

# 체류 지점 식별을 통한 POI 기반 사용자 이동 경로 분석 기법

정소진\*, 정동원\*\*<sup>1</sup>, 이석훈\*\*<sup>2</sup>

## Method of User Route Analysis based on POI through Stay Point Identification

Sojin Jeong\*, Dongwon Jeong\*\*<sup>1</sup>, and Sukhoon Lee\*\*<sup>2</sup>

---

이 연구는 2019년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임  
(NRF-2019R1I1A3A01060826).

---

### 요 약

개인화된 위치기반 서비스 제공을 위해 사용자의 이동 궤적으로부터 관심 지점(POI, Point of Interest)을 추출하는 기술이 요구된다. 하지만 기존의 연구들은 POI를 위치정보로만 추출할 뿐, 실제 체류한 장소의 의미적 정보를 알 수 없다는 문제를 지닌다. 본 논문은 사용자 이동 궤적으로부터 체류 지점을 식별하고 POI에 기반하여 사용자의 이동 경로를 분석하는 기법을 제안한다. 이를 위하여 모바일 디바이스로부터 수집된 사용자 이동 궤적에 DBSCAN을 적용하고 사용자가 체류한 지점을 식별한다. 이후 정확한 POI 추출을 위해 POI 후보 리스트를 생성하고 Open API의 역지오코드 서비스를 통하여 상호 기반의 POI로 이루어진 이동 경로를 도출한다. 평가를 위해 사용자의 실제 이동 궤적을 분석한 결과, 제안 기법은 정밀도 92.08%와 재현율 96.21%로 평가되었다.

### Abstract

In order to provide a personalized location-based service, a technology for extracting point of interest(POI) from user trajectories is required. However, existing studies have a problem in that they only extract POI as location information and cannot know the semantic information of the actual place they stayed. This paper proposes a method for identifying the stay point from the user trajectory and analyzing the user route based on the POI. This paper applies DBSCAN algorithm to the user trajectory of the user collected from mobile device and the point where the user stayed is identified. After that, a POI candidate list is generated for accurate POI extraction and a route composed of business name based POI is derived through the reverse geocoder service of the Open API. As a result of analyzing the user's actual trajectory for evaluation, the proposed method was evaluated with a precision of 92.08% and a recall of 96.21%.

### Keywords

user route analysis, point of Interest, stay point, reverse geocoder, GPS trajectory

---

\* 군산대학교 소프트웨어융합공학과 학부생

- ORCID: <https://orcid.org/0000-0002-4863-2176>

\*\* 군산대학교 소프트웨어융합공학과 교수 (교신저자)

- ORCID<sup>1</sup>: <https://orcid.org/0000-0001-9881-5336>

- ORCID<sup>2</sup>: <https://orcid.org/0000-0002-3390-5602>

· Received: Aug. 25, 2021, Revised: Sep. 30, 2021, Accepted: Oct. 03, 2021

· Corresponding Authors: Dongwon Jeong and Sukhoon Lee

Dept. of Software Convergence Engineering, Kunsan National University,  
Korea

Tel.: +82-63-469-8912, Email: {djeong, leha82}@kunsan.ac.kr

## 1. 서 론

최근 스마트폰을 비롯한 사용자의 위치정보를 수집하는 다양한 기기의 증가에 따라 개인화된 위치 기반 서비스(LBS, Location-Based Service)들이 발전되고 있다[1]. 이러한 서비스를 제공하려면 사용자가 머물렀던 장소(Place), 방문 지점과 같이 사용자의 관심 지점에 대한 다양한 정보가 필요하다. 이에 따라 현재 개인의 이동 동선을 분석하여 사용자의 관심 지점을 추출하는 연구가 활발히 진행되고 있다[2]-[5]. 관심 지점(POI, Point of Interest)이란 사용자가 관심을 가질 수 있을 만한 위치에 대한 정보로 사람들이 자주 이동하는 장소나 사용자에게 의미 있는 장소를 뜻한다[6]. 예를 들어 관광지, 문화 시설, 음식점 등과 같은 주요 장소에 관련된 정보들을 POI라고 할 수 있다.

사용자 개인의 POI를 추출하기 위해서는 사용자들의 이동 궤적을 분석하여 체류 지점(Stay point)을 식별하는 것이 필수적으로 요구된다[7][8]. 이때 본 논문에서 체류 지점은 사용자가 일정 시간(Time) 동안 머무른 위치를 지칭한다[9]. 또한, 이동 궤적은 어떤 시점에서 수집된 사용자의 GPS 좌표로부터 연속적으로 좌표가 이동되는 데이터를 의미한다.

GPS를 기반으로 사용자의 이동 궤적을 수집할 때 사용자가 건물 내부와 같이 어느 한 지점을 들어가거나 나오면 위성으로부터 신호를 수신받지 못해 다량의 잡음이 발행하는 문제를 지닌다[10][11]. 이와 같은 잡음들이 포함된 사용자 이동 궤적으로부터 체류 지점을 식별하기 위해서는 클러스터링 기법을 적용할 수 있다. 클러스터링 기법을 이용하여 사용자의 이동 궤적 중 뭉쳐있는 부분을 하나의 군집으로 생성한다면 이러한 군집을 체류 지점으로 식별할 수 있다.

하지만 현재 식별된 체류 지점은 GPS의 위치정보만을 표현할 수 있으며, GPS 위치정보는 사용자가 실제 머무른 POI와 연결할 수 없다. 예를 들어 사용자가 한 상가 건물에 있는 음식점에 들어갔다고 할 때, GPS 위치정보만으로는 해당 건물의 위치만 알 수 있고 사용자가 실제 어떤 매장을 이용했는지는 알 수 없다.

사용자가 이전에 방문했던 POI 정보, 카드 사용

내역, 일정 정보 등과 같은 다양한 정보를 활용하면 식별된 사용자의 체류 지점과 정확한 POI를 연결할 수 있다[12]. 즉, 이러한 정보들은 사용자가 체류한 위치정보에서 실제 머무른 장소의 의미적 정보를 구체적으로 도출하는데 활용될 수 있다.

따라서, 본 논문은 사용자의 이동 궤적으로부터 체류 지점을 식별하고 POI에 기반하여 사용자의 이동 경로를 분석하는 기법을 제안한다. 먼저, 사용자의 이동 궤적으로부터 체류 지점을 식별하기 위해 클러스터링 기법의 하나인 DBSCAN(Density-Based Spatial Clustering of Applications with Noise)을 이용한다. 또한, 정확한 POI 추출을 위하여 사용자의 이전 POI 정보 및 카드 사용 내역 분석 등을 통한 POI 후보 리스트를 생성하고 식별된 체류 지점과 매칭시킨다. 이때 카카오 Open API 중 지도 API로부터 체류 지점의 상호 및 업종 정보를 받아온다. 그 결과 상호 기반의 POI로 이루어진 이동 경로가 도출된다.

모바일 디바이스를 이용하여 실제 사용자의 이동 궤적을 수집하였으며, 이러한 데이터를 기반으로 4가지 클러스터링 기법들(K-means, G-means, DBSCAN, Sequenced DBSCAN)의 체류 지점 식별의 정확성을 비교 평가한다. 또한, 제안 기법을 통해 사용자 이동 경로 분석의 정확성을 검증한다.

본 논문의 구성은 다음과 같다. 제2장에서는 관련 연구를 소개하고 해당 연구의 문제점을 제시한다. 제3장에서는 본 논문에서 제안하는 사용자 이동 궤적에서 체류 지점을 식별하고 사용자 POI 기반 이동 경로를 분석하는 기법에 대해 자세하게 기술한다. 제4장에서는 실험 결과 및 평가를 기술한다. 마지막으로 제5장에서는 결론을 기술한다.

## II. 관련 연구

이 장에서는 체류 지점 식별 기법과 POI 추출 기법에 관한 연구를 소개하고 해당 연구의 문제점을 기술한다.

### 2.1 체류 지점 식별 기법 연구

POI 추출을 위해서는 체류 지점을 식별하는 기

법이 요구된다. 이에 따라 본 논문은 사용자의 GPS 좌표들을 분석하고 체류 지점을 식별하는 기법을 이용한다. 따라서 GPS 좌표들을 클러스터링하여 사용자 체류 지점을 식별한다.

[13]은 GPS 좌표를 이용하여 단일 혹은 다중 사용자의 의미 있는 지점을 찾고 위치 이동을 예측하는 연구로서 체류 지점 식별을 위하여 K-means 알고리즘을 이용한다. 하지만 K-means 알고리즘은 클러스터 수를 사전에 결정해야 하는 한계가 있다.

[14]는 K-means가 지니는 한계를 극복하기 위하여 클러스터의 수를 자동으로 결정하는 G-means 클러스터링 기법을 이용한다. 이 연구는 스마트폰 사용자의 의미 있는 방문 POI 추출을 위하여 Naive Bayes classifier를 사용해 체류 지점을 추출하고 분할된 장소에 G-means 클러스터링 기법을 통해 관심 지점을 식별한다. 하지만 G-means 클러스터링 기법은 K-means를 반복적으로 수행해야 하므로 수행시간이 다른 기법에 비해 오래 걸린다는 단점이 있다.

[15]는 GPS 좌표를 기반으로 체류 지점과 이동 패턴을 분석하는 연구이다. 이 연구는 일정 시간 이상 머문 지역을 탐지한 후 체류 지점 식별을 위해 밀도 기반 군집화 알고리즘인 DBSCAN 기법을 이용한다. DBSCAN 기법은 본 논문의 제안 기법에서도 체류 지점 식별을 위하여 사용하는 알고리즘이다.

[16]은 임계값을 만족하는 구간의 가장 가까운 곳에 있는 좌표를 체류 지점으로 지정하는 nearSP 기법과 시간 계산으로 실내에 머무른 위치를 추출하는 indoorSP 기법을 제안한다. 이 연구는 일반적으로 사용자가 건물 안에 들어가면 GPS 좌표 데이터는 수신받지 못하므로 해당 시간부터는 데이터가 없는 것을 전제로 한다. 하지만 실제로 데이터를 수집 시 발생하는 잡음 등을 고려하지 않는다는 문제를 지닌다.

## 2.2 POI 추출 기법 연구

[17]은 GPS 좌표 정보를 기반으로 관심 지점과 이동 경로를 자동 추출하는 방법을 제안한다. 이 연구는 관심 지점 탐지를 위해 밀도 기반 DBSCAN 알고리즘을 사용한다. 하지만 이 연구는 GPS 위치 정보만을 이용하여 POI를 추출하기 때문에 POI의

의미적 정보를 알 수 없다.

[18]은 모바일 사용자의 대용량 이동 궤적 정보를 기반으로 시맨틱 POI 발견 방법을 제안한다. 이 연구는 WiFi 맵을 이용한 GPS 클러스터링 알고리즘을 통해 체류 지점으로 식별되면 이를 시맨틱 노드로 추출한다. 하지만 이 연구는 시맨틱 POI를 3가지 부류로만 구별할 수 있고 해당 POI가 실제 어떤 장소인지는 알 수 없다.

[19]는 GPS 로그 데이터에서 K-means 알고리즘과 수정한 Canopy 알고리즘을 적용하여 후보 POI를 뽑아 가장 오래 머문 GPS 지점을 최종 POI를 추출하는 방안을 제안한다. 또한, POI 추출에 있어 GPS 좌표와 시간, 행동 데이터 중 멈춤 상태를 특징으로 사용한다. 하지만 이 연구는 단순한 행동 상태와 가장 오래 머문 지점으로 POI로 추출하기 때문에 이를 사용자의 POI로 볼 수 없다.

이러한 연구들은 대부분 사용자의 POI를 추출하지만, 해당 POI에 대한 물리적 위치정보만 파악할 수 있고 의미적 정보가 없어 실제 사용자가 어떤 장소에 머물렀는지는 알 수 없다.

따라서 본 논문은 사용자 POI가 실제 장소의 상호와 같이 의미 있는 정보를 제공하여 사용자의 이동 경로를 분석하는 것을 목표로 한다.

## III. 체류 지점 식별을 통한 POI 기반 사용자 이동 경로 분석 기법

이 장에서는 제안 기법의 프로세스를 기술하고, 주요 프로세스 중 사용자 체류 지점 식별 방법, POI 후보 리스트 생성 방법, 그리고 이를 통한 POI 기반 사용자 이동 경로 분석 기법을 기술한다.

### 3.1 제안 프로세스

그림 1은 본 논문에서 제안하는 사용자 POI 기반 이동 경로의 분석 기법의 프로세스를 보이며 이는 크게 체류 지점 식별, POI 후보 리스트 생성, 이를 이용한 POI 기반 사용자 경로 분석으로 구성된다.

사용자의 이동 경로를 분석하기 위해서는 먼저 사용자의 데이터를 기반으로 체류 지점을 식별해야 한다.

#### 4 체류 지점 식별을 통한 POI 기반 사용자 이동 경로 분석 기법

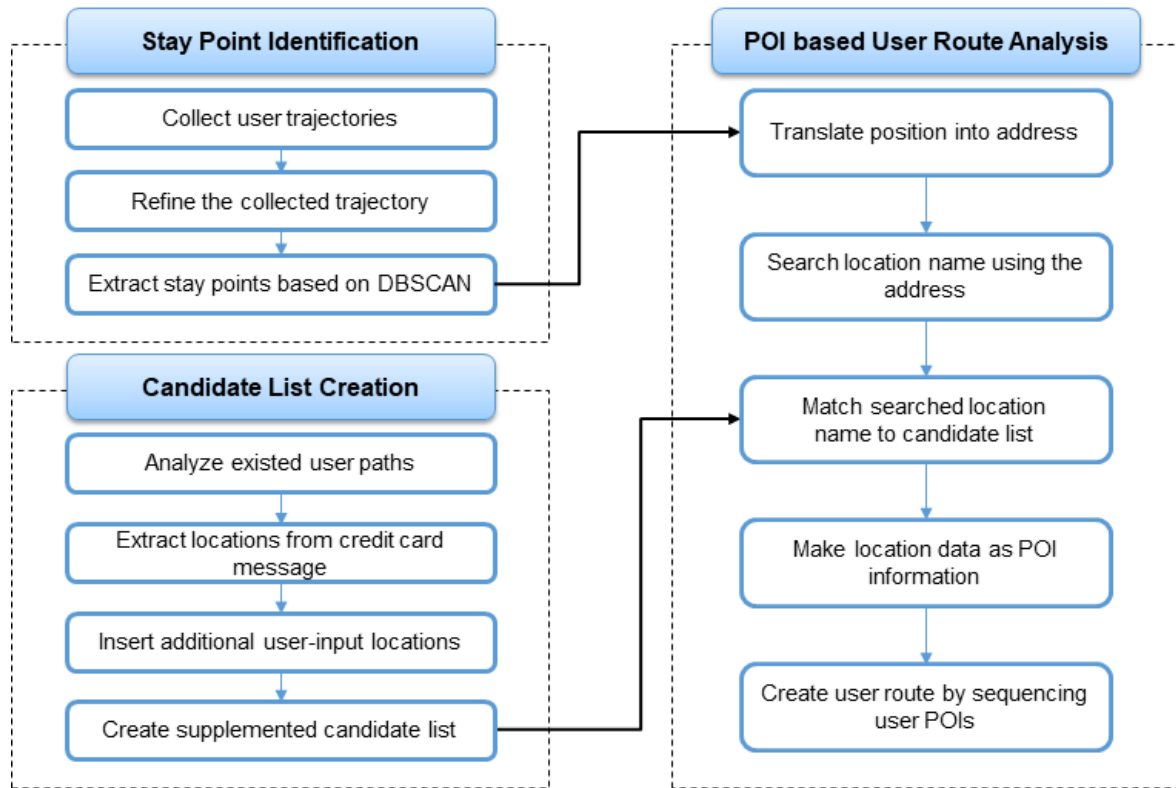


그림 1. 체류 지점 식별을 통한 POI 기반 사용자 이동 경로 분석 프로세스  
 Fig. 1. A process of user route analysis based on POI through stay point identification

이를 위해 모바일 디바이스에 내장된 GPS 센서를 이용하여 사용자의 이동 궤적을 수집한다. 수집된 이동 궤적은 데이터 정제 기법 등을 통하여 노이즈를 제거하고 날짜별로 구분하여 저장된다. 이후 DBSCAN 기법을 적용하여 사용자가 체류한 지점을 식별한다.

이러한 체류 지점들은 어떤 위치만을 나타낼 뿐 실제 상호나 의미적인 정보가 포함된 POI로 표현할 수 없다. 따라서 제안 기법은 식별된 체류 지점으로 부터 정확한 POI를 매칭하기 위해서 POI 후보 리스트를 이용한다. POI 후보 리스트는 사용자가 이전에 방문했던 POI 정보, 카드 사용 내역, 사용자 직접 입력 정보를 활용하여 생성된다.

POI 후보 리스트는 먼저 기존에 사용자가 이동했던 POI 분석 결과를 활용하여 생성한다. 또한, 사용자의 카드 사용 내역을 받아와 사용자가 이전에 방문했던 장소의 실제 상호를 추출하여 POI 후보 리스트에 추가한다. 집, 회사와 같이 상호 등으로 추출할 수 없는 정보들은 사용자로부터 위치정보와

함께 직접 입력받아 사용자 개인의 POI 후보 리스트를 보완한다. 이러한 후보 리스트는 사용자의 위치정보만 가지고 추출할 수 없는 실제 머무른 장소를 의미적 정보로써 도출하는 역할을 한다.

사용자의 POI에 기반한 이동 경로 분석은 앞서 식별한 체류 지점과 POI 후보 리스트를 매칭시킴으로 이루어진다. 식별된 체류 지점은 GPS 위치정보로 표현되고 POI 후보 리스트는 상호 및 의미적인 이름으로 표현되므로 두 정보의 매칭을 위해서는 각각을 주소(Address) 정보로 변환하는 것이 필요하다.

이를 위하여 먼저 사용자의 이동 궤적으로부터 식별된 체류 지점의 좌표를 Open API를 이용하여 주소로 변경하고 해당 주소로 검색되어 나오는 상호 리스트를 생성한다. 이렇게 생성된 사용자 체류 지점의 상호 리스트와 사용자 POI 후보 리스트를 비교한다. 이 두 개 리스트의 주소와 상호를 기반으로 매칭시켜 POI를 선정한다. 마지막으로 사용자의 이동 경로 분석 결과는 하루 동안 사용자가 이동한 순차적인 POI 리스트로 도출된다.

### 3.2 사용자 체류 지점 식별

사용자의 이동 궤적을 생성하기 위해 모바일 디바이스에 내장되어있는 GPS 센서를 기반으로 사용자의 GPS 좌표를 수집한다. 이는 수집한 날짜의 이름으로 텍스트 파일이 생성되며 모바일 디바이스 내부저장소에 저장된다. 저장된 텍스트 파일을 받아와 DB에 저장하고 해당 날짜의 데이터에 노이즈를 제거하는 정제 기법 등을 적용하여 사용자 이동 궤적으로 생성한다.

그림 2는 실제 수집한 사용자의 이동 궤적 중 일부분을 나타낸 그림이다. 이는 사용자가 건물 내부와 같이 어느 한 지점에 머무르기 위해 들어가거나 나올 때, 위성으로부터 신호를 수신받지 못해 그림 2와 같이 다량의 잡음을 포함하여 데이터가 수집된다. 이러한 데이터에 클러스터링 기법을 사용하여 잡음 부분을 사용자가 체류한 지점으로 식별할 수 있다.

사용자 이동 궤적에서 체류 지점을 식별하기 위해 체류 지점 식별 클러스터링 기법 중 DBSCAN 기법을 이용한다. DBSCAN은 데이터의 밀도에 기반하는 클러스터링 알고리즘으로 클러스터의 반경(Eps)과 최소 점의 개수(Minpts)가 매개변수로 주어진다[20]. 이는 하나의 점으로부터 EPS 내에 Minpts 만큼 점이 있으면 하나의 클러스터로 묶는다. 만약 클러스터에 속하지 않은 점들은 잡음(Noise)이라고 분류한다. 이처럼 DBSCAN의 두 가지 매개변수는 클러스터를 결정짓는 중요한 역할을 한다.

따라서 적절한 매개변수의 값을 찾기 위해 Elbow 기법의 군집 사이의 제곱 오차의 합(SSE)을 활용한다.



그림 2. 건물 체류 시 발생하는 잡음들  
Fig. 2. Noise generated while staying in the building

Elbow 기법이란 K-means 알고리즘의 적정 K값을 찾아주는 기법으로 K의 수를 하나씩 늘려 SSE 값이 감소하는 정도가 낮아질 때의 K를 구하는 기법이다[21]. 이와 같이 DBSCAN 기법의 적절한 매개변수를 찾아 사용자가 체류한 지점을 식별한다.

그림 3은 실제 수집된 사용자의 이동 궤적과 식별된 체류 지점을 보인다. 이는 사용자 이동 궤적에 DBSCAN 기법을 적용하여 클러스터링 된 지점을 시각화한 것이다. 이렇게 식별된 체류 지점에서 정확한 POI를 추출하여 사용자의 이동 경로를 분석할 수 있다.

### 3.3 사용자 POI 후보 리스트 생성

식별된 사용자의 체류 지점은 GPS 좌표만을 가지고 있다. 그러므로 사용자의 의미적 정보를 포함하는 POI를 추출하기 위해서 사용자 개인의 POI 후보 리스트를 사용해야 한다. 이를 위해 제안 기법은 사용자가 이전에 방문했던 POI 정보, 카드 사용 내역, 사용자 직접 입력 정보를 활용한다. 이는 사용자로부터 카드 사용 내역을 파일로 받아 자동으로 분석하여 사용자가 카드를 사용한 장소에 대한 내용을 받아온다.

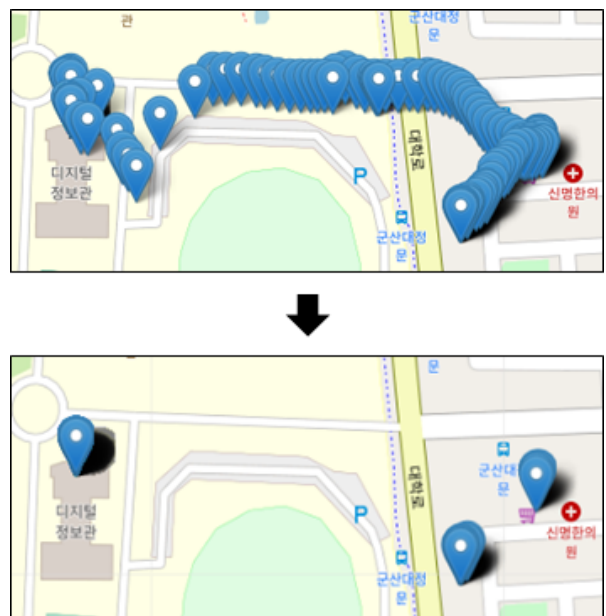


그림 3. 수집된 이동 궤적에서 체류 지점 식별  
Fig. 3. Identification of the stay point in the collected trajectory

Maps 기반 Open API를 활용하여 카드를 사용한 장소의 실제 상호, 유형(Type), 주소, 위도(Latitude), 경도(Longitude)와 같은 정보를 추가한다. 이를 사용자가 이전에 방문했던 장소들에 대한 사용자 POI 후보 리스트라고 명칭하고 생성한다. 이 리스트는 개별적 관리를 위해 CSV 파일로 저장된다.

표 1은 사용자의 카드 사용 내역을 기반으로 생성한 사용자 POI 후보 리스트의 스키마를 보인다. 장소는 상호와 같은 대표 이름을 나타내고 유형은 해당 장소의 카테고리를 나타낸다. 또한, 주소는 행정구역상 지번 주소, 위도와 경도는 해당 장소의 GPS 좌표를 나타낸다.

그림 4는 사용자의 다양한 정보를 기반으로 생성한 사용자 POI 후보 리스트의 예시이다. 이 리스트를 활용하여 사용자 체류 지점에서 사용자의 POI를 추출할 수 있다.

사용자가 추가하고자 하는 장소가 있을 시 사용자 POI 후보 리스트에 추가할 수 있어 좀 더 보완된 리스트를 생성할 수 있다. 만약 Open API를 통해 추출할 수 없는 개인적인 장소는 사용자로부터 위치정보와 장소 이름, 주소와 같은 추가적인 정보를 직접 입력받아 사용자 개인의 POI 후보 리스트를 보완한다.

표 1. 사용자 POI 후보 리스트의 스키마  
Table 1. Schema of user POI candidate list

Column	Definition
place	name of place
type	category of place
address	address of place
latitude	GPS coordinates
longitude	GPS coordinates

```
[[ '군산대학교 디지털정보관',
  '교육, 학문 > 학교부속시설',
  '전북 군산시 미룡동 290-2',
  '126.682784733311',
  '35.9449743765379',
  [ '다이소 군산대점',
    '가정, 생활 > 생활용품점 > 다이소',
    '전북 군산시 미룡동 879-1',
    '126.686322480081',
    '35.9449764638461' ],
  [ '익스태PC클럽',
    '가정, 생활 > 여가시설 > 게임방, PC방',
    '전북 군산시 미룡동 879-1',
    '126.686392528654',
    '35.945200153859' ] ] ]
```

그림 4. 사용자 POI 후보 리스트의 예  
Fig. 4. Example of user POI candidate list

하지만 상업적인 장소는 사용자로부터 해당 장소의 이름만 입력받아 Open API를 통해 자동으로 실제 상호, 유형, 주소, 좌표를 얻어와 리스트를 보완한다. 이러한 사용자 POI 후보 리스트는 사용자의 체류 지점에서 정확한 POI를 추출하여 실제 머무른 장소를 도출하는 역할을 한다.

### 3.4 POI 기반 사용자 이동 경로 분석

사용자의 POI에 기반한 이동 경로 분석은 앞에서 식별된 체류 지점과 사용자 POI 후보 리스트를 매칭시킴으로 이루어진다. 식별된 체류 지점은 GPS 위치정보만 가지고 있어 해당 건물의 위치만 알 수 있고 사용자가 실제 어떤 매장을 이용했는지는 알 수 없다. 따라서 식별된 체류 지점의 상호 리스트를 생성해야 한다.

식별된 체류 지점의 상호 리스트를 얻기 위해 카카오 Open API 중 지도 Local API를 사용하여 해당 체류 지점의 좌표를 지번 주소로 변환한다. 키워드로 장소 검색을 하는 API를 사용하여 앞에서 변환한 체류 지점의 주소를 기반으로 체류 지점의 상호 리스트를 생성한다.

그림 5는 Open API로 받아온 사용자의 체류 지점 좌표의 상호 리스트를 보인다. 먼저, 체류 지점의 좌표를 주소로 변경한 후 해당 주소에 있는 상호 리스트를 추출한다. 이는 사용자 체류 지점의 상호 리스트로 사용자가 머무른 장소의 주소에 대한 상호 후보군을 제공하고 있다.

체류 지점의 상호 리스트에서 실제 사용자가 머문 장소를 추출하기 위해 앞에서 생성한 사용자 체류 지점의 상호 리스트와 사용자 POI 후보 리스트를 비교하여 해당 체류 지점의 상호 리스트에서 POI를 추출한다.

먼저 생성된 사용자 체류 지점의 상호 리스트와 사용자 POI 후보 리스트를 비교한다. 두 리스트의 주소와 상호를 기반으로 매칭시킨다. 식 (1)의 Haversine 공식을 사용하여 사용자 체류 지점(S)의 좌표와 매칭된 사용자 POI 후보 리스트(P)에 있는 장소의 좌표 간 거리를 구하여 가장 가까운 장소를 최적의 POI로 추출한다.

```

1. 체류 지점 좌표 ->( 35.94525453228349 , 126.68274919212601 )
1. 체류 지점 주소 -> 전북 군산시 미룡동 290-2
상호명: ['군산대학교', '전북은행 군산대지점', 'GS25 군산대본점',
'OU 군산대도서관', 'OU 군산대학생회관점', '군산대학교 황흥도서관',
'군산대학교 제2학생회관', '군산대학교 대학본부', '군산대학교
학생생활관 관리동', '군산대학교 박물관', '군산대학교 디지털정보
관', '한울미디어', '동원문고', '군산대학교 황흥문화원', 'IBK기업은
행365 군산대학교출장소', '더큰도시락 군산대점', '군산대학교 음악
관', '군산대학교 중앙도서관', '군산대학교 문무관', '군산 청년해외
취업지원센터', '군산대학교 종합교육관', '군산대학교 평생교육원제2
동', '군산대학교 산업대학원', '사회관', '베토벤안정원', '군산시 쿨
어귀출지원센터', '군산대학교 교육대학원', '체험과여행사업단', '군
산대학교 인문대학사회과학대학', '황흥점', '군산대학교정류소',
'전북은행365코너 군산대지점', '군산대학교 자연과학대학4호관', '군
산대학교 간호학과', '군산대생협그라제에', '군산대학교 진리관', '군
산대학교 누리관', '군산대학교 테니스장1', '군산대학교 자연과학대
학 3호관', '군산대학교 남문', '카페디림 군산대학교중앙도서관점',
'군산대학교 자연과학대학1호관', '바른교육', '군산대학교 농구장2',
'군산대학교 축구장']

2. 체류 지점 좌표 ->( 35.94504822222222 , 126.68635173333331 )
2. 체류 지점 주소 -> 전북 군산시 미룡동 879-1
상호명: ['다이소 군산대점', '미니스톱 군산대점문점', '라라코스트
군산대점', '네이처리퍼블릭 군산대점', '익스타PC클럽', '머슬마켓
니스클럽', '천원세상D마트', '밀크밥버거 군산대점', '60계치킨 군산
나문미점점', '해피데이PC클럽', '감탄떡볶이 군산대점', '바이킹노래
연습장', '주차장', '후면빌딩']

3. 체류 지점 좌표 ->( 35.944606038095245 , 126.68576155714284 )
3. 체류 지점 주소 -> 전북 군산시 미룡동 882-4
상호명: ['no results']
    
```

그림 5. 체류 지점의 상호 리스트  
Fig. 5. List of business names of stay points

Haversine 공식이란 둥근 지구에 있는 두 지점의 위도, 경도를 이용하여 측정된 가장 짧은 거리를 구하는 공식을 말한다[22].

$$d = 2r \cdot \arcsin \sqrt{\sin^2\left(\frac{S_x - P_x}{2}\right) + \cos(P_x)\cos(S_x)\sin^2\left(\frac{S_y - P_y}{2}\right)} \quad (1)$$

추출된 사용자의 POI에 머무른 시간, 상호, 유형, 주소, 위도, 경도와 같이 다양한 정보를 추가하여 의미 있는 정보를 가진 POI로 생성한다. 이렇게 생성된 POI를 기반으로 사용자의 이동 경로를 분석할 수 있다.

그림 6은 추출된 POI에 다양한 정보를 추가하여

p_num	time	place	type	addr	lat	lon
0	1 2021-03-17 21:02:57 ~ 2021-03-17 21:13:37	군산대학교 디지털정보관	교육,학문 > 학교부속시설	전북 군산시 미룡동 290-2	35.944974	126.682785
1	2 2021-03-17 21:18:50 ~ 2021-03-17 21:19:25	다이소 군산대점	가정,생활 > 생활용품점 > 다이소	전북 군산시 미룡동 879-1	35.944976	126.686323
2	3 2021-03-17 21:21:58 ~ 2021-03-17 21:23:16	집	기타	전북 군산시 미룡동 882-4	35.944743	126.685730

그림 6. 생성된 의미 있는 POI 테이블  
Fig. 6. Created meaningful POI table

생성된 의미 있는 POI 테이블을 보인다. ‘p\_num’은 사용자의 POI가 추출된 순서로 해당 장소를 방문한 순서를 의미한다. 이를 통해 사용자의 이동 경로를 도출해 낼 수 있다. ‘time’은 해당 장소에 사용자가 머무른 시간을 의미하며 식별된 체류 지점의 시간 정보를 나타낸다. ‘place’는 해당 장소의 실제 상호를 의미하고 ‘type’은 해당 장소의 유형, 카테고리 의미한다. ‘addr’은 해당 장소의 지번 주소를 의미하고 ‘lat’과 ‘lon’은 해당 장소의 위치 좌표를 의미한다.

그림 7은 최종적으로 도출된 POI 기반 사용자 이동 경로를 보인다. 사용자의 POI를 마커로 표시하고 해당 마커에 팝업창을 띄워 사용자 방문 순서와 실제 장소의 상호를 보여준다. 이를 통해 시각적으로 사용자의 이동 경로를 확인할 수 있다.

따라서 사용자의 이동 경로는 사용자 이동 궤적에 클러스터링 기법을 사용하여 체류 지점을 식별하고, 사용자의 POI가 추출되는 순서를 기반으로 분석할 수 있다.

#### IV. 실험 및 평가

이 장에서는 실험 및 평가를 위해 모바일 디바이스를 이용하여 실제 수집한 이동 궤적을 기반으로 체류 지점 식별 실험 및 POI 추출 실험을 수행한다.

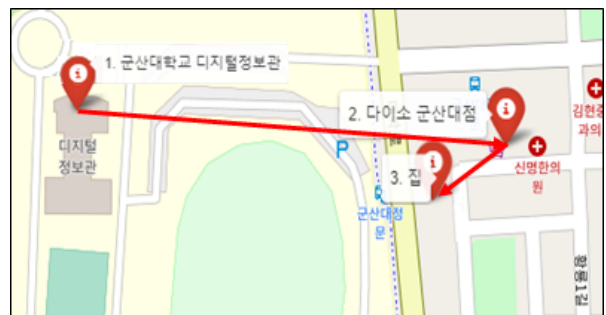


그림 7. POI 기반 사용자 이동 경로 도출  
Fig. 7. Deduction of POI based user route

먼저, 체류 지점 식별 실험을 통하여 클러스터링 기법들을 비교 평가한다. 또한, POI 추출 실험을 통하여 제안 기법을 통해 추출된 POI의 정확성을 검증한다.

4.1 실험 방법

먼저 사용자의 이동 궤적을 수집하기 위해 자체 개발한 모바일 앱을 사용하여 모바일 디바이스에 내장된 센서로 사용자의 GPS 좌표를 수집하였다. 또한, 성능 측정을 위하여 사용자가 직접 이동 경로 및 머물렀던 장소와 그 수를 기록하는 것으로 데이터셋을 구성한다.

그림 8은 실험을 자원한 피실험자들의 스마트폰을 이용하여 5초 주기로 수집된 데이터셋의 일부분을 보인다. 수집된 데이터는 모바일 디바이스에 내장된 센서로부터 측정된 위도, 경도, 3축 가속도(x, y, z) 데이터 및 데이터의 수집 시간을 포함하고 있다. 그 결과, 총 40일 동안 7명이 실험을 위해 자유롭게 이동하면서 수집하였으며 수집된 데이터는 총 44개의 이동 궤적이며 34,997행으로 이루어져 있다.

표 2는 실험 동안 사용자가 직접 머무른 장소의 위치와 수를 기록한 표 중 일부를 보인다.

time	lat	lon	x	y	z	speed	acttype
2021-03-08 21:04:41	35.9450527	126.6828516	1.901054144	5.877944469	5.49007225	8.264691849	RUNNING
2021-03-08 21:04:46	35.9450034	126.6828709	3.763799906	3.32804203	7.750267506	9.236270887	RUNNING
2021-03-08 21:04:51	35.9449446	126.6828889	0.974469841	4.704750061	6.114977837	7.776710008	RUNNING
2021-03-08 21:04:56	35.9448937	126.6829725	2.916226625	8.748680115	6.672843933	11.38290945	RUNNING
2021-03-08 21:05:01	35.9448555	126.6830365	1.613741279	6.656084061	4.764606953	8.343206539	RUNNING
2021-03-08 21:05:06	35.9448438	126.683076	0.577020228	5.229096413	5.736682415	7.783696215	RUNNING
2021-03-08 21:05:11	35.9448444	126.6831012	0.936161458	5.81090498	6.373559475	8.675556198	RUNNING
2021-03-08 21:05:16	35.9448128	126.6831698	0.090982445	6.584255695	6.313702583	9.122693746	RUNNING
2021-03-08 21:05:21	35.9447711	126.683219	1.027143836	5.411061287	7.620976925	9.402866478	RUNNING
2021-03-08 21:05:26	35.9447346	126.6832627	3.445361376	5.655277252	9.931451797	11.93676718	RUNNING
2021-03-08 21:05:31	35.9447211	126.6833228	4.580247879	2.365543604	12.82134151	13.81887352	RUNNING
2021-03-08 21:05:36	35.9447281	126.6833888	3.919427872	4.17800951	6.347222328	8.550140911	RUNNING
2021-03-08 21:05:41	35.9447543	126.683447	3.598595142	2.585816866	8.047157288	9.186570429	RUNNING

그림 8. 데이터 수집 결과  
Fig. 8. Data collection results

표 2. 사용자가 기록한 POI 이동 경로 및 개수의 예  
Table 2. Examples of POI route and count recorded by users

No.	POI route	POI count
1	A(식당) → B(카페) → C(피씨방)	3
2	C(피씨방) → F(집)	2
3	F(집) → B(카페) → E(학교 건물)	3
4	E(학교 건물) → D(생활용품점) → F(집)	3
5	E(학교 건물) → F(집) → D(생활용품점)	3
6	G(마트) → H(마트) → F(집)	3

이는 제안 기법을 통해 추출된 사용자의 POI와 정확성을 검증하기 위해 사용자로부터 직접 수집하였으며, 실험 동안 총 113개의 POI에 체류하였음을 확인할 수 있다.

4.2 체류 지점 식별 기법 비교 평가

이 실험은 수집된 사용자 이동 궤적에서 사용자 체류 지점을 식별하기 위한 4가지 체류 지점 식별 기법(K-means, G-means, DBSCAN, Sequenced DBSCAN)을 조사하고 실제 데이터에 적용해 본 결과를 바탕으로 비교 평가를 수행한다. 이때 4가지 체류 지점 식별 기법은 기존 연구에서 체류 지점 식별을 위해 쓰인 클러스터링 기법들로 선정하였다.

K-means는 임의의 K를 결정하여 데이터를 K개의 군집으로 형성하는 클러스터링 알고리즘이다[13]. 지정된 K개의 임의의 중심점으로부터 점들의 거리를 구해 가까운 점들을 하나의 클러스터로 생성하고, 각 클러스터에 속한 점들을 기반으로 중심점을 업데이트하는 방식이다.

G-means는 K-means를 반복적으로 수행하여 단점을 극복한 군집화 방법으로 데이터에 대해 임의의 K, 즉 클러스터의 수를 자동으로 결정하여 군집화하는 알고리즘이다[14].

DBSCAN은 본 논문에서 활용하는 기법으로 데이터의 밀도에 기반하여 밀도가 높은 부분을 클러스터링한다. 이 기법은 본 논문의 3.2절에서 자세히 기술하였다.

Sequenced DBSCAN은 기존에 있는 DBSCAN에서 시간이나 순서를 고려하여 클러스터링을 수행한다. 이는 처음에 무작위로 한 점을 선택한 후 그 점을 기준으로 슬라이딩 윈도우를 수행한다. 이웃 반경을 설정하고 설정된 이웃 반경과 반경 내에 최소 점 개수를 매개변수로 설정해 클러스터링하는 알고리즘이다[23].

성능 평가를 위하여, 각 기법의 정밀도(Precision), 재현율(Recall), F-measure를 측정하고, 어떤 기법이 가장 성능이 우수한지를 실험한다. 정밀도는 해당 기법이 식별한 체류 지점 중 사용자가 실제로 체류한 지점과 일치하는지 비율을 계산하며 식 (2)와 같다. 재현율은 사용자가 실제로 체류 지점의 정답 중



해당 기법이 맞춘 체류 지점 개수의 비율을 계산하며 식 (3)과 같다. F-measure는 정확률과 재현율의 조화를 평가하며 식 (4)와 같다.

$$precision = \frac{\text{기법이 맞춘 체류 지점 수}}{\text{기법이 식별한 체류 지점 수}} \quad (2)$$

$$recall = \frac{\text{기법이 맞춘 체류 지점 수}}{\text{사용자 실제 체류 지점의 정답}} \quad (3)$$

$$F\text{-measure} = 2 * \frac{precision * recall}{precision + recall} \quad (4)$$

그림 9는 각 기법의 정밀도, 재현율, F-measure를 측정된 결과를 보인다. 실험 결과 DBSCAN 기법의 정밀도는 90.78%, 재현율은 93.85%, F-measure은 92.27%로 다른 클러스터링 기법들보다 가장 우수한 성능을 보였다. 따라서 본 논문은 가장 좋은 성능을 보이는 DBSCAN 기법을 체류 지점 식별 기법으로 선정하였다.

### 4.3 POI 추출의 정확성 평가

이 실험은 체류 지점 식별 실험 결과에 따라 수집된 사용자 이동 궤적에 DBSCAN 기법을 적용하여 체류 지점을 식별한다. 이후 식별된 체류 지점에서 제안 기법을 통해 사용자의 POI를 추출하여 사용자의 이동 궤적을 분석한다. 이때 제안 기법의 정확성을 검증하기 위해 추출된 사용자의 POI와 실제 사용자가 기록한 POI를 비교하여 정밀도, 재현율을 측정한다.

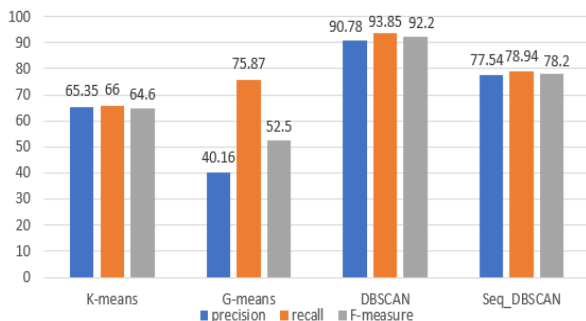


그림 9. 클러스터링 기법들의 평가 결과  
Fig. 9. An evaluation results of clustering algorithms

표 3은 제안 기법을 사용하여 추출한 POI와 사용자가 직접 기록한 실제 POI를 비교하여 나타낸 일부를 보인다. 이를 통해 제안 기법을 사용하여 추출된 POI가 실제 사용자의 POI를 도출할 수 있는지 검증하였다.

그림 10은 제안 기법을 사용하여 추출한 날짜별 POI의 정밀도와 재현율을 측정한 결과를 보인다. 정밀도는 해당 기법을 통해 추출된 POI 중 제안 기법을 통해 추출된 POI가 맞춘 개수의 비율을 계산하며 식 (5)와 같다. 재현율은 실제 사용자의 POI 중 제안 기법을 통해 추출된 POI가 맞춘 개수의 비율을 계산하며 식 (6)과 같다. 그 결과 평균 정밀도는 92.08%, 재현율은 96.21%를 보였다.

$$precision = \frac{\text{제안 기법이 맞춘 POI 수}}{\text{해당 기법을 통해 추출된 POI 수}} \quad (5)$$

$$recall = \frac{\text{제안 기법이 맞춘 POI 수}}{\text{실제 사용자의 POI 수}} \quad (6)$$

## V. 결 론

기존의 POI 추출에 관련된 연구들은 POI의 GPS 위치 정보만을 추출해 제시할 뿐 해당 지점의 의미적인 정보를 알 수 없다. 이러한 문제점을 해결하기 위해 본 논문은 체류 지점 식별을 통한 POI의 의미적 위치를 추출하고 이를 이용하여 사용자 이동 경로를 분석하는 기법을 제안하였다.

표 3. 실제 이동 POI와 제안 기법의 POI 비교의 예  
Table 3. A comparison example of actual moving POI and POI of the proposed method

No.	Seq	Labeled POI	Extracted POI
1	1	A(식당)	A(식당)
	2	B(카페)	B(카페)
	3	C(피씨방)	C(피씨방)
2	1	C(피씨방)	도로
	2	F(집)	F(집)
3	1	F(집)	F(집)
	2	B(카페)	B(카페)
	3	E(학교 건물)	E(학교 건물)
4	1	E(학교 건물)	E(학교 건물)
	2	D(생활용품점)	D(생활용품점)
	3	F(집)	F(집)

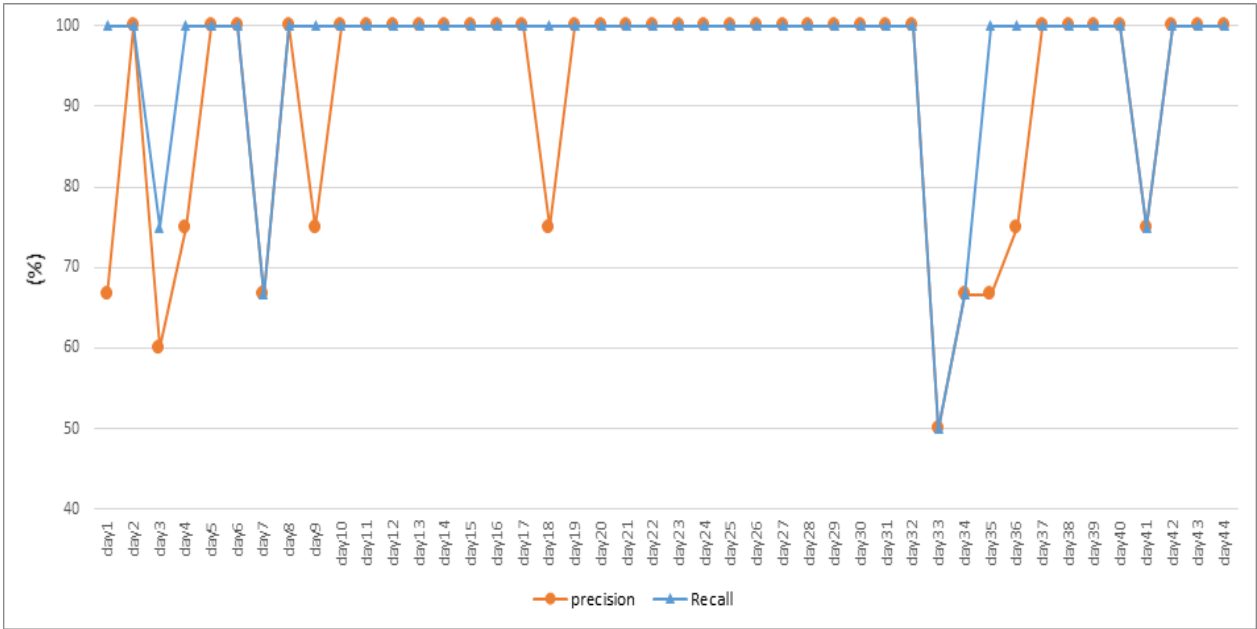


그림 10. 제안 기법의 날짜별 정밀도, 재현율 측정 결과  
 Fig. 10. Measurement results of precision and recall by date of the proposed method

이를 위해 수집된 사용자 이동 궤적에 DBSCAN 기법을 적용하여 사용자가 체류한 지점을 식별하였다. 이후 사용자의 다양한 정보를 활용하여 사용자 POI 후보 리스트를 생성하고 식별된 체류 지점과 매칭시켜 사용자의 POI를 추출하였다. 마지막으로 사용자의 POI가 추출되는 순서를 기반으로 사용자의 이동 경로를 분석하였다.

평가를 위하여 기존 연구들이 사용하고 있던 4가지 클러스터링 기반의 체류 지점 식별 기법을 비교 평가하였으며 본 논문에서 사용하는 DBSCAN 기반의 체류 지점 식별 기법이 가장 우수함을 검증하였다.

또한, 제안 기법의 성능을 평가하기 위하여 실제 사용자로부터 수집된 이동 궤적과 제안 기법을 통해 추출된 POI를 비교하여 정확성을 측정하였다. 그 결과 평균 정밀도는 92.08%, 재현율은 96.21%로 측정되어 제안 기법이 실제 사용자 이동 경로를 비교적 높은 정확도로 분석할 수 있음을 검증하였다.

제안 기법은 사용자 POI에 기반하여 지도 검색, 장소 추천 및 경로 예측 등과 같은 개인화된 다양한 위치 기반 서비스를 제공할 수 있을 것으로 기대된다. 또한, 감염병 등으로 인한 개인의 이동 동선 파악이 필요할 경우 자동으로 물리적 위치뿐만 아니라 의미적인 장소 정보를 도출해낼 수 있어 수

월하게 이동 경로 분석이 가능하다.

향후 연구로 사용자의 일정 정보, 예약 정보, SNS 정보 등과 같이 사용자의 다양한 추가 정보를 적용한 POI 추출 연구를 진행할 예정이다.

### Acknowledgement

2021년도 한국정보기술학회 하계종합학술대회에서 발표한 논문(사용자 이동 궤적에서 체류 지점 식별 기법들에 대한 연구)을 확장한 것임

### References

- [1] E. Cho, S. Park, Y. Park, J. Yoo, C. Park, Y. Lee, S. Lee, and K. Lee, "Mobile location based service and platform status", The Journal of the Korean Institute of Communication Sciences, Vol. 35, No. 4, pp. 67-76, Mar. 2018.
- [2] Y. Heo, "A Study on Method of Density Based Mixing POI Extracting System", Master Thesis, Mokpo University, Aug. 2015.
- [3] S. Kim, A. Bilal, and J. Kang, "A Survey on Sequential POI(Point-Of-Interest) Recommendation

- System Using Location-based Social Network Data", Proceedings of the Korean Information Science Society Conference, pp. 280-283, Dec. 2015.
- [4] R. A. Hamid and M. S. Croock, "A developed GPS trajectories data management system for predicting tourists' POI", TELKOMNIKA (Telecommunication, Computing, Electronics and Control), Vol. 18, No. 1, pp. 124-132, Feb. 2020. <http://dx.doi.org/10.12928/telkommnika.v18i1.13006>.
- [5] J. Suzuki, Y. Suhara, H. Toda, and K. Nishida, "Personalized Visited-POI Assignment to Individual Raw GPS Trajectories", ACM Transactions on Spatial Algorithms and Systems, Vol. 5, No. 3, pp. 1-28, Aug. 2019. <https://doi.org/10.1145/3317667>.
- [6] K. Kim, "A Disambiguation Method for Point of Interest Detection", International Journal of Tourism and Hospitality Research, Vol. 28, No. 11, pp. 25-36, Nov. 2014.
- [7] M. Jang, M. Yoon, and J. W. Chang, "A Survey on Moving Object Trajectory Mining Techniques in Location-based Services", The Korean Institute of Information Scientists and Engineers, Vol. 28, No. 1, pp. 67-88, Apr. 2012.
- [8] J. G. Park, and S. L. Oh, "Extraction method of Stay Point Using a Statistical Analysis", smart Media Journal, Vol. 5, No. 4, pp. 26-40, Dec. 2016.
- [9] G. Stylianou, "Stay-point Identification as Curve Extrema", Department of Computer Science European University Cyprus, Jan. 2017.
- [10] J. M. Kim, S. K. Yang, H. J. Baek, M. J. Jeon, and Y. T. Park, "GPS Noise Reduction and Trajectories Simplification for Personal Routes Learning in Close Range", Journal of KIISE : Computer system and theory, Vol. 39, No. 4, pp. 260-269, Aug. 2012.
- [11] J. M. Kim and Y. T. Park, "GPS-based Noise Reduction Approach for Trajectory Recognition", Journal of KISS:Software and Application, Vol. 39, No. 4, pp. 328-337, Apr. 2012.
- [12] J. Lin, O. Wolfson, and H. Yin, "Extracting Semantic Location from Outdoor Positioning Systems", Proceedings of the 7th International Conference on Mobile Data Management, Nara, Japan, May 2006. <https://doi.org/10.1109/MDM.2006.87>.
- [13] D. Ashbrook and T. Starner, "Using GPS to learn significant locations and predict movement across multiple users", In Personal and Ubiquitous Computing, Vol. 7, No. 5, pp. 275-286, Oct. 2003. <https://doi.org/10.1007/s00779-003-0240-0>.
- [14] S. Kim and S. Cho, "A Meaningful Visited-POI Extraction Method based on Spatial G-means clustering for Smartphone User", The Korean Institute of Information Scientists and Engineers, Busan, Korea, pp. 644-646, Jun. 2014.
- [15] M. Zignani and S. Gaito, "Extracting Human Mobility Patterns From GPS-based Traces", Proceedings of the 3rd IFIP Wireless Days Conference, Venice, Italy, Oct. 2010. <https://doi.org/10.1109/WD.2010.5657695>.
- [16] J. Park, S. Lee, and M. Jung, "Stay Point Extraction Method that Improve Accuracy of Location and Distinguish Between Indoors & Outdoors", Journal of the Institute of Electronics and Information Engineers, Vol. 52, No. 6, pp. 95-104, Jun. 2015. <https://doi.org/10.5573/ieie.2015.52.6.095>.
- [17] J. Kim and Y. Park, "POI Detection and Route Identification for Building Route Models for Smartphone Users", Journal of KISS: Software and Applications, Vol. 40, No. 12, pp. 799-808, Dec. 2013.
- [18] H. Lee, J. Choi, and Y. Park, "Semantic Point of Interest Detection from Large-scale GPS Data of Mobile Users", Journal of KISS: Software and Applications, Vol. 39, No. 3, pp. 175-184, Mar. 2012.
- [19] C. Kim, "Clustering algorithm for finding Point of Interests using GPS data", Master Thesis, Soongsil University, Dec. 2013.

[20] M. Ester, H. Kriegel, J. Sander, and X. Xu, "A Density based Algorithm for Discovering Clusters in Large Spatial Databases with Noise", In Proceeding of KDD-96, pp. 226-231, 1996.

[21] T. M. Kodinariya and P. R. Makwana, "Review on determining number of Cluster in K-Means Clustering" Int. J. of Advanced. Researches in Computer Science and Managment Studies, Vol. 1, No. 6, pp. 90-95, Nov. 2013.

[22] I. Yang, W. Jeon, and J. Moon, "A study on a distance based coordinate calculation method using Inverse Haversine Method", Journal of Digital Contents Society, Vol. 20, No. 10, pp. 2097-2102, Oct. 2019. <http://dx.doi.org/10.9728/dcs.2019.20.10.2097>.

[23] S. Kim and S. Lee, "Clustered Noise Filtering Based Trajectory Refinement Method", Journal of KIIT. Vol. 18, No. 3, pp. 11-20, Mar. 31, 2020. <http://dx.doi.org/10.14801/jkiit.2020.18.3.11>.

이 석 훈 (Sukhoon Lee)



2009년 2월 : 고려대학교  
전자및정보공학부(학사)  
2011년 2월 : 고려대학교  
컴퓨터·전파통신공학과(공학석사)  
2016년 2월 : 고려대학교  
컴퓨터·전파통신공학과(공학박사)  
2016년 3월 ~ 2017년 3월 :

아주대학교 의료정보학과 연구강사  
2017년 4월 ~ 현재 : 군산대학교 소프트웨어융합공학과  
부교수  
관심 분야 : 사물인터넷, 메타데이터, 센서 레지스트리,  
시맨틱 웹, 경로 예측

저자소개

정 소 진 (Sojin Jeong)



2018년 3월 ~ 현재 : 군산대학교  
소프트웨어융합공학과 학부생  
관심 분야 : 데이터 분석, 위치  
기반 서비스

정 동 원 (Dongwon Jeong)



1997년 2월 : 군산대학교  
컴퓨터과학과(학사)  
1999년 2월 : 충북대학교  
전자학과(석사)  
2004년 2월 : 고려대학교  
컴퓨터학과(박사)  
2005년 4월 ~ 현재 : 군산대학교

소프트웨어융합공학과 교수  
관심 분야 : 데이터베이스, 시맨틱 서비스, 빅데이터,  
사물인터넷, 엣지컴퓨팅, 지능형 융합 서비스