성정시장에서 16곳을 입지 후보지로 선정했다. 또한 해당 결과에서 최적의 입지를 선정하기 위해, 최대한 많은 수요를 커버할 수 있도록 하는 MCLP 모델을 적용했다. 그 결과, 각 시장의 주차장 수요를 최대한 충족하는 최적의 입지를 선정할 수 있었다.

이번 제안은 크롤링을 통해 사람들이 천안시 전통시장에 가진 인식을 파악할 수 있게 하였으며, 청년층의 전통시장 무관심 문제, 시장의 주차장 혼잡 문제를 해결하는 데 기여할 것으로 예상된다. 향후 이 제안의 분석에 필요한 천안시의 구체적인 데이터가 추가된다면 더욱 적합한 연구로 발전할 수 있으리라 기대한다.

참고 문헌

- 1) 정해운, 이성호, “소셜 빅데이터를 활용한 전통시장

활성화 방안 연구”, 유통연구, 2024.01.

2) 박준성, 이준, 주용진, “빅데이터와 공간분석을 통한 서울시 전통시장 활성화 방안”, 대한공간정보학회 학술대회, 2023.12.

3) [Python] BeautifulSoup을 이용한 네이버 블로그 크롤링 - 2. <https://url.kr/cbc6yh>

4) [12일차] ABC부트캠프(유튜브 댓글 크롤링... 네이버 블로그.

https://blog.naver.com/a1egre_/223506914386

5) KoNLPy 한국어 처리 패키지 - 데이터 사이언스 스쿨. <https://url.kr/h6ntio>

6) [NLP] LDA 토픽 모델링을 활용한 앱 리뷰 분석 프로젝트. <https://heytech.tistory.com/401>

7) Semantic Network Analysis | chaelist. <https://url.kr/u4wpnb>

8) 이해빈, 김다혜, 남귀연, 정유지, 한정규, 박영진, “공공데이터 기반 공영주차장 최적입지 선정을 위한 최대 커버지역 문제(MCLP) 해결 기법”, 멀티미디어학회논문지, 2023.02.

9) 최낙현, 김정화, “빅데이터 기반 공유형 마이크로 모빌리티의 주차시설 입지 최적화 연구”, 대한토목학회논문집(국문), 2023.04.

10) 오현석, 이주용, “MCLP를 활용한 수원시 급속 전기차충전소 입지 최적화”, 대한교통학회 학술대회지, 2023.03.

11)

COMPAS_SUWON_3rd/COMPAS_최종코드.ipynb at master wow-kim. <https://url.kr/y7z3y7>