

A Technique for Generating Semantic Trajectories by Using GPS Positions and POI Information

Yuhee Jang[†] · Juwon Lee^{**} · Hyo-Sang Lim^{***}

ABSTRACT

Recently, semantic trajectories which combine GPS positions and POIs(Point of Interests) become more popular in order to expand location based services. To construct semantic trajectories, the existing algorithms exploit the extent information of POIs described as polygons and find overlapping regions between GPS positions and the extents. However, the algorithms are not applicable in the condition where the extent information is not provided such as in Google Map, Naver Map, OpenStreetMap and most of the open geographic information systems. In this paper, we provide a novel algorithm to construct semantic trajectories only with GPS positions and POI points but without POI extents.

Keywords : GPS Trajectory, Point of Interest, Semantic Trajectory

GPS 이동 궤적과 관심지점 정보를 이용한 시맨틱 궤적 생성 기법

장 유 희[†] · 이 주 원^{**} · 임 효 상^{***}

요 약

최근 위치기반서비스의 확장을 위해 GPS 위치정보에 관심지점(POI: Point of Interest) 정보를 결합한 시맨틱 궤적(Semantic Trajectory)이 주목받고 있다. 기존 연구의 경우 GPS 궤적과 POI의 면적정보(polygon)가 겹치는 경우를 찾아내어 시맨틱 궤적을 생성하였다. 하지만 구글 지도, 네이버 지도, OpenStreetMap 등과 같은 공개된 지리 정보 시스템에서는 POI의 면적정보를 제공하지 않기 때문에 기존 방법으로는 시맨틱 궤적을 생성하지 못하는 문제가 있다. 본 논문에서는 POI의 면적정보가 없는 제한적인 상황에서도 GPS 위치정보와 POI의 좌표값(points)만을 이용하여 시맨틱 궤적을 생성할 수 있는 기법을 제안한다.

키워드 : GPS 이동 궤적, 관심지점, 시맨틱 궤적

1. 서 론

최근 무선 통신 기술의 발달과 모바일 컴퓨팅 기술의 발전으로 사용자의 위치정보 수집이 용이하게 되었다. 수집된 위치정보를 이용하여 사용자의 이동 패턴 인식 및 위치 추적이 가능하게 되었고, 이를 통해 위치기반서비스(Location Based Service)가 발달하게 되었다. 다양한 서비스를 제공하

기 위해 GPS 위성으로부터 수신 받은 사용자의 물리적인 GPS 궤적을 분석하여 내포된 지식이나 새로운 정보를 발견하려는 노력이 이루어져 왔다[1, 2].

더욱 풍부한 지식의 발견을 위해 GPS 궤적이 어떤 의미를 내포하고 있는지를 설명하기 위한 추가적인 정보인 의미정보(contextual data)를 활용하려는 노력이 진행되었다. GPS 궤적에 의미정보를 더하여 생성한 궤적을 시맨틱 궤적이라 부른다[3]. 의미정보의 예로는 GPS 궤적에서 이동 객체가 정지했던 구간을 식별하는 어노테이션(annotation) 정보, 도로망(Road Network)과의 연계성, 날씨 정보, 이동 수단, 머물렀던 관심 지점(POI: Point of Interest)의 업종정보(category) 등이 있다. 이렇게 생성된 시맨틱 궤적을 분석에 활용함으로써 단순히 물리적인 위치만을 나타내는 GPS 궤적으로부터 얻을 수 없었던 다양한 지식을 발견할 수 있게 되었다. 간단한 예로 Fig. 1(a)와 같이 GPS 궤적으로 표현된 사용자의 이동경로를 통해 사용자가 움직인 목적을 추측한다고 가정했을 때, GPS 궤적만을

* 이 논문은 2014년도 정부(미래창조과학부)의 재원으로 한국연구재단의 일반연구지원사업 지원을 받아 수행된 것임(2012R1A1A1042875).
** 이 논문은 2015년도 한국정보처리학회 춘계학술발표대회에서 'GPS 이동 궤적과 관심지점 정보를 이용한 시맨틱 궤적 생성 기법'의 제목으로 발표된 논문을 확장한 것임.
† 준회원: 연세대학교 전산학과 석사과정
** 비회원: 연세대학교 전산학과 석사과정
*** 정회원: 연세대학교 컴퓨터정보통신공학부 교수
Manuscript Received: July 6, 2015
First Revision: August 17, 2015
Second Revision: August 24, 2015
Accepted: August 26, 2015
* Corresponding Author: Hyo-Sang Lim(hyosang@yonsei.ac.kr)

사용하여 사용자의 이동 목적을 판단하는 것은 한계가 있다. 하지만 Fig. 1(b)와 같이 GPS 궤적과 POI의 업종정보를 결합한 시맨틱 궤적을 이동 목적 추측에 활용할 경우, 사용자가 식사를 하기 위해 음식점으로 이동했다는 직관을 얻을 수 있다.

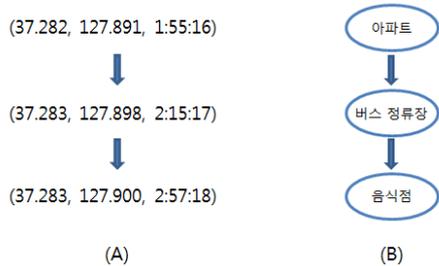


Fig. 1. GPS Trajectory and Semantic Trajectory

앞의 예에서와 같이, 의미정보로 사용자가 머물렀던 POI의 업종정보를 사용하는 시맨틱 궤적이 다양한 응용 분야에서 중요하게 사용될 수 있어 최근 많은 주목을 받고 있다. 응용의 한 예로 GPS 궤적을 사용자가 방문했던 장소의 타입 시퀀스로 변환하여 사용자 개인의 라이프 패턴을 찾으려는 연구가 있었다[4]. 또 다른 응용으로는 사용자의 이동 패턴을 발견하고자 GPS 궤적에 공간 정보의 타입(예: 도심, 여행지, POI 업종정보 등)을 결합한 시맨틱 궤적을 이용하였다[5]. 또 다른 연구에서는 사용자가 방문한 POI의 업종정보와 방문했던 시간정보를 통해 사용자의 행동을 예측하고자 하였다[6, 7]. 이외에도 사용자가 방문했던 POI의 업종정보의 패턴을 분석하여 사용자의 목적지를 예측하려는 연구 등이 있었다[8].

GPS 궤적과 POI 정보를 이용해 시맨틱 궤적을 생성하기 위해서는 사용자가 방문했던 POI를 추측하기 위한 방법이 필요하다. 기존의 방법[9]은 POI의 면적정보와 사용자의 GPS 궤적이 겹치는 경우 사용자가 그 POI를 방문했다고 판단하였다. 그러나 공개된 POI 정보 대부분은(예: 구글 지도, 네이버 지도, OpenStreetMap) 면적정보를 제공하지 않으며 POI의 위치를 표시하기 위한 중심 좌표값만을 제공한다. 따라서 기존의 방법으로는 공개된 POI 정보를 사용해 사용자가 방문했던 POI를 판단하는 것이 어렵다는 문제가 있다. 이를 해결하기 위해서 연구 [6]에서는 GPS 궤적과 POI 좌표값 간의 거리에 기반하여 사용자가 방문했던 POI를 판단하는 방법을 제시하였다. 하지만 이 방법은 GPS 수신시 발생하는 오차나 잡음(noise)을 고려하지 않았기 때문에 실제 환경에서 사용할 때 정확도가 떨어지는 문제가 있다.

본 논문은 모바일 사용자의 이동경로를 나타내는 GPS 궤적과 POI의 중심점 정보를 이용하여, 면적정보가 없는 제한적인 상황에서도 시맨틱 궤적을 생성할 수 있는 기법을 제안한다. 제안하는 기법은 먼저 GPS 궤적으로부터 속력을 계산하고, 계산된 속력으로 GPS 궤적에서 사용자가 정지했을만한 구간을 판별한다. 다음으로, 판별된 정지 구간의 각 지점을 주변의 POI 밀도에 따라 영역으로 확장시키고, 영역에 포

합되는 POI들을 사용자가 방문했을 것이라고 예상되는 후보 POI로 선정한다. 그리고 후보 POI들 중에서 정지구간과 거리상 가장 가깝고 가장 오랜 시간 영역에 포함되었던 POI를 사용자가 실제로 방문했을 것이라고 최종적으로 판단한다.

2. 관련 연구

물리적인 위치를 나타내는 GPS 궤적을 통해 사용자가 실제 특정 장소(POI)를 방문하였는지 판별하는 방법들은 궤적 마이닝(trajjectory mining)에서 지속적으로 연구되는 분야이다[3]. 시맨틱 궤적의 개념을 제시한 관련 연구[9]에서는 사용자 GPS 궤적으로부터 정지한 구간을 판별하는 데 면적정보를 가진 POI를 사용하였다. 특정 POI의 면적 내에 사용자 GPS 궤적이 임계시간 이상 포함된 경우 이를 정지 구간으로 식별하였고, 식별된 정지 구간에 POI 정보를 의미정보로 추가(annotation)하여 시맨틱 궤적을 생성하였다. 하지만 대부분의 공개된 지도정보에서와 같이 POI 정보에 면적정보가 없는 경우, 이 방법으로는 정지 구간과 사용자가 어떤 POI를 방문했는지 판단할 수 없다는 단점이 있다.

관련 연구[6]에서는 면적정보 없이 POI의 좌표정보만을 이용해 실제 사용자가 방문한 POI를 판별하는 방법을 제시하였다. POI의 면적정보를 유추하기 위해 각각의 POI 좌표정보들 간의 거리에 따라 영역을 분할하는 보로노이 다이어그램(voronoi diagram)을 사용하였다. 분할된 영역은 단 하나의 POI를 포함하는 다각형으로 구성된다. 사용자 궤적이 다각형을 지나가게 되면 사용자가 다각형에 포함된 POI를 방문했다고 유추하였다. 하지만 이 방법은 사용자의 궤적과 가장 가까운 하나의 POI만을 사용자가 방문했던 POI로 고려하고, 같은 방문 POI를 갖는 연속된 GPS 위치를 묶어서 하나의 정지구간으로 식별하기 때문에 POI가 밀집된 환경에서는 GPS의 오차로 인해 사용자가 실제 방문한 POI를 정확하게 유추하지 못하는 문제가 있다. 이를 개선하고자 본 논문에서는 단순히 거리가 가까운 하나의 POI만을 사용하는 것이 아니라 방문 가능성 있는 여러 개의 POI들 중 거리가 가장 가깝고 오랫동안 머물렀을만한 POI를 방문한 POI로 유추함으로써 GPS의 오차가 발생하더라도 사용자가 방문한 POI를 보다 정확하게 판별하는 기법을 제안한다.

3. 제안하는 기법

본 연구는 관련 연구들과는 순서를 반대로 하여, 먼저 사용자가 정지했을 것으로 예상되는 구간들을 식별하고, 식별된 정지 구간에서 사용자가 방문했을만한 POI들을 찾는다. 다음으로 정지 구간에서 사용자가 실제 방문했을 POI를 추정하는 순서로 시맨틱 궤적을 생성한다. 이는 POI의 면적정보를 구할 수 없는 현실적인 환경을 가정한 방법이다. 제안하는 방법은 Fig. 2와 같이 4단계를 거쳐 수행된다.

Step 1에서는 GPS 궤적으로부터 예측 정지 구간들을 식별한다. Step 2에서는 각각의 예측 정지 구간에서 사용자가

방문했을만한 후보 POI들을 선별한다. Step 3에서는 후보 POI들 중에서 비교적 오랜 시간 머물렀으며 거리가 가까운 POI를 실제 사용자가 방문했을 POI로 선정한다. Step 4에서는 방문했을 것이라고 판단되는 POI를 이용해 좌표 간의 이동으로 표현되었던 GPS 궤적을 방문한 POI 업종정보로 나타내는 시맨틱 궤적을 생성한다.

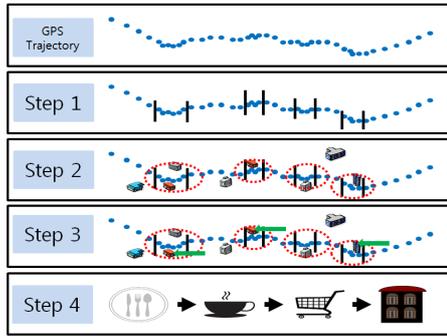


Fig. 2. Overview of the Proposed Method

3.1 Step 1 : 예측 정지 구간 식별

예측 정지 구간을 설명하기에 앞서 본 논문에서 사용하는 GPS 궤적과 POI 정보에 대해 서술한다. 사용자의 이동경로를 나타내는 GPS 궤적(GPS Trajectory)은 $GT = \{ P_1, \dots, P_i, \dots, P_n \}$ 로 표현한다. 궤적의 한 지점(position) P_i 는 $P_i = (id, (x, y), t)$ 로 나타내며, id 는 지점의 식별정보, (x, y) 는 각 지점의 좌표를 의미한다. 그리고 t 는 GPS 정보를 수신한 시점(timestamp)을 의미한다. POI는 $(pid, (x, y), c)$ 로 표현되며, pid 는 POI 식별정보를 나타내고, (x, y) 는 POI의 좌표 정보, c 는 업종정보를 의미한다. 그리고 사용자가 방문 가능한 모든 POI의 집합은 $POIset = \{ POI_1, \dots, POI_m \}$ 으로 표현된다.

예측 정지 구간(PSI: Predictive Stop Interval)이란 GT 에서 사용자가 정지했을 것이라고 예측되는 구간을 의미한다. 정지의 사전적 의미는 움직임을 멈춘 상태를 뜻하지만 GPS의 수신 오차로 인해 실제 사용자가 멈춰있더라도 속력이 0이 되는 경우는 드물다[3]. 또한 실제 사용자의 이동 수단은 차량, 대중교통, 도보 등 다양하지만 사용자가 원하는 장소를 최종적으로 방문할 때는 일반적으로 걸어서 이동한다. 따라서 제안하는 기법에서는 예측 정지 구간을 사람이 평균적으로 걷는 속도 이하로 움직인 구간으로 가정한다. 본 논문에서는 평균 걷는 속도로 1.3m/s를 사용하였는데, 이는 한국교통연구원의 2002년 조사자료[10]에서 한국인 성인의 평균 걷는 속도가 4.8km/h라고 한 것에 따른 것이다. 이에 따라 PSI 는 GT 의 P_i 에서 P_{i+i} 까지의 평균 속력이 $1.3m/s \cdot a$ 이하가 되는 구간으로 정의하며 Equation (1)을 만족하는 임의의 k 번째 예측 정지 구간은 $PSI_k = (P_i, P_{i+1}, \dots, P_{i+i})$ 로 표현된다. 그리고 GT 에서 식별된 PSI_k 의 집합은 $PSIset = \{ PSI_1, \dots, PSI_m \}$ 으로 나타낸다. 이때, a 는 $0 < a \leq 1$ 인 실수로 응용 환경에 따라 걷는 속도 기준을 조정하는 목적으로 사용한다. 또한 a 가 작아질수록 PSI 를 정밀하게 식별

하는 것이 가능하지만, a 이 너무 작아지게 되면 사용자가 실제 정지해 있었던 구간이 여러 PSI 로 분할되는 문제가 발생하기 때문에 a 의 값은 응용 환경에 따라 결정된다.

$$\frac{\sqrt{(P_{i+i}.x - P_i.x)^2 + (P_{i+i}.y - P_i.y)^2}}{P_{i+i}.t - P_i.t} \leq 1.3m/s \cdot a \quad (1)$$

3.2 Step 2 : 후보 POI 선별

다음 단계에서는 각각의 예측 정지 구간에서 사용자가 실제로 방문했다고 추정되는 후보 POI(CPOI: Candidate POI)들을 선별하기 위해 예측 정지 구간의 지점들을 영역으로 확장하고, 확장된 영역 내에 포함된 POI들을 후보 POI로 선별한다. GPS의 지점들은 GPS 오차로 인해 실제 사용자의 정확한 위치를 나타내지 않기 때문에 확장된 영역은 사용자가 머물렀을 가능성이 있는 모든 점을 포함한 영역이라고 가정한다. 설명의 편의성을 위해 k 번째 예측 정지 구간 PSI_k 에 포함된 한 점 P_i 를 $P_{k,i}$ 로 표현하고, $P_{k,i}$ 의 좌표 값을 중심으로 반지름이 R 인 원을 $Circle_{k,i} = (P_{k,i}, R)$ 로 표현한다.

지점들을 영역으로 확장하기 위해 PSI_k 의 모든 P_i 를 중심으로 $Circle_{k,i}$ 를 생성하고 생성된 $Circle_{k,i}$ 내부에 포함되는 POI들을 후보 POI로 선별한다. $Circle_{k,i}$ 를 모두 동일한 크기로 생성할 경우 POI가 밀집해 있는 지역에서는 너무 많은 후보 POI가 선별되어 사용자가 실제 방문한 POI를 판별할 때 비용(cost)이 증가하게 된다. 반대로 POI가 드문 지역에서는 후보 POI를 선별하지 못하거나 너무 적은 수의 후보 POI를 선별하여 사용자가 실제 방문한 POI가 후보 POI에 포함되지 않을 수 있다. 따라서 제안하는 기법에서는 일정 수의 후보 POI를 확보하는 것을 보장하기 위해 P_i 주변의 POI 개수에 따라 반비례하게 R 을 결정한다.

R 을 주변 POI의 개수에 따라 결정하기 위해서 먼저 전체 지역을 길이가 L 인 정사각형 셀(cell)로 분할하고, 셀에 포함된 POI의 개수를 POI 밀집도(POI density)로 정의한다. $Circle_{k,i}$ 를 생성하기 위해 $P_{k,i}$ 가 포함된 셀을 찾고, $P_{k,i}$ 가 포함된 셀의 POI 밀집도를 $POI density_{k,i}$ 로 표현한다. R 은 셀 내부에 포함된 $P_{k,i}$ 로부터 생성된 $Circle_{k,i}$ 가 셀에 포함된 POI들 중 최소 하나의 POI를 포함할 수 있도록 결정한다. 이를 위해 셀 내부의 POI들의 위치는 균일하게 분포되어 있다고 가정하고 R 을 결정하게 된다. 따라서 R 은 $P_{k,i}$ 주변에 POI가 많을수록 작게, 반대로 POI가 적을수록 크게 결정된다. Equation (2)는 $POI density_{k,i}$ 와 L 에 따라 R 이 어떻게 결정되는지를 나타낸다.

$$R = \frac{L\sqrt{2}}{2\sqrt{POI density_{k,i}}} \quad (2)$$

3.3 Step 3 : 후보 POI집합에서 방문 POI 선정

방문 POI(VPOI: Visited POI)는 $CPOIset_k$ 중 사용자가 실제로 머물렀을 가능성이 가장 높은 CPOI를 의미한다. 제안하는 기법에서는 VPOI 선정을 위해 CPOI마다 근접도

(proximity)와 머문 시간(duration)을 고려하여 결정한다.

근접도는 각각의 CPOI가 PSI_k 와 얼마나 가까운 거리에 있었는가를 나타내며, CPOI가 가깝게 위치할수록 근접도는 높다. 따라서 근접도 값은 $P_{k,i}$ 와 $P_{k,i}$ 를 중점으로 생성된 $Circle_{k,i}$ 내부에 포함된 CPOI 간 거리의 합으로 정의되며, 이 값이 작을수록 근접도는 높고, 이 값이 클수록 근접도는 낮다. Fig. 3에서 CPOI₁의 근접도 값은 $P_{k,1}$ 와 CPOI₁ 간의 거리와 $P_{k,2}$ 와 CPOI₁ 간의 거리 합이 된다. Equation (3)은 근접도의 정의를 나타낸다.

$$CPOI_i.proximity = \sum_i^{#of P_{k,i}} Euclidean\ Distance(P_{k,i}, CPOI_i) \quad (3)$$

머문 시간은 Equation (4)와 같이 CPOI가 $P_{k,i}$ 를 중심으로 생성된 각각의 원들에 포함되어 있었던 예측 시간(predictive time)의 합으로 정의한다.

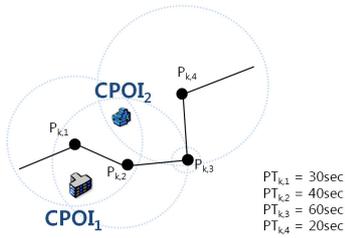


Fig. 3. Candidate POIs

$$CPOI_i.duration = \sum_i^{#of P_{k,i}} PT_{k,i} \quad (4)$$

예측 시간($PT_{k,i}$)은 PSI_k 의 한 지점 $P_{k,i}$ 에서 사용자가 멈추어 있었던 개략적인 시간을 의미하며, Equation (5)를 통해 계산된다. 예를 들어, Fig. 3에서 CPOI₁의 duration은 90sec (30sec + 40sec + 20sec)이다.

$$PT_{k,i} = \frac{(P_{k,i+1}.t - P_{k,i}.t) + (P_{k,i}.t - P_{k,i-1}.t)}{2} \quad (5)$$

$VPOI_k$ 는 $CPOIset_k$ 에서 근접도가 작으면서 머문 시간이 긴 CPOI가 선정된다. 하지만 근접도와 머문 시간을 나타내는 단위는 각각 거리, 시간으로 서로 다른 스케일을 가지기 때문에 Equation (6)과 같이 각각의 총합으로 정규화하여 사용한다. 또한 Equation (7)과 같이 $VPOI_k$ 의 머문 시간이 최소 임계시간($\Delta minDuration$) 미만인 경우 실제 사용자가 $VPOI_k$ 에서 머물렀다고 판단하기 어렵기 때문에 제거하게 된다.

$$VPOI_k = \text{MAX}_{CPOI_k \in CPOIset_k} \left(2 - \frac{CPOI_i.proximity}{\sum_{j=1}^{|CPOIset_k|} CPOI_j.proximity} + \frac{CPOI_i.duration}{\sum_{j=1}^{|CPOIset_k|} CPOI_j.duration} \right) \quad (6)$$

$$VPOI_k.duration < \Delta minDuration \quad (7)$$

3.4 Step 4 : 시맨틱 궤적 생성

본 논문에서의 시맨틱 궤적 ST (Semantic Trajectory)는 GPS 궤적 GT 로부터 식별된 PSI_k 에서 방문 POI로 선택된

POI 의 업종정보의 시퀀스를 의미하고 $ST = (VPOI_{1,c}, \dots, VPOI_{i,c}, \dots, VPOI_{m,c})$ 으로 표현된다. 시맨틱 궤적은 $PSIset$ 에서 각 PSI_k 를 $VPOI_i$ 의 업종정보로 변환시켜 생성한다.

제안하는 알고리즘

입력: $GT, POIset, L, \alpha, \Delta minDuration$

출력: ST

알고리즘

- 01: GT 로부터 속력 계산하여 V 에 저장
- 02: FOR EACH $v = V[i]$
- 03: $v[i] \sim v[i+1]$ 까지의 평균속력이 $1.3m/s \cdot \alpha$ 이하인 구간 식별
- 04: 식별된 구간으로 PSI 로 생성
- 05: $PSIset$ 에 PSI 를 추가
- 06: i 에 $i+1$ 저장
- 07: END FOR EACH
- 08: FOR EACH $psi = PSIset[i]$
- 09: FOR EACH $point = PSI[i]$
- 10: $point$ 가 포함된 셀의 $density$ 계산
- 11: $L, density$ 를 이용하여 R 결정
- 12: $point$ 를 중심으로 반지름이 R 인 $circle$ 생성
- 13: IF $PSIset$ 중 $circle$ 내에 포함되는 POI가 존재 THEN
- 14: $CPOIset_i$ 에 POI를 추가
- 15: END IF
- 16: END FOR EACH
- 17: END FOR EACH
- 18: FOR EACH $psi = PSIset[i]$
- 19: FOR EACH $cpoi = CPOIset[i]$
- 20: $cpoi$ 의 근접도와 머문시간 계산
- 21: END FOR EACH
- 22: MAX($cpoiset$)를 이용 $VPOI_i$ 선정
- 23: IF ($VPOI_i.duration > \Delta minDuration$) THEN
- 24: ST 에 $VPOI_i.c$ 추가
- 25: END IF
- 26: END FOR EACH

Fig. 4. Process of the Proposed Method

Fig. 4는 지금까지 설명한 4단계의 수행과정을 하나의 알고리즘으로 정리한 것이다. 먼저, GT 로부터 속력을 계산한다(line 1). 그 다음으로 PSI 를 식별하기 위해 Equation (1)을 만족시키는 모든 구간을 찾아 PSI 를 생성하고, 생성된 모든 PSI 를 순차적으로 $PSIset$ 에 추가한다(line 2~7). 그 다음으로 PSI 별 CPOI를 선별하기 위해 PSI 의 각 P 를 중점으로 원을 생성한다. 생성되는 원의 반지름 R 은 Equation (2)를 통해 결정되며, 원 내부에 포함되는 POI를 CPOI로 선별한다(line 8~17). 다음으로 Equation (6)과 같이 머문 시간이 길고 근접도 값이 작은 CPOI를 방문 POI로 선정하고, Equation (7)과 같이 선정된 방문 POI의 머문 시간이 최소 임계시간 미만인 경우 제거한다(line 18~23). 마지막으로, 최종 선정된 방문 POI의 업종정보를 ST 에 추가한다(line 24).

4. 실험 평가

본 절에서는 제안하는 기법의 유용성을 평가하기 위해 α 값을 변화시켜가며 실제 사용자가 정지해있었던 구간과 방문 POI에 대해 각각의 정확도(precision)과 재현율(recall)을

측정하는 간단한 실험을 수행한다.¹⁾ 그리고 보로노이 다이어그램(voronoi diagram)을 이용해 시맨틱 궤적을 생성하는 관련 연구 [6]의 방법과 정성적 비교를 통해 제안하는 기법의 우수성을 입증한다.

4.1 실험 환경

본 실험에서는 GPS 궤적 수집 디바이스로 스마트폰을 사용하였고, Google Map API를 사용해 POI 정보를 수집하였다. 또한 예측 정지 구간을 필터링 하기 위한 최소 임계 시간을 10분으로 설정하였고, POI 밀집도를 계산하기 위해 셀의 한 변의 길이 L 을 100m로 설정하였다. 즉, 전체 지역을 $100 \times 100m^2$ 크기의 셀로 분할하여 실험을 수행하였다. 반지름 R 은 Equation (2)에 따라서 계산된 값을 사용하였다. Table 1은 L 이 100m일 때 POI 밀집도에 따라 계산된 R 의 값의 예를 나타낸 것이다.

Table 1. Radius Example by Varying POI Density

POI 밀집도(POI density)	1	2	3	4	5	6
반경(R)	70m	50m	40m	35m	31m	28m

4.2 실험 데이터

실험 참가자의 이동 궤적을 수집하기 위해 참가자의 안드로이드 스마트폰을 사용하였고, 안드로이드 마켓에서 무료로 설치 가능한 GPSLogger 애플리케이션을 이용하였다. 수집된 GPS 정보는 GPX 파일 포맷으로 수집하고, GPX 파일로부터 위도, 경도, 수신 시점을 추출하여 실험에 사용하였다. Table 2는 실제 실험에 사용하였던 GPS 궤적의 예다.

Table 2. GPS Trajectory Example

위도	경도	수신 시점
37.351087	127.926989	2015-05-07T15:04:39Z
37.351122	127.927002	2015-05-07T15:06:43Z
...		
37.351263	127.927073	2015-05-08T14:52:12Z
37.351325	127.927133	2015-05-08T14:54:41Z
37.351267	127.927103	2015-05-08T14:57:08Z

실험에는 총 두 명의 사용자가 각각 하루 동안 수집한 이동 경로를 사용하였다. Table 3은 사용자들의 이동 경로 정보이다. 사용자 1은 POI가 밀집되어있는 지역에서 이동 경로를 수집하였고, 사용자 2는 POI가 드문 지역에서 이동 경로를 수집하였다. Fig. 5는 실제 면적이 $1,300m \times 650m$ 인 지역에서 사용자 1이 이동한 전체 경로 중 일부를 구글 지도 위에 나타낸 것이다.

본 실험에서는 사용자들이 활발하게 움직였던 지역에 대해 POI를 수집하였고, 총 5470개의 POI를 실험에 사용하였

Table 3. Summary of Users' Trajectories

	수집 개수	수신 간격	수집 시간	이동 수단	평균 POI 밀집도	실제 방문 POI
사용자 1	593개	120초	약 24시간	걷기, 차량	11.81	7개
사용자 2	635개	120초	약 20시간	걷기	3.65	10개

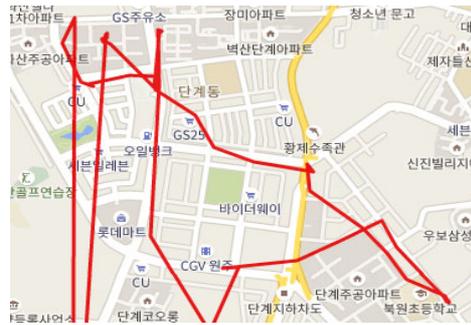


Fig. 5. Visualization of User1's Trajectory

다. 구글에서는 다양한 POI 정보를 제공하지만 실험에서는 POI의 식별정보(PID), 좌표(위도, 경도), 이름, 종류(업종 정보) 총 4가지의 정보만을 사용하였다. Table 4는 실험에 사용된 POI 정보의 예다.

Table 4. POI Example

PID	위도	경도	이름	종류
ChIJH2UdRlB0YzURs_w7N124CJM	37.345601	127.927461	펠리아	beauty_salon, establishment
ChIJH2UdRlB0YzURZ69vIcvHowg	37.345601	127.927461	롯데마트 원주점	store, establishment
...				
ChIJNSuWilB0YzURgGX PycXGXR0	37.346801	127.926745	북스타문고	book_store, store, establishment

4.3 실험 분석

첫 번째로, a 값을 변화시키면서 예측 정지 구간(PSI)의 정확도 및 재현율을 측정하였다. 이를 위해 전체 사용자의 궤적 중에서 이동 경로 수집 시에 사용자가 실제 방문했던 POI마다 기록했던 도착 시간과 떠난 시간 사이에 수신된 1135개의 점들을 정답 집합으로 사용하였다. 정확도는 알고리즘으로 구해진 PSI들의 점들 중 정답 집합의 비율을 측정하였고, 재현율은 정답 집합 중 알고리즘이 구한 PSI들 점들의 비율을 측정하였다.

Fig. 6은 a 값 변화에 따른 PSI의 정확도와 재현율의 변화를 보여준다. a 값이 0.5일 때 알고리즘에 의해 식별된 PSI 중 정답은 1118개, 오류는 51개, a 값이 0.2일 때의 정답은 1078개, 오류는 31개로 높은 정답률을 보임을 알 수 있다. a 값 변화에 따라서 결과가 달라지는 이유는 다음과 같다. 실제

1) 본 실험은 알고리즘의 정확성을 확인하기 위해서 제한적인 데이터와 환경에서 수행한 것으로, 알고리즘의 성능을 개선한 후 대량 데이터에 대한 실험을 수행할 계획이다.

GPS 궤적의 경우 위치정보 수신 시에 오차가 발생한다. 제안하는 알고리즘에서는 PSI를 연속된 점들 간 거리 차가 초속 $1.3 \cdot a$ m/s 이하인 구간으로 식별한다. 따라서 a 값이 작아지면 상대적으로 오차가 작은 점들도 PSI로 식별하지 않게 되어 높은 정확도를 얻게 된다. 하지만 반대로 실제 사용자가 정지해 있었던 구간에서 발생한 오차가 작은 점들 역시 PSI로 식별되지 않으므로 재현율은 낮아진다.

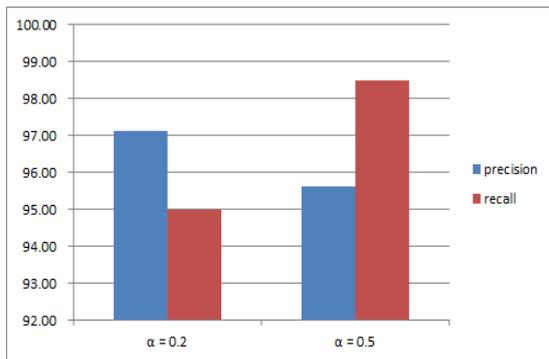


Fig. 6. Result of PSI by Varying Alpha

두 번째로 시맨틱 궤적의 정확도와 재현율을 측정하였다. 시맨틱 궤적의 정답 집합의 경우 사용자가 실제 방문했던 16개의 POI를 정답 집합으로 사용하였다. 정확도는 알고리즘이 생성한 시맨틱 궤적에 포함된 POI들 중 정답 집합의 비율을 측정하였고, 재현율은 정답 집합 중 시맨틱 궤적에 포함된 POI의 비율로 측정하였다.

Fig. 7은 a 값 변화에 따라서 생성된 시맨틱 궤적의 정확도와 재현율을 나타낸다. a 값이 0.5일 때 알고리즘에 의해 식별된 시맨틱 궤적 중 정답은 9개, 오류는 9개, a 값이 0.2일 때의 정답은 9개, 오류는 8개이다. a 값이 0.5일 때보다 0.2일 때 시맨틱 궤적의 정확도가 약 8% 정도 높게 측정되었다. a 값이 작아질수록 PSI 식별 시에 수신 오차로 인해 생기는 점들이 PSI에 포함되지 않기 때문에 정확도는 증가하게 된다. 생성된 시맨틱 궤적의 정확도와 재현율은 약 50%로 측정되었지만, 이는 복층 건물에 여러 POI가 함께 존재하는 경우가 많기 때문에 발생한 문제로, 이를 보정하기 위해 추가적인 실험을 수행한다.

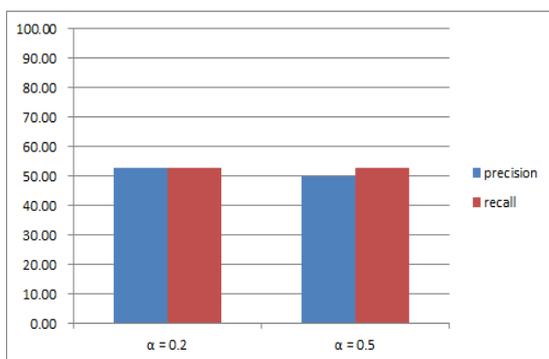


Fig. 7. Result of Semantic Trajectory by Varying Alpha

Fig. 8은 a 값을 0.2로 설정하고 생성된 시맨틱 궤적의 정확도와 재현율을 측정할 때, CPOI 중 상위 1~3순위의 CPOI를 사용하였을 때의 결과이다. 상위 3순위의 CPOI를 사용하였을 경우 알고리즘을 통해 발견된 정답은 10개, 오류는 3개로 정확도가 76.92%까지 높아지는 것을 확인할 수 있었다.

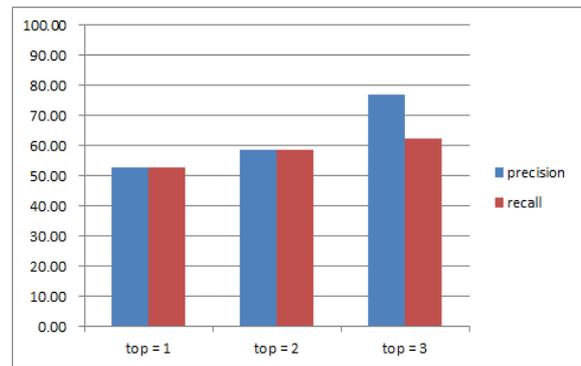


Fig. 8. Result of Semantic Trajectory by Ranking($\alpha = 0.2$)

4.4 관련 연구 [6]과의 비교 분석

본 절에서는 제안하는 기법의 우수성을 입증하기 위해 관련 연구 [6]에서 제안한 기존의 방법과 비교 분석을 수행한다. 기존 방법의 경우 사용자 궤적의 각 지점에서 거리가 가장 가까운 하나의 POI를 사용자가 방문한 POI로 유추하고 같은 방문 POI를 갖는 연속된 지점을 묶어 사용자가 정지한 구간으로 판별한다. 하지만 실제 사용자가 방문한 POI는 GPS 오차로 인해 거리가 가장 가까운 POI가 아닌 경우가 발생하여 실제 사용자가 정지했던 구간이 분할되는 문제가 있다. 이 문제를 개선하고자 제안하는 기법에서는 예측 정지 구간 단위로 사용자가 방문한 POI를 유추하였다. 이는 사용자가 방문한 POI를 유추할 때 여러 지점을 함께 고려함으로써 각 지점에서 발생한 GPS 오차를 구간 단위로 보정하기 위함이다.

비교 분석을 위해 기존의 방법은 보로노이 다이어그램의 특성을 반영해 GPS 궤적의 각 지점에서 거리가 가장 가까운 하나의 POI를 사용자가 방문한 POI로 유추하도록 구현하였다. 보다 상세한 비교 분석을 위해 실험에 사용된 전체 궤적 중 일부에 대해 분석을 수행한다. Fig. 9는 비교 분석에 사용된 궤적을 구글 지도에 나타낸 것이다. Restaurant A의 위치와 Restaurant B의 위치를 나타내는 두 점의 거리는 약 13m로 실제 두 POI는 인접해 있으며, 사용자는 Restaurant A에서 약 1시간 동안 머물렀다.

제안하는 기법의 경우 비교 분석에서 사용한 궤적에 대해 93%의 정확도로 예측 정지 구간을 판별하였고, 후보 POI로 선별된 9개의 POI 중 방문 POI를 Restaurant A로 판별하였다. 하지만 기존 방법의 경우 실제 사용자가 Restaurant A에 머물렀던 구간 중 4개의 지점에서 Restaurant B를 방문한 것으로 유추하여 9개의 구간으로 분할되었다. 이 결과를 통해 제안하는 기법이 기존의 방법보다 GPS의 오차나 노이

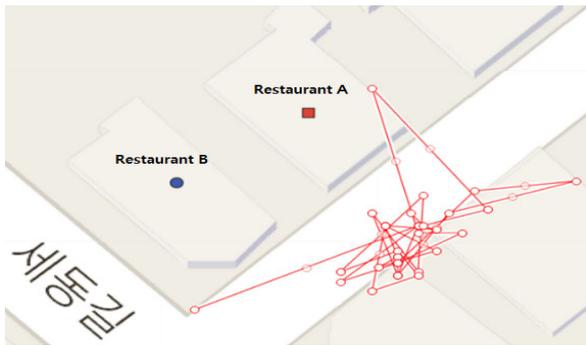


Fig. 9. Visualization of Semantic Trajectory Used Case Study

즈가 발생하였을 경우 비교적 정확하게 사용자가 방문한 POI를 식별해내는 것을 확인하였다.

5. 결론

본 논문에서는 공개된 지리 정보 시스템에서 제공하는 POI 정보와 사용자의 GPS 궤적을 이용해 사용자가 방문한 POI의 업종정보로 이루어진 시맨틱 궤적을 생성하는 기법을 제안하였다. POI 정보를 이용해 시맨틱 궤적을 생성하기 위해서는 사용자가 방문한 POI를 추정해야 한다. 기존의 방법 [9]은 사용자의 GPS 궤적과 POI의 면적정보가 겹칠 경우 해당 POI를 사용자가 방문한 POI로 추정하였다. 하지만 일반적으로 공개된 지리 정보 시스템에서 제공하는 POI 정보에는 위치를 나타내는 좌표정보만을 제공하기 때문에 이 방법으로는 시맨틱 궤적을 생성할 수 없는 문제가 있다. 또 다른 기존의 방법[6]은 POI의 좌표정보를 이용해 사용자가 방문한 POI를 추정하여 시맨틱 궤적을 생성하였다. 하지만 GPS 오차가 발생할 경우 사용자가 방문한 것으로 추정된 POI의 정확도가 낮아지는 문제가 있으며 실제 GPS 정보의 대부분은 오차가 포함되어 있다. 본 논문에서는 이런 두 가지 문제를 개선하고자 POI의 면적정보가 없는 제한적인 상황에서도 사용자가 방문한 POI를 보다 정확하게 추정하여 시맨틱 궤적을 생성하는 기법을 제안하였다.

제안하는 기법의 우수성을 검증하기 위해 구글에서 제공하는 좌표정보만을 가진 POI와 실제 수집한 사용자의 GPS 궤적을 사용하여 실험을 수행하였다. 실험의 결과로 약 50% 이상의 정확도로 사용자가 방문한 POI를 추정하였으며, 비교 분석을 통해 기존의 방법[6]보다 GPS 오차에 강인함을 보였다. 향후 연구로 더 많은 수의 GPS 궤적을 수집하여 실험 대상 지역을 확장하고 사용자가 방문한 POI를 정확하게 추정하기 위한 방법을 연구할 것이다.

References

[1] Giannotti, Fosca, et al., "Trajectory pattern mining," *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, 2007

- [2] Li, Quannan, et al., "Mining user similarity based on location history," *Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems*, ACM, 2008.
- [3] Parent, Christine, et al., "Semantic trajectories modeling and analysis," *ACM Computing Surveys (CSUR)*, Vol.45, No.4, (2013): 42.
- [4] Ye, Yang, et al., "Mining individual life pattern based on location history," *Mobile Data Management: Systems, Services and Middleware, 2009. MDM'09. Tenth International Conference on*, IEEE, 2009.
- [5] Alvares, Luis Otavio, et al., "Dynamic modeling of trajectory patterns using data mining and reverse engineering," *Tutorials, posters, panels and industrial contributions at the 26th international conference on Conceptual modeling—Volume 83, Australian Computer Society, Inc.*, 2007.
- [6] Xie, Kexin, Ke Deng, and Xiaofang Zhou, "From trajectories to activities: a spatio-temporal join approach," *Proceedings of the 2009 International Workshop on Location Based Social Networks*. ACM, 2009.
- [7] Renso, Chiara, et al., "How you move reveals who you are: understanding human behavior by analyzing trajectory data." *Knowledge and Information Systems*, Vol.37, No.2, pp.331-362. 2013.
- [8] Ying, Josh Jia-Ching, et al., "Semantic trajectory mining for location prediction," *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, ACM, 2011.
- [9] Alvares, Luis Otavio, et al., "A model for enriching trajectories with semantic geographical information," *Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems*, ACM, 2007.
- [10] 김희평, "걷는 속도," *문화일보 신문기사* [Internet], <http://www.munhwa.com/news/view.html?no=2007050401033037076002>. 5, 2007.



장유희

e-mail : yuheejang@yonsei.ac.kr

2014년 연세대학교 컴퓨터정보통신공학부
(학사)

2014년~현 재 연세대학교 전산학과
석사과정

관심분야: 데이터베이스, 정보보안, 궤적
마이닝



이 주 원

e-mail : juwonlee@yonsei.ac.kr
2015년 연세대학교 컴퓨터정보통신공학부
(학사)
2015년~현 재 연세대학교 전산학과
석사과정
관심분야: 데이터베이스, 정보보안, GPU를
사용한 데이터 처리



임 효 상

e-mail : hyosang@yonsei.ac.kr
1998년 연세대학교 컴퓨터과학과(학사)
1999년 한국과학기술원 전산학과(석사)
2007년 한국과학기술원 전산학과(박사)
2007년~2011년 Purdue Univ. Computer
Science Department(박사후 과정)
2011년~현 재 연세대학교 컴퓨터정보통신공학부 조교수
관심분야: 모바일 데이터베이스, 데이터스트림, 정보보안, 데이터
신뢰도 분석, GPU를 사용한 데이터 처리