실내 환경에서 사용자 이동 경로 보간 및 클러스터링 기반 동선 분석

이재웅, 김장겸*, 김태엽
세종대학교, *세종대학교, (주) 파파야

Jaewoong0310@gmail.com, *jk.kim@sejong.ac.kr, tyrus.kim@papaya.camp

Network-Based Path Interpolation and Clustering for User Trajectory Analysis in Indoor Environments

Lee Jae Woong, Kim Jang Kyum*, Taeyeab Kim
Sejong Univ., * Sejong Univ., Papaya Co., Ltd

요 약

본 논문은 스마트폰 위치 데이터를 활용한 실내 공간 분석을 위해 네트워크 기반 지리분석 방법론을 확장하는 새로운 프레임워크를 제안한다. 2024 코엑스 의료기기 전시회에서 수집된 사용자 위치 데이터를 통해 비동기적 위치를 보간하고 방문 패턴에 따른 클러스터링을 수행하였다. 분석 결과 전시회장 이용자는 집중관람형, 적극탐색형, 빠른 순회형의 3 가지의 그룹으로 클러스터링 되었으며, 이는 향후 다양한 실내공간의 방문객 분석에 적용 및 활용될 수 있다.

Ⅰ. 서 론

최근 스마트폰의 광범위한 보급과 무선통신기술의 발전으로 사용자 위치 데이터의 수집이 용이해지면서, 이를 기반으로 한 다양한 서비스가 상용화되고 있다 [1,2]. 특히, 실내 환경에서의 위치 기반 서비스는 쇼핑몰, 박람회, 공항 등에서 사용자 동선 분석을 통한 맞춤형 서비스 제공이 가능하다는 이점을 지닌다. 이러한 배경에서 사용자 위치 데이터의 효율적 수집과 분석은 다양한 산업에서 중요한 연구 분야로 자리잡고 있다.

하지만, 기존의 위치 데이터 수집 및 분석 방식은 주로 실외 공간을 대상으로 개발되어, 건물 실내에서는 적용에 한계가 존재한다. 또한, 사용자별로 수집된 위치 데이터는 시간적 비동기성을 가지며, 데이터 누락이 빈번히 발생하는 문제가 있다. 이를 해결하기 위해 본 연구에서는 비동기적 수집 시간에 따른 사용자 위치를 보간하는 방법과 방문 위치 기반의 실내 환경 특화 대표 동선 분석 프레임워크를 제안한다.

Ⅱ. 본론

A. 데이터 설명 및 전처리

본 연구에서는 코엑스에서 진행한 2024 Korea International Medical & hospital Equipment Show 의 이용객 중 행사장 안내 어플리케이션을 활용하여 이용객의 위치데이터를 수집한 데이터를 사용하였다.
어플리케이션에서 수집된 특성상 위치데이터는 사용자가 어플리케이션을 사용하거나 어플리케이션 내에서 부스 검색 기능을 활용하기 위해 접속한 시간 등 사용하는 시간에서만 데이터가 수집되어 시간에 따른 비동기성을 포함하고 있다. 2024 년 3 월 14 일부터 17 일까지 4 일간의 데이터 중 행사 운영시간인 10 시부터 18 시 중 최소한 두번 이상 위치 좌표가 수집되며 그 차이가 2 분 이상인 사용자의 데이터만 한정하여 사용했다. 각 일자별 수집된 사용자 위치 좌표 데이터와 사용자수는 Table.1 과 같다. 또한 코엑스 행사장 1 층의 A 홀과 B 홀의 홀 내 사거리 좌표 데이터를 활용하였다. 사거리 좌표 데이터의 A 홀 48 개, B 홀 40 개, A 홀과 B 홀 통로 4 개로 총 개수는 92 개이다.

날짜	위치데이터 개수	사용자수
2024-03-14	3942	238
2024-03-15	4548	266
2024-03-16	2470	164
2024-03-17	1546	96
Total	12506	764

Table. 1 일자별 데이터 수

B. 사용자 위치 데이터 보간

1 층 내 A, B 홀의 사거리 좌표 데이터를 통해 각 좌표를 노드로 사용하고 노드와 노드 사이 통로를 네트워크로 하여 1 층 내 이동 동선을 Fig.1 과 같이 양방향 네트워크화 하여 나타내었다 [3]. 각 노드를 활용하여 각 홀 안에 위치해있는 사용자 위치 좌표 데이터를 최근접거리 노드에 포함시켜 각 노드당 밀집인구를 계산 후 역수로 활용하여 양방향 네트워크 링크의 가중치로 사용하였다. 네트워크 내에서 경로 보간을 위해 사용자의 위치 데이터를 통해 얻은 출발 노드 경유지 노드, 도착 노드를 설정한 후 총 경로의 가중치의 합이 가장 작은 경로를 이용하는 다익스트라 알고리즘을 활용하여 사용자의 이동

경로를 보간하였다 [4].

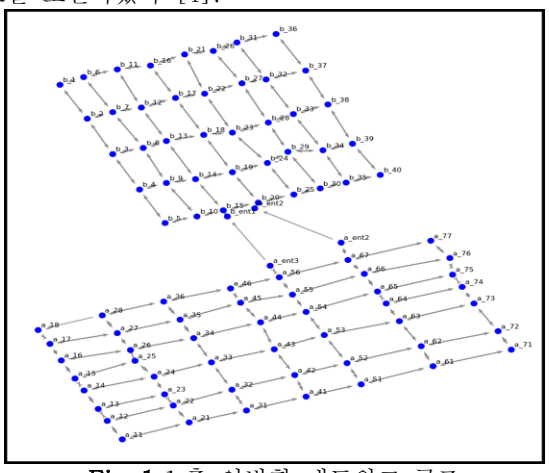

Fig. 1 1 층 양방향 네트워크 구조

C. 방문지 기반 클러스터링

노드의 방문 여부를 One-hot encoding 으로 변환하여 0,1 로 표기하였다. 각 사용자당 사용한 네트워크 길이, 각 노드당 방문여부 변수를 K-means 클러스터링을 위한 변수로 선정함. 해당 데이터를 바탕으로, 클러스터링을 진행하였으며, 이 때, 최적의 k 를 선정하기 위해 Fig.2 의 Elbow-method 와 Calinski-Harabasz Index 를 활용하였다.

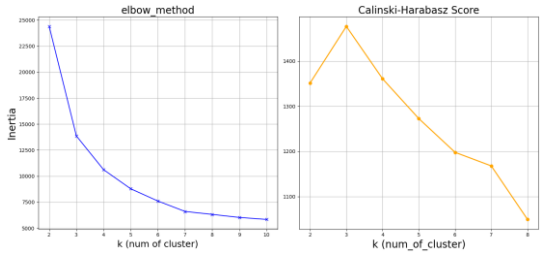

Fig. 2 Elbow method 와 Calinski-Harabasz Score

Fig. 2 에서와 같이, 2 가지 Index 를 통해 분석한 결과, 클러스터의 개수가 3 인 경우가 optimal 한 클러스터 수인 것으로 판단하였으며, 본 연구에서는 해당 결과를 바탕으로 고객들의 이동 동선을 분석하였다.

D. 각 클러스터별 분석

각 클러스터링 결과와 특징은 Table.2 와 같다.

클러스터	N	평균경로길이	평균 체류 시간 (분)
Group 1 (집중탐색)	470	3.74	51
Group 2 (적극탐색)	220	14.72	110
Group 3 (빠른순회)	74	30.05	82

Table. 2 각 클러스터별 개수 및 특징

각 클러스터링 그룹의 특징으로는 Group1 의 경우 이동 경로가 짧으며 하나의 부스에서 평균적으로 13 분이상 소요하였다. 즉 소수의 부스를 집중적으로 관찰 후 퇴장하는 집중탐색형이라 추정할 수 있다.
Group2 의 경우 체류시간이 가장 길지만 상대적으로 이동 경로는 짧아 평균 부스 시간이 8 분으로 나타났다. 이러한 점을 통해 적극탐색형 이용자라 추정할 수 있다.
Group3 는 가장 긴 평균 경로가 나타난 그룹으로 전반적인 관람을 위해 많은 이동이 나타났고 부스당 평균 소요시간이 3 분으로 많은 부스를 보기 위해 빠르게 이동하여 개별 부스에서의 관람 시간이 매우 짧았을 것이라 예상하여 빠른 순회형 이용자로 추정한다.
각 클러스터 별 가장 많은 방문이 일어난 상위 15 개의 노드는 Fig.3 의 그림과 같다.

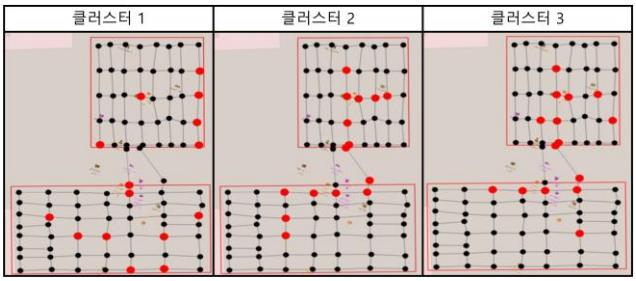

Fig. 3 각 클러스터별 주요 방문 노드

고객들이 가장 많은 방문이 발생한 상위 15 개의 노드를 Fig. 3 에 시각화하였다. 해당 결과를 통해 고객군별로 관심도가 높은 부스를 파악할 수 있으며, 각 클러스터의 주요 이동 경로와 방문 지점을 분석할 수 있다. 이러한 분석을 바탕으로, 고객이 자주 방문하는 위치에 관련 부스를 배치하여 방문자 경험을 개선하고, 추가적인 수익을 극대화하는 전략을 수립할 수 있을 것으로 기대된다.

Ⅲ. 결론

본 연구에서는 기존의 대규모 야외지역에서 시행되었던 네트워크 분석을 선물 실내에서 적용하기 위한 프레임워크를 제안한다. 이를 통해 코엑스 전시회장의 사용자 이동경로와 방문 패턴을 효과적으로 분석하였으며 그룹별 주요 방문 위치, 이동, 체류 특성을 확인하였다. 향후 연구에서는 이러한 프레임 워크를 통해 백화점, 복합문화공간, 먹자골목 등 실내 밀집도가 높은 곳에서 또한 확대 적용이 가능할 것이며 사용자 행동 및 위치 예측에 적용할 수 있을 것이라 기대한다.

ACKNOWLEDGMENT

이 논문은 2024 년 대한민국 교육부와 한국연구재단의 인문사회분야 신진연구자지원사업의 지원을 받아 수행된 연구임(NRF-2024S1A5A802363012). 또한, 파파야 주식회사의 지원을 통해 연구가 수행됨.

참 고 문 헌

[1] 김영롱, 윤메솔, and 이혜령. 빅데이터를 이용한 코로나 시대 수도권유동인구 네트워크 분석. 경기연구원 기본연구, pages 1–73, 2021.

[2] Gagandeep Kaur and Ruchika Gupta. A study on location based service sand ttp based privacy preserving techniques. In 2021 InternationalConference on Advances in Computing and Communications (ICACC),pages 1–5, 2021.

[3] Peter Haggett and Richard J. Chorley. Network Analysis in Geography.Edward Arnold, London, 1969. Bibliography: p. 319–335.

[4] Edsger W Dijkstra. A note on two problems in connexion with graphs. Numerische mathematik, 1(1):269–271, 1959.