

# 이동 객체를 위한 시공간 궤적의 질의 처리

<sup>1\*</sup>오병우

## Query Processing of Spatio-temporal Trajectory for Moving Objects

<sup>1\*</sup>Byoungwoo Oh

### 요약

최근 전염성이 높은 COVID-19 바이러스의 영향으로 접촉자 추적에 사용할 수 있는 시공간 궤적에 대한 중요성이 커지고 있다. 시공간 궤적은 객체가 이동한 시간 및 공간 데이터를 저장한다. 본 논문에서는 이동 객체의 시공간 궤적에 대한 질의 처리를 제안한다. 이동 객체의 시공간 궤적 모델은 위치를 저장하기 위하여 포인트 형태의 공간 데이터를 가지고 타임스탬프 형태의 시간 데이터를 가진다. 궤적 질의는 같은 버스에 탑승하여 서로 밀접하게 접촉한 사용자들의 쌍을 검색하는 질의이다. 궤적 질의를 처리하기 위하여 데이터셋으로는 마이크로소프트의 Geolife 데이터를 사용한다. 제안된 궤적 질의 처리는 궤적 데이터를 날짜별로 나누고 각 날짜에 대해 사용자의 궤적들이 근처에 있었는지 검사하고 접촉자들에 대한 정보를 결과로 생성한다.

### Abstract

*The importance of spatio-temporal trajectories for contact tracing has increased due to the recent COVID-19 pandemic. Spatio-temporal trajectories store time and spatial data of moving objects. In this paper, I propose query processing for spatio-temporal trajectories of moving objects. The spatio-temporal trajectory model of moving objects has point type spatial data for storing locations and timestamp type temporal data for time. A trajectory query is a query to search for pairs of users who have been in close contact by boarding the same bus. To process the trajectory query, I use the Geolife dataset provided by Microsoft. The proposed trajectory query processing method divides trajectory data by date and checks whether users' trajectories were nearby for each date to generate information about contacts as the result.*

**Keywords:** Trajectory, Moving objects, Spatio-temporal data, Query processing, Contact tracing

---

<sup>1\*</sup> 국립금오공과대학교 컴퓨터공학과 교수 (bwoh@kumoh.ac.kr)

## I. 서론

최근 COVID-19 바이러스 사태로 전세계가 어려움을 겪고 있고, 정부 부처는 감염자 선별 및 관리에 집중하고 있다. 바이러스의 확산 경로, 접촉자 추적 등을 위해서는 이동 객체의 시공간 궤적에 대한 기술에 대한 중요성이 증대되고 있다[1][2][3]. 이동 객체가 시간과 공간상에서 움직인 객체의 변화를 저장하는 시공간 궤적에 대한 처리는 바이러스 접촉자 추적에 사용될 뿐만 아니라, 치안, 안전, 재난 관리, 물류, 생태, 환경, 도시 계획, 디지털 트윈 등 광범위한 분야에서 활용할 수 있는 중요한 연구 분야이다.

이동 객체는 시간과 공간상에서 움직이므로 시간과 위치 데이터로 구성된다. 시간 데이터는 일반적으로 수집하기 쉽고, 위치 데이터는 주로 GPS 를 통해 수집하는 경우가 많다. GPS 이외에도 접속된 WiFi 나 기지국의 위치를 사용하는 경우도 있다. 이동 객체는 사람인 경우가 많지만 한랭전선, 철새 떼, 도시 경계 등과 같이 다양한 경우도 존재하므로 활용 목적에 따라서 적절한 모델을 사용한다. 공통적으로 시간 데이터, 공간 데이터, 나머지 속성 데이터로 구성되는 경우가 많다. 시간 데이터로는 시작 시간과 종료 시간으로 구성되는 기간을 사용할 수도 있지만 일반적으로 타임스탬프를 사용하는 경우가 많다[4][5].

궤적 질의는 이동 객체에 대한 시공간 질의로서 이동 객체의 위치 및 시간을 검색하는 질의, 공간 및 시간 축에 대한 중첩, 교차, 포함 등의 조인 질의, 가장 가까운 k 개의 이동 객체를 검색하는 kNN 질의 등이 있다. 이동 객체의 방향이나 속도 질의와 예측 질의도 사용된다.

이동 객체는 시공간상에서 움직이면서 다른 이동 객체와 가까이 접근하기도 하고 멀어지기도 한다. 심각하게 전염성이 높은 바이러스가 유행하는 시기에는 접촉자를 추적하는 것이 중요하다. 특히, 많은 사람들이 좁은 공간에 모일 수밖에 없는 대중교통, 음식점 등이 바이러스 감염의 취약지대가 될 수 있다. 본 논문에서는 서로 가까이 근접한 이동 객체들을 검색하는 시공간 궤적의 질의 처리를 제안한다.

## II. 시공간상에서 객체의 이동

### 2.1 이동 객체

이동하는 객체는 공간상에서 움직이면서 위치가 변경되는데 객체가 위치했던 시간을 포함하여 시공간 상에서 움직인다고 간주한다. 이동 객체의 좌표를 수집하기 위해서는 GPS 수신 데이터, WiFi 접속 기록, 기지국 접속 기록 등이 사용된다. 교통카드 사용 내역, 항공권 결제, 택시 호출 앱, 쇼핑몰 결제, 음식점 결제 등의 정보도 기록 시간과 공간상의 좌표를 특정할 수 있으므로 넓은 의미의 시공간 궤적 데이터로 볼 수 있다. 시공간 궤적 데이터는 개인의 좌표로서 매우 민감한 개인 정보이므로 보호가 필요하다. 이동 객체는 활용 목적에 따라 다양하게 모델링할 수 있다.

이동 객체는 기하학적인 모양의 공간 데이터와 타임스탬프를 사용하는 시간 데이터를 결합하여 시공간 데이터로 표현할 수 있다. 공간 데이터를 처리할 때 기하학적인 모양은 주로 포인트, 라인, 폴리곤의 형태로 표현한다[6]. 그림 1 은 시간의 변화( $t_0$  및  $t_1$ )에 따른 기하학적인 모양의 예시를 보여준다.

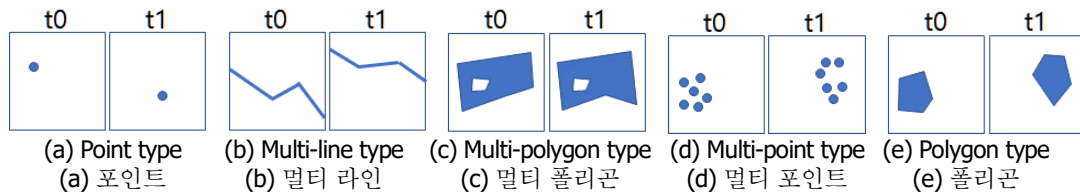


Figure 1. Geometry types of spatio-temporal data

그림 1. 시공간 데이터의 기하학적인 유형

사람, 차량, 선박 등과 같이 단일 위치 좌표에 집중하는 포인트 형태인 경우에는 그림 1(a)와 같이 2 차원 평면상에서의 포인트 기하 좌표와 타임스탬프로 이동 객체를 표현한다. 시간이 지나면서 모양이나 길이가 변경되는 해안선, 한랭전선 등과 같이 라인 형태인 경우에는 그림 1(b)와 같이 멀티 라인 기하 정보와 타임스탬프로 표현할 수 있다. 시간이 지남에 따라 모양과 면적이 변경되는 도시 경계, 재난 지역 등과 같이 다각형 형태인 경우에는 그림 1(c)와 같이 멀티 폴리곤 기하 정보와 타임스탬프로 표현할 수 있다.

여러 개체들이 집단을 구성하는 철새 떼, 보호 동물의 군락 등은 그림 1(d)와 같이 멀티 포인트 형태로 기하 정보를 표현할 수도 있지만, 각 개체의 위치보다 집단의 경계가 중요할 경우에는 그림 1(e)와 같이 폴리곤 형태로 기하 정보를 표현하고 타임스탬프로 시간 정보를 표현할 수도 있다. 활용에 따라서는 도시의 인구 변화와 같이 공간 데이터는 변화하지 않지만 기하 데이터가 아닌 속성 데이터가 시간에 따라 변화할 경우에도 넓은 의미의 이동 객체로 간주할 수도 있다. 도시의 경계(기하학적인 멀티 폴리곤 형태)가 시간에 따라 변경될 가능성이 존재하므로 실제로 변경되지는 않았더라도 이동 객체로 간주하면 더욱 유연한 모델링이 될 수 있다.

시간 데이터는 타임스탬프처럼 각각의 위치에 대한 한 개의 시각으로만 구성할 수도 있지만, 활용 목적에 따라 이동 시작 시각으로부터 이동 종료 시각까지의 시간 구간을 사용할 수도 있다. 이 경우에 포인트 형태의 이동 객체는 시간 구간에서 이동한 경로에 따라 라인 형태의 기하 데이터로 표현할 수도 있다. 시간 구간을 작은 단위로 나눌 수 있다면 작은 단위의 시간 구간에 대해 공간 데이터는 라인 세그먼트 형태로 표현할 수 있고 전체 구간에 대해서는 멀티 라인 형태로 표현할 수도 있다.

일반적으로 포인트 형태의 이동 객체에 대한 활용이 많다[1][3][4][7]. 본 논문에서도 이동 객체는 공간 데이터를 위해서 기하학적으로 포인트 형태를 갖고 시간 데이터를 위해서는 타임스탬프를 갖는 모델을 사용한다.

## 2.2 이동의 종류

이동 객체는 공간 및 속성 데이터가 시간에 따라 변하는 광의로 해석될 수도 있지만 본 논문에서는 사람을 이동 객체로 한정한다. 사람이 시공간 상에서 이동하는 궤적을 저장하기 위하여 포인트 타입의 기하 데이터와 타임스탬프 타입의 시간 데이터를 저장한다.

사람이 시공간 상에서 이동하는 경우는 크게 정기적으로 반복되는 이동과 비정기적인 이동으로 나눌 수 있다. 정기적으로 반복되는 이동은 반복 주기에 따라 매일, 매주, 매달, 매년 등으로 나눌 수 있다. 매일 반복되는 이동은 출근이나 등교가 있다. 매주 종교 모임에 참석하거나 주말 부부인 경우 금요일 퇴근 후에 집에 가기 위한 정기적인 이동이 있다. 매달 동호회 모임에 참석하거나, 매년 명절이나 부모님 생신에 참석하기 위한 이동도 정기적인 이동이다. 비정기적인 이동은 여행, 콘서트 참가, 쇼핑 등과 같이 정해진 반복 주기가 없는 이동이다.

사람이 이동하다 보면 자연스럽게 다른 사람과 이동 경로가 겹칠 수 있다. COVID-19, 메르스 등의 전염병이 확산되는 사태를 겪으면서 밀접 접촉자 파악 및 관리에 대한 중요성이 증대되었다. 접촉의 관점에서 이동은 사람의 인원과 위치 이동 거리에 따라 분류할 수 있다. 적은 인원이 모이고 가까운 위치로 같이 이동한 뒤에 흩어지는 이동은 대표적으로 마을 버스를 타는 것을 예로 들 수 있다. 더 많은 인원이 모여서 도시 내에서 이동하는 시내버스나 지하철, 국가 내에서 이동하는 고속버스나 기차를 이용하는 예제도 있다. 매우 많은 인원이 모여서 국제선 항공기를 타고 국가 밖으로 이동하는 경우도 있다. 택시 기사가 지속적으로 운전하는 택시의 경우는 한 명에서 세 명의 매우 적은 인원이 비교적 짧은 거리를 이동하지만 다른 승객들이 공통적으로 이용한다. 반면에 사람들이 위치가 이동하지 않는 장소에 모였다가 흩어지는 이동 패턴을 보이는 음식점, 카페, 찜질방, 직장, 학교 등도 존재한다. 이동 객체를 활용하려는 목적에 따라 다양한 이동이 추가될 것이다.

본 논문에서는 접촉의 관점에서 분류한 이동 중에서 많은 인원이 모여서 같이 이동한 뒤에 흩어져서 밀접 접촉자가 발생하기 쉬운 이동의 종류에 대해서 시공간 궤적의 질의 처리를 제안한다.

### III. 시공간 궤적의 질의 처리

객체가 시간과 공간 상에서 이동한 경로인 시공간 궤적은 해당 객체가 언제 어디를 이동하였는지를 표현한다. 시간은 연속적으로 흐르지만 위치는 이산적으로 측정하여 해당 객체의 궤적에 추가한다. 이렇게 누적된 이동 객체의 궤적은 다양한 용도로 활용할 수 있다. 궤적을 활용하기 위해서는 궤적 질의가 필요하다. 기존 시공간 궤적을 분석하기 위한 시스템은 크게 공간 데이터베이스, 시공간 빅 데이터 인프라스트럭처, 개발 도구로 구분할 수 있다[5]. Oracle Spatial, PostGIS, Spatial NoSQL (Neo4j Spatial) 등과 같은 공간 데이터베이스는 공간 데이터에 시간 데이터를 추가하여 같이 질의할 수 있도록 지원한다. 시공간 빅 데이터 인프라스트럭처에는 Hadoop 에 기반한 ST-Hadoop, Spark 에 기반한 TrajSpark, NoSQL 에 기반한 TrajMesa 등이 있고, Geopandas 파이썬 라이브러리, 별도의 시스템인 AsterixDB 등도 있다. 개발 도구에는 GDAL, GeoSpark, RPyGeo 등이 있다. 본 논문에서는 Pandas 및 Geopandas 파이썬 라이브러리를 사용하여 제안하는 시공간 궤적의 질의 처리를 구현한다.

#### 3.1 시공간 궤적 질의

시공간 궤적 질의는 이동 객체에 대해 사용자가 원하는 시간 및 공간 정보를 검색하는데 사용된다. 기존의 시스템에서 일반적으로 사용하는 시공간 궤적 질의는 다음과 같다[7]. 특정 이동 객체가 주어진 시간(정확한 시각 또는 시간 범위)에 어떤 공간상의 위치에 있는지 검색하는 질의, 또는 주어진 위치(정확한 공간상의 좌표 또는 좌표 범위)에 시간상 언제 위치했는지 검색하는 질의가 가능하다. 공간 및 시간 축을 기준으로 이동 객체와 공간 객체 사이에서는 중첩, 교차, 통과, 포함함, 포함됨 등의 질의를 수행할 수 있다. 예를 들어, 궤적, 도로, 기간 데이터를 공간 조인하여 특정 도로를 지났던 객체들과 그 시간을 검색할 수 있다. 주어진 공간상의 좌표에 가장 가까운 k 개의 이동 객체를 검색하는 kNN (k Nearest Neighbor) 질의도 가능하다. 이동 객체는 시공간상에서 이동하므로 방향과 속도를 가지고 있어서, 주어진 시간의 방향이나 속도를 검색하거나 특정 속도 이상의 구간을 검색하는 등의 질의가 가능하다. 이동 객체의 방향과 속도를 활용하여 위치를 예측하는 질의도 가능하다.

기존의 전통적인 질의뿐만 아니라 시공간 궤적을 활용하는 목적에 따라서 다양한 질의가 추가될 수 있다. COVID-19 나 메르스와 같이 비말, 공기 전파, 접촉을 통해 전파되는 바이러스로 인해 사회적인 문제가 발생했다. 특히, COVID-19는 매우 전염성이 높아서 같은 시간에 같은 공간에서 감염자 주변에 있었던 사람은 감염 위험성이 있는 밀접 접촉자가 된다. 바이러스의 감염 경로를 파악하고 감염자와 접촉한 사람들을 선별하기 위해서 시공간 궤적을 활용할 수 있다. 2.2 절에서 제안한 접촉의 관점에서 분류한 인원과 이동 거리에 대한 분류는 접촉자 파악을 위해 유용하게 활용할 수 있다. 많은 사람들은 버스, 지하철, 비행기, 선박 등의 이동 수단을 통해 출발지에서 목적지까지 이동하면서 밀접 접촉할 수 있으므로, 본 논문에서는 이동의 종류 중에서 많은 사람이 같이 이동하여 경로가 겹치는 이동 유형에서 접촉자들을 검색하는 질의를 처리하는 방법을 제안한다.

#### 3.2 궤적 데이터

사람이 이동한 궤적 데이터는 개인 정보를 포함하고 있어서 프라이버시 보호를 염두에 두어야 한다. 궤적 데이터 축적이 어렵고 민감한 개인 정보이므로 공개된 궤적 데이터가 많지 않다. 연구를 위해서는 시뮬레이션을 통해 궤적 데이터를 생성하여 사용하기도 한다. 본 논문에서는 공개된 실제 궤적 데이터인 Geolife 데이터셋을 사용한다.

마이크로소프트가 제공하는 Geolife 데이터셋은 주로 중국 베이징시에 거주하는 182 명이 2007년부터 2012년까지 5년 동안 이동하면서 GPS 장치와 핸드폰을 통해 수집한 좌표를 개인별로 저장한 광범위한 대용량 궤적 데이터이다 [8][9][10]. 실제 궤적이므로 정기적으로 반복되는 출근 등의 일상과 비정기적인 쇼핑, 여행 등을 모두 포함한다. 궤적 데이터는 182명에 대해 '000'부터 '181'까지의 개인 폴더에 별도로 GPS 로그를 저장한다. 해당 폴더에는 파일명은 궤적의 시작 시간이고 확장자는 PLT를 사용하는 파일들이 저장되어 있다. PLT 확장자 파일은

메타데이터인 처음 6 줄과 그 뒤에 위도, 경도, 고도, 타임스탬프 등의 심표로 구분된 데이터로 구성된다. 메타데이터를 제외하고 CSV (Comma-Separated Values) 파일로 처리할 수 있다.

182 명의 기록 중에서 73 명에 대한 데이터는 이동에 사용한 방법도 저장하고 있다. 저장된 이동 방법으로는 걷기, 뛰기, 자전거, 오토바이, 승용차, 택시, 버스, 지하철, 기차, 항공기, 보트의 11 개가 있다. 이동에 사용한 방법에 대한 데이터는 개인 폴더에 “labels.txt” 파일에 저장한다. 텍스트 파일에는 시작 시간, 종료 시간, 이동 방법 데이터가 저장된다.

본 논문에서는 Geolife 데이터셋의 궤적 데이터 및 이동 방법 데이터를 모두 읽어서 중국 베이징시 인근 지역에서 이동 방법이 버스인 데이터만을 필터링하여 저장하고 사용한다. 본 논문에서 사용하는 필터링된 궤적 데이터는 1,060,747 개이다. 데이터 전처리 및 질의 처리를 위한 개발 언어로는 파이썬을 사용했다. 그림 2 는 Geopandas 라이브러리를 사용하여 논문에서 사용하는 궤적 데이터를 © OpenStreetMap (openstreetmap.org/copyright) 배경지도 위에 출력한 예제를 보여준다.

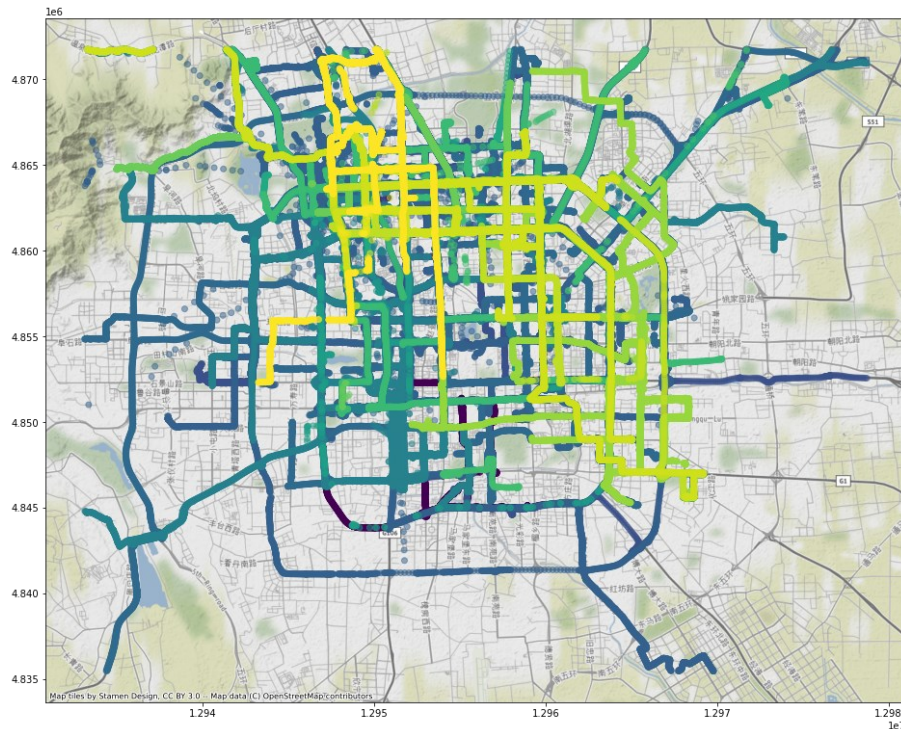


Figure 2. Trajectory data (background map: © OpenStreetMap)

그림 2. 궤적 데이터 (배경지도: © OpenStreetMap)

### 3.3 궤적 질의 처리

Geolife 데이터셋으로부터 중국 베이징시 인근 지역의 이동 방법이 버스인 궤적들을 필터링한 뒤에 Geopandas 의 데이터프레임을 생성한다. 그 뒤에 groupby() 함수를 사용하여 날짜별로 그룹을 생성하여 데이터를 준비한다. 준비된 데이터에 대해 궤적 질의를 처리한다. 많은 사람이 모여서 같이 이동한 뒤에 흩어지는 이동 유형 중에서 동일한 버스에 탑승한 접촉자를 추적하기 위한 궤적 질의 처리를 의사 코드로 표현하면 그림 3 과 같다. 코드에서 배열 result 는 검색된 접촉자 정보를 저장한다. 논문에서는 효율적으로 접촉자를 검색하기 위하여 날짜별로 나눠서 처리하는 방법을 사용한다. 코드에서 3 라인의 for 문은 궤적 데이터가 기록된 모든 날짜에 대해 각각의 날짜 그룹을 만들고 해당하는 날짜를 반복적으로 처리한다. 날짜 day 에 대한 날짜 그룹인 day\_group 은 해당하는 날짜(day)에 기록된 모든 궤적 데이터를 저장한다.

```

1 result = []
2
3 for day, day_group in 날짜별로 분류한 모든 그룹:
4     user_group = day_group에 대해 사용자별로 분류한 모든 그룹들
5     if user_group의 그룹이 한 개만 존재:
6         continue
7
8     users = user_group에 속한 사용자
9     finished_users = {}
10    for user in users:
11        finished_users.add(user)
12        user_others = users - finished_users
13
14    for trajectory in day_group에서 user의 궤적:
15        if user_others가 공집합:
16            break
17        for other in user_others:
18            for item in day_group에서 other의 궤적:
19                if 두 궤적 데이터간 시간 차이 < 1분 & 평면 좌표계상의 거리 < 20미터:
20                    result.append(other 접촉자의 정보)
21                    finished_users.add(other)
22                    user_others = users - finished_users
23                break

```

Figure 3. Trajectory query processing for contact tracing

그림 3. 접촉자 추적을 위한 궤적 질의 처리

코드에서 4 라인의 `user_group` 은 `day_group` 을 사용자별로 나눈 그룹들을 저장한다. 만약 `user_group` 에 한 개의 그룹만 존재한다면 해당하는 날짜에는 단 한 명의 사용자에 대한 궤적 데이터만 기록되었다는 의미이다. 그러므로, 해당 날짜인 `day` 에는 접촉할 다른 사용자가 없어서 검색할 필요도 없으므로 다음 날짜로 넘어간다.

한 날짜에 동시에 여러 사용자의 궤적 데이터가 기록되어 있는 경우에는 한 사용자의 모든 궤적 데이터에 대해 다른 사용자들 각각의 궤적 데이터가 가까이 있는지 반복적으로 검사한다. 코드에서 8 라인의 `users` 는 날짜 `day` 에 궤적 데이터가 기록된 모든 사용자를 저장한다. 코드 9 라인에서는 검색이 종료된 사용자를 저장하는 `finished_users` 집합을 초기화한다.

코드에서 10 라인의 `for` 문은 `day` 에 해당하는 날짜에 기록된 궤적 데이터를 가지고 있는 사용자들 각각에 대해 반복적으로 검색을 수행한다. 먼저 현재 `user` 의 값을 `finished_users` 에 추가한 뒤에, 날짜 `day` 에 궤적 데이터가 기록된 모든 사용자인 `users` 에서 `finished_users` 를 뺀 차집합을 `user_others` 에 저장한다.

코드에서 14 라인의 `for` 문은 사용자 `user` 에 해당하는 모든 궤적들에 대해 다른 사용자의 궤적과 가까이 있는지를 반복적으로 검사한다. 이 때 만약 해당 날짜 `day` 에 대한 궤적을 가진 모든 사용자인 `users` 에서 `finished_users` 를 뺀 차집합인 `user_others` 가 공집합이라면 이미 모두 검색을 마쳤거나 접촉자로 분류된 경우이므로 궤적에 대한 검사를 종료한다. 코드에서 17 및 18 라인의 반복문은 `user_others` 집합에 속한 모든 사용자의 모든 궤적과 검사한다.

본 논문에서 같은 버스에 탑승했는지 판단하는 기준은 비교하는 두 궤적(`user` 의 궤적과 `other` 의 궤적)간의 시간 차이가 1 분 이내이고 평면 좌표계로 변환한 거리가 20 미터 이내로 설정하였다. 궤적의 저장 시간은 궤적을 저장하는 GPS 장치나 핸드폰에 따라 상이할 수 있으므로 완전히 같은 시각이나 위치를 사용한다면 일치하지 않을 수 있으므로 기간과 거리를 지정하였다. 만약 같은 버스에 동승한 사용자라고 판단되면, 접촉자로 분류하여 `result` 리스트에 궤적 정보를 저장하고 현재 `other` 의 값을 `finished_users` 에 추가한 뒤에, `users` 에서 `finished_users` 를 뺀 차집합을 `user_others` 에 저장하여 이미 접촉자로 검색된 사용자와 더 이상 비교하지 않도록 만들고 반복을 중지한다. 그 이후에는 `user_others` 에 속한 다음 사용자 `other` 와 비교를 반복한다.

검색의 결과로 41 건의 접촉을 확인할 수 있었으며 중복이 없는 접촉자 집합은 10 건이 검색되었다. 그림 4 는 검색된 결과의 일부를 보여준다. 언제 어디서 사용자들이 버스에서 서로



접촉했는지 확인할 수 있다. 시간과 공간 좌표는 정확히 일치하는 경우도 있고, 시간차와 거리가 같은 버스에 탑승했다고 판단하는 기준을 만족하는 경우도 있다.

```
[{'user': [139, 108],
  'time': [Timestamp('2007-10-07 10:08:55'), Timestamp('2007-10-07 10:08:55')],
  'lat': [39.9907166666667, 39.9907166666667],
  'lon': [116.312, 116.312],
  'time_diff': 0.0,
  'distance': 0.0},
...
{'user': [167, 126],
  'time': [Timestamp('2008-07-29 12:30:34'), Timestamp('2008-07-29 12:30:19')],
  'lat': [39.973169, 39.973005],
  'lon': [116.418446, 116.418488],
  'time_diff': 0.25,
  'distance': 18.58387964530111},
...]
```

Figure 4. Result of query processing

그림 4. 질의 처리 결과

#### IV. 결론

COVID-19, 메르스 등과 같은 전염병이 확산하는 상황에서는 접촉자 추적이 필요하므로 접촉자를 파악하기 위해서 시공간 질의 처리가 중요하다. 본 논문에서는 시간과 공간상에서 이동하는 객체의 특성 및 표현 방법에 대해 살펴보고 이동하는 객체가 사람인 경우의 종류에 대해 분류하였다. 이동의 종류 중에서 많은 인원이 같은 공간에 모일 수 있고 해당 인원들의 위치가 같이 변경되는 버스 탑승의 경우를 고려하였고, 동일한 버스에 탑승한 접촉자를 추적하기 위한 궤적 질의 처리 방법을 제안하였다. 시공간 궤적의 데이터로는 Geolife 데이터셋의 중국 베이징시 인근 지역에서 이동 방법이 버스인 데이터를 사용하였고, 개발을 위해서는 파이썬 언어와 Pandas 및 Geopandas 라이브러리를 사용하였다. 접촉자 검색을 위해서는 Geolife 궤적 데이터를 날짜별로 그룹을 만들어서 각 날짜에 대해 궤적 데이터가 기록된 사용자들 간에 시간 및 거리를 비교하였다. 검색된 결과를 통해서 언제 어디서 사용자들이 버스에서 서로 접촉했는지 확인할 수 있었다. 제안된 질의 처리 방법은 감염자 확산과 전파자 추적에 활용할 수 있다.

본 논문에서 제안하는 질의 처리 방법의 특징은 다음과 같다. 첫째, 전염병 상황에서 중요성이 증대된 접촉자 검색을 시공간 궤적 질의로 처리하였다. 둘째, 인원과 이동 거리에 따라 이동의 종류를 정의하고 이동 수단에 따른 질의 처리를 지원하였다. 셋째, 실제 데이터인 Geolife 궤적 데이터에 질의 처리 방법을 적용하여 접촉자를 선별할 수 있음을 보여주었다. 제안하는 방법의 제한점은 질의 처리를 위해 이동 수단에 대한 정보가 필요하며 민감한 개인 정보를 사용한다는 점이다.

향후 연구 방향으로서는 기존의 개별 이동 객체 중심의 시공간 인덱스 구조 또는 공간 분할 인덱스를 발전시켜서 접촉자 추적에 효율적인 성능을 제공할 수 있는 이동 방법을 고려한 시공간 인덱스에 대한 연구가 필요하다. 이동 수단에 대한 정보가 제공되지 않을 때에도 질의 처리를 수행하기 위하여 같이 이동하는 사람의 위치와 속도를 사용하여 이동 수단을 추측하는 연구도 가능하다. 정책적으로는 개인 정보 보호를 보장하면서도 전염병이나 재난 재해 상황에서 공공의 목적을 위해 휴대폰 위치 정보, 교통 카드 이용 내역, 신용카드 결제 내역 등을 활용하여 시공간 궤적 데이터를 활용할 수 있는 방안에 대한 연구가 필요하다.

#### V. 감사의 글

이 연구는 금오공과대학교 학술연구비로 지원되었음 (과제번호: 202002400001)

## VI. 참고문헌

- [1] B. T. Nixon, R. Alseghayer, B. Graybill, X. Zhang, C. Costa, and P. K. Chrysanthis, "Efficient Detection of COVID-19 Exposure Risk," in Proc. of the IEEE International Conference on Mobile Data Management (MDM), Jun. 2022, pp. 310-313.
- [2] P. Vepakomma, S. N. Pushpita, and R. Raskar. (2021). DAMS: Meta-estimation of private sketch data structures for differentially private COVID-19 contact tracing. Tech. Rep. [Online]. Available: [https://www.researchgate.net/profile/Praneeth-Vepakomma-2/publication/344238306\\_DAMS\\_Meta-estimation\\_of\\_private\\_sketch\\_data\\_structures\\_for\\_differentially\\_private\\_contact\\_tracing/links/5f5f6f8392851c0789652448/DAMS-Meta-estimation-of-private-sketch-data-structures-for-differentially-private-contact-tracing.pdf](https://www.researchgate.net/profile/Praneeth-Vepakomma-2/publication/344238306_DAMS_Meta-estimation_of_private_sketch_data_structures_for_differentially_private_contact_tracing/links/5f5f6f8392851c0789652448/DAMS-Meta-estimation-of-private-sketch-data-structures-for-differentially-private-contact-tracing.pdf)
- [3] M. E. Ali, S. S. Eusuf, and K. A. Islam. (2020). An Efficient Index for Contact Tracing Query in a Large Spatio-Temporal Database. [Online]. Available: <https://arxiv.org/pdf/2006.12812.pdf>
- [4] S. Wang, Z. Bao, J. S. Culpepper, and G. Cong, "A survey on trajectory data management, analytics, and learning," ACM Computing Surveys (CSUR), Vol. 54, No. 2, pp. 1-36, 2021.
- [5] M. M. Alam, T. Luis, and B. Albert. "A survey on spatio-temporal data analytics systems," ACM Computing Surveys, Vol. 54, No.10s, pp. 1-38, Nov. 2022.
- [6] H. Butler, M. Daly, A. Doyle, S. Gillies, S. Hagen, and T. Schaub. (2016). The geojson format (rfc7946). [Online]. Available: <https://www.rfc-editor.org/rfc/rfc7946>
- [7] B. Mohamed, S. Mahmoud, and Z. Esteban, "Distributed spatiotemporal trajectory query processing in SQL," in Proc. of the 28th International Conference on Advances in Geographic Information Systems, Nov. 2020, pp.87-98.
- [8] Y. Zheng, L. Zhang, X. Xie, and W. Y. Ma, "Mining interesting locations and travel sequences from GPS trajectories," in Proc. of International conference on World Wild Web (WWW 2009), Madrid, Spain, pp. 791-800.
- [9] Y. Zheng, Q. Li, Y. Chen, X. Xie, and W. Y. Ma, "Understanding Mobility Based on GPS Data," in Proc. of ACM conference on Ubiquitous Computing (UbiComp 2008), Seoul, Korea, pp. 312-321.
- [10] Y. Zheng, X. Xie, and W. Y. Ma, "GeoLife: A Collaborative Social Networking Service among User, location and trajectory," Invited paper, in IEEE Data Engineering Bulletin. Vol. 33, No. 2, pp. 32-40, 2010.

## 저자소개



**오병우(Byoungwoo Oh)**

1999년 2월 건국대학교 대학원 컴퓨터공학과 박사

1999년 6월~2004년 2월 한국전자통신연구원 텔레매틱스연구단 선임연구원

2004년 3월~현재 국립금오공과대학교 컴퓨터공학과 교수

관심분야: 공간 데이터베이스, GIS, 위치 기반 서비스