

GPS 이동 궤적과 관심지점 정보를 이용한 시맨틱 궤적 생성 기법

장유희, 이주원, 임효상
연세대학교 원주캠퍼스 컴퓨터정보통신공학부
e-mail : {yuheejang, juwonlee, hyosang}@yonsei.ac.kr

A Technique for Generating Semantic Trajectories by Using GPS Moving Trajectories and POI information

Yuhee Jang, Juwon Lee, and Hyo-Sang Lim
Computer and Telecommunications Engineering Division, Yonsei University

요 약

모바일 환경에서 사용자의 GPS 궤적은 위치기반서비스(Location Based Service)에서 새로운 자원으로써 활용되고 있다. 위치기반서비스의 확장을 위해 단순히 사용자의 위치를 지도에 표시하는 것뿐만 아니라 사용자들이 위치했던 장소들이 내포하고 있는 의미를 발견해 내는 것이 필요하다. 이를 위해 최근 사용자의 위치정보에 관심지점(POI: Point of Interest)의 정보를 결합하여 시맨틱 궤적(Semantic Trajectory)을 생성하고 분석하는 연구들이 진행되고 있다. 이러한 기존연구의 경우 시맨틱 궤적을 생성하기 위해, 사용자의 GPS 궤적과 POI의 면적 정보(polygon)가 겹칠 경우를 찾아내서 이를 시맨틱 궤적으로 생성하였다. 하지만 대부분 공개된 POI 정보는 실제 장소들의 면적 정보를 제공하지 않고 좌표(point) 값만을 제공하기 때문에 기존의 방법으로는 시맨틱 궤적을 생성하지 못하는 문제가 있다. 본 논문에서는 사용자의 GPS 궤적과 POI의 좌표 값을 이용하여 사용자가 실제 방문했을 것으로 예상되는 POI를 추정하고 이를 시맨틱 궤적으로 생성해 내는 방법을 제안한다. 제안하는 기법은 GPS 궤적의 속력 정보를 사용하여 사용자가 정지했었던 구간을 판별하고, 정지 구간 주변의 POI 밀도에 따라 정지 구간을 영역으로 확장한다. 그리고 영역에 포함된 POI 중 정지 구간과의 거리가 가장 가깝고, 가장 오랜 시간 포함되었던 POI를 사용자가 방문했던 POI로 판단한다. 이 방법은 POI의 면적정보가 없는 제한적인 상황에서도 시맨틱 궤적을 생성할 수 있다는 장점을 가진다.

1. 서론

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 제공하기 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. 의미정보의 예로는 GPS 궤적에서 이동 객체가 정지했던 구간, 머물렀던 관심 지점(POI: Point

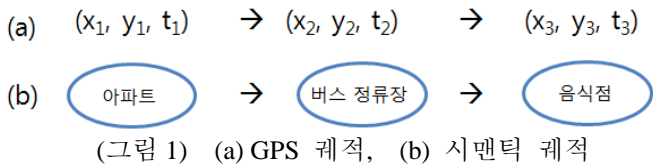
of Interest)의 업종정보, 도로망(Road Network)과의 연계성, 날씨 정보 등이 있으며, 물리적인 위치만을 나타내는 GPS 궤적 분석을 통해 얻을 수 없었던 새로운 지식을 발견하는데 유용하게 사용될 수 있다. 이때, GPS 궤적에 의미정보를 더하여 생성한 궤적을 시맨틱 궤적(semantic trajectory)이라 부른다[3]. 본 논문에서는 특히 사용자가 머물렀던 POI의 업종정보를 사용하여 시맨틱 궤적을 생성하는 것에 초점을 둔다.

예를 들어, 그림 1의 (a)와 같이 GPS 궤적으로 표현된 사용자의 이동경로를 통해 사용자가 움직인 목적을 추측한다고 가정을 했을 때, GPS 궤적만을 사용하여 사용자의 이동 목적을 판단하는 것은 한계가 있다. 그림 1의 (b)는 POI를 의미정보로 사용한 시맨틱 궤적의 예이다. 시맨틱 궤적을 목적 추측에 활용할 경우, 사용자가 식사를 하기 위해 음식점으로 이동했다는 것이 알 수 있다.

GPS 궤적과 POI 정보를 이용해 시맨틱 궤적을 생성하기 위해서는 사용자가 방문했던 POI를 예상하기

· 이 논문은 2014년도 정부(미래창조과학부)의 재원으로 한국연구재단의 일반연구지원사업 지원을 받아 수행된 것임(2012R1A1A1042875).

위한 방법이 필요하다. 기존의 방법은 POI 의 면적정보와 사용자의 GPS 궤적이 겹치는 경우에 사용자가 그 POI 를 방문했다고 판단하였다. 그러나 공개된 POI 정보의 대부분은(예: 구글 지도, 네이버 지도, OpenStreetMap) 면적정보를 제공하지 않으며 POI 의 위치를 표시하기 위한 좌표 값 만을 제공한다. 따라서 기존의 방법으로는 공개된 POI 정보로는 사용자가 방문했던 POI 를 판단하는 것이 어렵다는 문제가 있다. 또 다른 연구에서는 GPS 궤적과 POI 좌표 값 간의 거리에 기반하여 사용자가 방문했던 POI 를 판단하는 방법을 제시하였으나, 이때는 POI 의 면적정보가 고려되지 않기 때문에 정확도가 떨어진다는 문제점이 있다.



본 논문은 모바일 사용자의 이동 경로를 나타내는 GPS 궤적과 POI 좌표 정보만을 이용하여, 면적정보가 없는 제한적인 상황에서도 시맨틱 궤적을 생성하는 기법을 제안한다. 제안하는 기법은 GPS 궤적으로부터 속력을 계산하고, 계산된 속력으로 GPS 궤적에서 사용자가 정지했을 만한 구간을 판별한다. 판별된 정지구간을 주변의 POI 밀도에 따라 영역으로 확장시키고, 영역에 포함되는 POI 들을 사용자가 방문했을 것이라고 예상되는 후보 POI 로 선정한다. 후보 POI 들 중에서 정지구간과 거리상 가장 가깝고 가장 오랜 시간 포함되었던 POI 를 사용자가 실제로 방문했을 것이라고 최종적으로 판단한다.

2. 관련 연구

물리적인 위치를 나타내는 GPS 궤적을 통해 사용자가 실제 특정 POI 장소를 방문하였는지 판별하는 방법들은 궤적 마이닝(trajjectory mining) 에서 지속적으로 연구되는 분야이다[3]. 관련연구 [4]에서는 사용자 GPS 궤적으로부터 정지한 구간을 판별하는데 면적정보를 가진 POI 를 사용하였다. 특정 POI 의 면적 내에 사용자 GPS 궤적이 임계시간 이상 포함된 경우 이를 정지구간으로 식별하였고, 식별된 정지구간에 POI 의 의미정보를 추가(annotation)하여 시맨틱 궤적을 생성하였다. 하지만, 대부분의 공개된 지도정보에서와 같이 POI 의 면적정보 없이 좌표 정보만을 가지고 있는 경우에는, 이 방법으로는 정지구간과 사용자가 어떤 POI 를 방문했는지 판단할 수 없다는 단점이 있다.

관련연구 [5]에서는 면적정보 없이 좌표정보만을 가진 POI 에서 실제 사용자가 방문한 POI 를 판별하는 방법을 제시하였다. GPS 궤적 중 특정 부분 궤적과 거리상으로 가장 가까운 POI 를 사용자가 방문한 POI 로 판별하였다. 하지만, 단순히 거리만을 고려하였기 때문에 각각의 POI 가 나타내는 물리적인 실제 장소의 면적 차이를 반영하지 못한다는 단점이 있다.

3. 제안하는 방법

본 연구는 기존 연구들과는 순서를 반대로 하여, 먼저 사용자가 정지했을 것으로 예상되는 구간들을 식별하고, 다음으로 식별된 정지 구간에서 어느 POI 에 방문했는지를 추정하는 순서로 시맨틱 궤적을 생성한다. 이는 POI 의 부피 정보를 구할 수 없는 현실적인 환경을 가정한 방법이다. 제안하는 방법은 4 단계를 거쳐 수행된다. Step 1 에서는 GPS 궤적으로부터 예측 정지 구간들을 식별한다. Step 2 에서는 각각의 예측 정지 구간에서 사용자가 방문했을 만한 후보 POI 들을 선별한다. Step 3 에서는 후보 POI 들 중에서 가장 오랜 시간 머물렀으며 거리가 최소인 것을 실제 사용자가 방문했을 POI 로 선정한다. Step 4 에서는 방문했을 것이라고 판단되는 POI 를 이용해 좌표간의 이동을 표현되었던 GPS 궤적을 방문한 POI 업종정보를 나타내는 시맨틱 궤적을 생성한다.

3.1 Step 1: 예측 정지구간 식별

예측 정지구간을 설명하기에 앞서 본 논문에서는 사용되는 GPS 궤적과 POI 정보에 대해 서술한다. 사용자의 이동경로를 나타내는 GPS 궤적 $GT = \{ P_1, \dots, P_i, \dots, P_n \}$ 로 표현한다. 궤적의 한 지점 P_i 는 $P_i = ((x, y), t)$ 로 나타내며, (x, y) 는 각 지점의 좌표를 의미한다. 그리고 t 는 좌표에서 사용자가 위치한 시간을 의미한다. POI 는 $(pid, (x, y), c)$ 로 표현되며, pid 는 POI 식별정보를 나타내고, (x, y) 는 POI 의 좌표정보, c 는 업종정보를 의미한다. 그리고 사용자가 방문 가능한 모든 POI 의 집합 $POIset = \{ POI_1, \dots, POI_m \}$ 으로 표현된다.

예측 정지 구간(PSI)이란 GT 에서 사용자가 정지했을 것이라고 예측되는 구간을 의미한다. 실제 사용자가 정지해 있더라도 GPS 의 오차 때문에 속력이 0 이 되는 경우는 드물다[3]. 따라서 본 논문에서는 사용자가 정지한 구간을 사람이 걷는 속도(1.3m/s)이하로 움직인 구간으로 간주한다. 이에 따라 PSI 는 GT 의 P_i 에서 P_{i+l} 까지의 평균속력이 $1.3m/s \cdot a$ 이하가 되는 구간으로 정의한다. 이때, a 는 $0 < a \leq 1$ 인 실수로 응용 환경에 따라 걷는 속도 기준을 조정하는 목적으로 사용한다.

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

식(1)을 만족하는 임의의 k 번째 예측 정지 구간은 $PSI_k = (P_i, P_{i+1}, \dots, P_{i+l})$ 로 표현되며, GT 에서 식별된 PSI_k 의 집합은 $PSIset = \{ PSI_1, \dots, PSI_m \}$ 으로 나타낸다.

3.2 Step 2: 후보 POI 선별

다음 단계에서는 각각의 예측 정지 구간에 대해서 사용자가 실제로 방문했다고 추정되는 후보 POI 들을 선별한다. 본 논문에서는 POI 의 면적 정보가 주어지지 않는 경우를 가정하기 때문에, PSI_k 의 P_i 로부터 POI 의 면적정보를 반영할 수 있는 방법이 요구된다. 면적 정보를 반영하기 위해 PSI_k 의 P_i 를 중심으로 하는 원을 생성하여 POI 의 면적정보를 반영한다. 앞으

로의 설명의 편의성을 위해 PSI_k 의 P_i 를 $P_{k,i}$ 로 표현하고, $P_{k,i}$ 를 중심으로 반지름이 R 인 원을 $Circle_{k,i} = (P_{k,i}, R)$ 로 표현한다.

본 논문에서는 POI의 면적을 추정하기 위해서 주변 영역의 POI 밀집도를 사용한다. 만약 POI의 밀집도가 높다면 상대적으로 각 POI의 면적이 작을 가능성이 크고, 반대로 낮다면 면적도 넓을 가능성이 크기 때문이다. 이런 성질을 반영하기 위해 $Circle_{k,i}$ 생성시 $P_{k,i}$ 주변의 POI 개수에 따라서 반지름 R 을 가변적으로 결정한다. $P_{k,i}$ 주변 POI 개수를 판단하기 위해 사용자가 실제 움직일 수 있는 지역을 길이가 L 인 정사각형 셀(cell)로 분할한다. 셀에 포함된 POI의 개수를 POI 밀집도(density)로 정의하고, $P_{k,i}$ 가 포함된 셀의 밀집도는 $POI_{density}_{k,i}$ 로 표현한다. 그리고 R 을 결정할 때 각 셀 내부의 POI 분포는 균일하다고 가정하고, R 은 $Circle_{k,i}$ 이 적어도 하나 이상의 POI 포함할 수 있도록 결정된다. 식(2)는 $POI_{density}_{k,i}$ 와 L 에 따라 R 어떻게 결정되는지를 나타낸다.

$$R = L\sqrt{2/POI_{density}_{k,i}} \quad \text{식(2)}$$

그리고 $P_{k,i}$ 들로부터 생성된 모든 원에 포함되는 POI를 후보 POI 집합 $CPOI_{set}_k = \{CPOI_1, \dots, CPOI_i, \dots, CPOI_m\}$ 로 표현한다.

3.3 STEP 3: 후보 POI 집합에서 방문 POI 선정

PSI_k 로부터 생성된 $CPOI_{set}_k$ 중 에서 방문 POI($VPOI_k$)는 $CPOI$ 의 근접도(proximity)와 머문 시간(duration)을 고려하여 결정한다. 근접도는 $CPOI_{set}_k$ 의 $CPOI_i$ 와 PSI_k 의 모든 P 와의 평균 거리를 의미하고, 머문 시간은 $CPOI_i$ 가 PSI_k 의 각 P_i 를 중심으로 생성된 원에 포함되어 있었던 시간을 의미한다. 식(3)을 통해 $CPOI_{set}_k$ 에서 $VPOI_k$ 를 선정한다.

IF($CPOI_i.duration > \Delta minDuration$) AND 식(3)

$$\text{MAX}_{CSL_k \in CSL_{set}_i} \left(2 - \frac{CSL_k.proximity}{average\ proximity} + \frac{CSL_k.duration}{average\ duration} \right)$$

식 (3)에서 먼저, 평균 근접도(average proximity)와 평균 머문 시간(average duration)으로 근접도와 머문 시간을 나누어 정규화 시킨다. 그리고 근접도는 가장 작으면서 머문 시간이 가장 긴 $CPOI_i$ 를 $VPOI_k$ 로 선정한다. $VPOI_k$ 의 머문 시간이 최소 임계값($\Delta minDuration$)이하인 경우 실제 사용자가 $VPOI_k$ 에서 머물렀다고 판단하기 어렵기 때문에 제거된다.

3.4 Step 4: 시맨틱 궤적 생성

본 논문에서의 시맨틱 궤적 ST 는 GT 로부터 식별된 PSI_k 에서 방문 POI($VPOI$)로 선택된 POI의 업종정보의 시퀀스를 의미하고 $ST = (VPOI_{1,c}, \dots, VPOI_{i,c}, \dots, VPOI_{m,c})$ 으로 표현된다. 시맨틱 궤적은 PSI_{set} 에서 각 PSI_i 를 $VPOI_i$ 의 업종정보로 변환시켜 생성한다.

그림 2 는 본 논문에서 제안하는 기법에 대한 알고

리즘을 의사코드로 나타낸 것이다. 먼저, GT 로부터 속력을 계산한다(line 1). 그 다음으로 PSI 를 식별하기 위해 식(1)을 만족시키는 모든 구간을 찾고 PSI 를 생성하고, 생성된 모든 PSI 를 순차적으로 PSI_{set} 에 추가한다(line 2~7). 그 다음으로 PSI 구간별 $CPOI$ 를 선별하기 위해 PSI 구간의 각 P 를 중심으로 원을 생성한다. 생성되는 원의 반지름 R 은 식(2)를 통해 결정되며, 원 내부에 포함되는 POI 를 $CPOI$ 로 선별한다(line 8~17). 다음으로 식(3)과 같이 근접도가 가장 낮고 머문 시간이 가장 긴 $CPOI$ 를 $VPOI$ 로 선정한다(line 18~24). 마지막으로, 선정된 $VPOI$ 의 업종정보를 ST 에 추가한다(line 24~26).

제안하는 알고리즘

입력: $GT, POI_{set}, L, \alpha, \Delta minDuration$

출력: ST

알고리즘

- 01: GT 로부터 속력 계산하여 V 에 저장
- 02: FOR EACH $v = V[i]$
- 03: $v[i] \sim v[i+l]$ 까지의 평균속력이 $1.3m/s \cdot \alpha$ 이하인 구간 식별
- 04: 식별된 구간으로 PSI 로 생성
- 05: PSI_{set} 에 PSI 를 추가
- 06: i 에 $i+l+1$ 저장
- 07: END FOR EACH
- 08: FOR EACH $psi = PSI_{set}[i]$
- 09: FOR EACH $point = PSI[j]$
- 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 POI_{set} 중 $circle$ 내에 포함되는 POI 가 존재 THEN
- 14: $CPOI_{set}_j$ 에 POI 를 추가
- 15: END IF
- 16: END FOR EACH
- 17: END FOR EACH
- 18: FOR EACH $psi = PSI_{set}[i]$
- 19: FOR EACH $cpoi = CPOI_{set}[j]$
- 20: $cpoi$ 의 근접도와 머문시간 계산
- 21: END FOR EACH
- 22: $MAX(cpoi_{set}_i)$ 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

(그림 2) 제안한 기법 알고리즘

4. 예제 분석

본 절에서는 예제 분석을 통해 제안하는 방법의 유용성을 확인한다. 그림 4 는 사용자가 이동한 경로를 지도상에 표현한 것이다. 지도상의 점들은 오차를 고려한 사용자의 위치좌표를 의미하고, 검은색 테두리 점들은 사용자가 실제 정지했을 것이라고 예측되는

구간이다. 예제 분석에서 사용된 POI 들은 좌표 정보만을 가지고 있고, 실제 사용자는 POI2 를 방문했다고 가정한다.

먼저, POI 의 면적 정보를 사용하는 기존연구 [4]의 방법은 이러한 면적이 주어지지 않는 경우에는 사용할 수 없다. 그리고, GPS 궤적과 거리상으로 가장 가까운 POI 를 사용자가 방문한 POI 로 판별하는 기존 방법[5]의 경우 POI1 를 사용자가 방문하였을 것이라고 판별할 것이다. 이 방법의 경우 단순히 거리만을 고려하여 판단하기 때문에 방문한 것으로 예측되는 POI 를 잘못 예측한 것을 알 수 있다.

본 논문에서 제안한 방법으로 사용자가 방문했을 POI 를 판별한다면, 정지 구간으로 판단된 점 주변의 POI 밀집도를 이용해 가상의 POI 면적정보를 고려한 영역을 생성한다. 초록색 원은 가상의 POI 면적정보를 고려하여 생성된 영역의 결과이다. 생성된 원들로 이루어진 영역 내에 포함된 POI1, POI2, POI3 은 사용자가 정지 구간에서 방문했을 것이라고 예측된 후보 POI 들이다. 후보 POI 들중 정지 구간의 점들과의 거리가 가장 가깝고, 생성된 원들 내에 가장 오랜 시간 동안 포함된 POI2 를 실제 사용자가 방문한 POI 로 판별한다.



(그림 4) 예제 분석을 위한 사용자 이동 경로의 예

5. 결론

본 논문에서는 POI 의 좌표정보와 사용자 궤적 주변의 POI 밀집도를 고려하여 사용자 GPS 로부터 사용자가 방문한 POI 의 업종정보로 이루어진 시맨틱 궤적을 생성하는 기법을 제안하였다. 제안한 기법은 사용자 궤적 주변의 POI 밀집도를 고려하여 GPS 궤적을 영역정보로 변환하여 사용자가 방문했을 만한 POI 를 판별한다. 또한 대부분의 공개된 지도정보에서 제공하는 면적정보가 없는 POI 를 이용하여 비교적 정확한 시맨틱 궤적을 생성하였다. 다음 단계의 연구로, 실제 지도 데이터와 실제 GPS 궤적을 사용하여 제안하는 방법의 정확도를 측정할 계획이다.

참고문헌

- [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) 45.4 (2013): 42.
- [4] 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.
- [5] 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.