

# GPS 데이터 분포를 고려한 실내 Stay Point 추출 방법

박진관\*, 최상길\*\*, 백종길\*\*, 정민아\*, 이성로\*\*

\*목포대학교 컴퓨터공학과

\*\*목포대학교 전자공학과

e-mail:chrjspj@mokpo.ac.kr

## Extraction Method of Indoor Stay Point considering the Distribution of GPS Time Data

Jin-Gwan Park\*, Sang-Gil Choi\*\*, Jong-gil Baek\*\*, Min-A Jeong\*, Seong-Ro Lee\*\*

\*Dept of Computer Engineering, Mok-po University

\*\*Dept of Electronics Engineering, Mokpo National University

### 요 약

최근 모바일 기기의 발전으로 사용자의 위치를 수집하고 분석하는 방법들이 연구되고 있다. 이러한 방법들 중 하나인 궤적 데이터 마이닝은 사용자의 궤적을 바탕으로 의미 있는 정보를 추출하기 위해 사용된다. 궤적 데이터 마이닝을 수행하기 위해서는 사용자의 GPS로그를 분석하여 Stay Point를 추출하는 과정이 선행되어야 한다. 기존의 Stay Point 추출 방법은 실내와 실외의 Stay Point를 구분하지 못한다. 본 논문에서는 기존의 Stay Point 알고리즘을 보완하기 위해 GPS 데이터 분포를 고려하여 실내에서 머무른 지점만을 추출하는 Stay Point 알고리즘을 제안한다.

### 1. 서론

최근 들어 모바일 사용, 특히 스마트폰이 우리 생활 곳곳에 확산되면서 모바일 기기에서 수집된 위치 데이터를 통해 사용자의 이동 패턴 인식 및 위치 내역 추적이 가능하게 되었다. 이를 통해 위치기반서비스(LBS: Location-Based Service)가 발전하고, 수많은 개인화 서비스들이 제공되고 있다[1]. 수집된 수많은 공간 데이터들은 매우 많은 의미를 내포하고 있는데, 중요한 점은 이러한 공간 데이터를 분석함으로써 사람들을 위한 의미 있는 장소를 추출해 낼 수 있다는 것이다[2]. 대부분의 사람들이 일정 공간에 머무는 곳이 의미 있는 장소라고 할 수 있는데, 의미 있는 장소란 문화적으로 중요한 장소임을 뜻한다. 예를 들어 뉴욕에 있는 자유의 여신상과 자주 찾는 공공장소(쇼핑 몰, 식당, 사무실, 학교, 공원 등등)가 의미 있는 장소가 될 수 있다. 이러한 정보는 사용자가 주변의 위치를 이해하는데 도움이 되고, 여행을 할 때 관광 명소를 추천 받을 수 있다[3,4]. 이러한 의미 있는 장소를 Place of Interest 혹은 Point of Interest(이하 PoI)[2]라고 하며 PoI를 추출해 내기 위해서는 우선 개인의 이동 궤적을 분석하여 머무른 지점(Stay Point or Stay location 이하 SP)을 알아내야 한다. SP는 개인이 일정 시간 동안 일정 장소에 머무른 곳을 뜻한다[3]. 이처럼 머무른 장소, 즉, SP는 개인의 의미 있는 장소를 추출한 것이고, PoI는 개인의 SP를 병합한 후 포괄적으로 의미 있는 장소를 선별한 것이다. 아울러, 이러한 정보는 경로

예측, 영업 홍보, 도로 확장 등의 다양한 분야에서 유용하게 사용 될 수 있다[5].

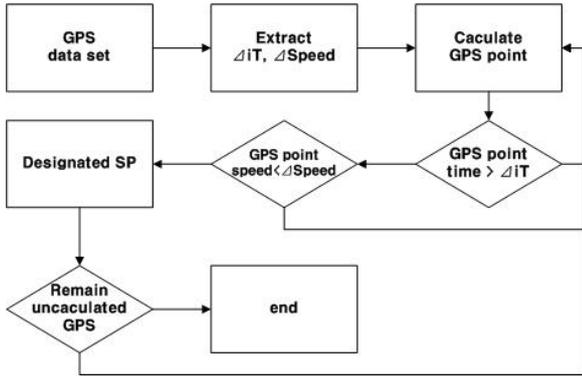
기존의 SP 추출 방법[6]은 실내와 실외를 구분하지 않고 모두 SP로 지정하기 때문에 육안으로 분간하기가 쉽지 않으며, 전체 궤적의 특성을 반영하지 않고 SP를 추출한다.

본 논문에서는 기존의 SP 추출 알고리즘을 보완하여 실내에서 머무른 지점을 추출할 수 있는 indoorSP 알고리즘을 제안한다. 제안하는 indoorSP 알고리즘은 전체 궤적의 GPS 포인트간 시간에 정규 누적분포함수(Cumulative Distribution Function : CDF)의 역함수를 통해 생성된 확률변수값을 시간 임계값으로 지정하고, 전체 궤적의 평균 속도를 속도 임계값으로 지정한다. 시간 임계값을 초과하는 GPS 포인트의 속도를 계산한 후 지정된 속도 임계값 미만이면 실내에서 머무른 지점으로 추출된다.

### 2. 본론

제안한 알고리즘은 전체 궤적에서 시간 임계값  $\Delta t$ , 속도 임계값  $\Delta speed$ 를 추출한 후, GPS 포인트간 시간이  $\Delta t$ 를 초과하고 속도가  $\Delta Speed$  미만이면 실내에서 머무른 지점으로 지정한다. 왜냐하면, 유저가 실내로 들어가서 GPS 신호가 끊긴 후, 다시 실외로 나와서 GPS 신호가 연결되는 상황이 실내에서 머무른 곳이 되기 때문이다[7]. 일정 간격으로 수집되는 시간 외의 시간은 유저가 실내로 들어갔거나, GPS 신호가 잡히지 않는 음영지역을 지나가

거나, 혹은 GPS 신호자체가 lost 된 경우이다. 유저가 건물 내부로 들어갔다가 다시 입구로 나오면 GPS 포인트 간격의 속도가 사람이 평균적으로 걷는 속도인 1.4 m/s 보다 낮게 측정된다. 하지만 음영지역을 지나가거나 GPS 신호가 lost 된 경우의 속도를 측정하면, GPS 신호가 lost 된 지점부터 다시 수신된 거리만큼 속도가 증가한다. 유저가 실내로 들어간 지점만을 추출하기 위해 Δspeed가 사용된다. 위와 같은 상황에서, 속도 임계값 ΔSpeed 보다 낮은 GPS 포인트가 유저가 실내에서 머무른 지점이 되는 것이다. 본 논문에서 제안하는 indoorSP 추출과정은 그림 1과 같다.



(그림 1) indoorSP 추출 과정

GPS 데이터 셋의 하루치 궤적마다 시간 임계값 Δt와 속도 임계값 Δspeed를 계산해야 한다. GPS 데이터 셋의 하루치 궤적을 모집단이라고 하면, 속도 임계값 Δspeed는 모집단의 평균 속도를 지정해준다. 시간 임계값 Δt는 모집단의 시간에 대한 평균과 표준편차를 이용한 정규 누적분포함수의 역함수를 통해 생성된 확률변수 값을 지정한다. GPS 포인트간 시간은 연속형 변수이기 때문에 연속형 정규 누적분포함수의 역함수를 사용한다. 연속형 정규 누적분포함수는 식(1)과 같고 식(1)의 역함수는 식(2)와 같다.

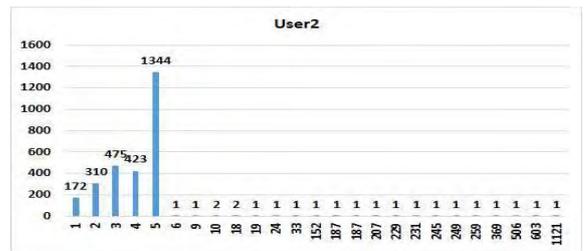
$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \quad (1)$$

$$(e = 2.718, \pi = 3.141)$$

$$F^{-1}(y) = \infty(x : F(x) \geq y), 0 \leq y \leq 1 \quad (2)$$

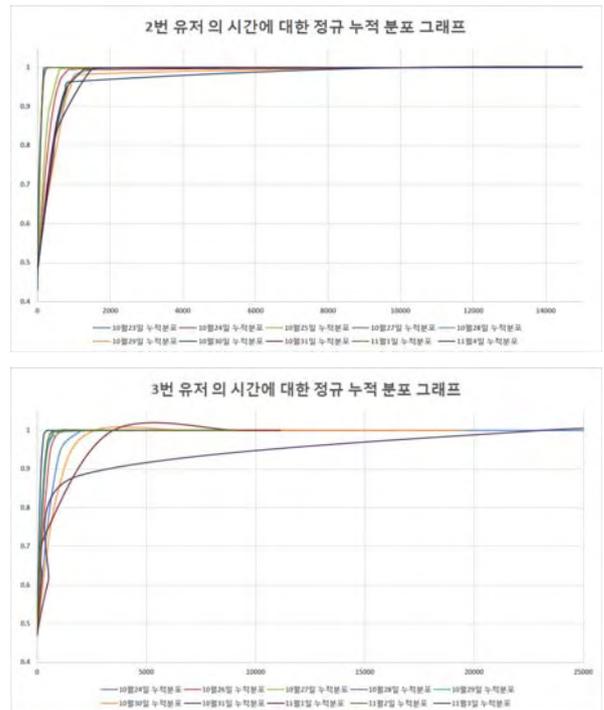
F(x)는 정규 누적분포함수이며 F<sup>-1</sup>(y)는 F(x)의 역함수이다. 역함수에 확률과 평균, 표준편차를 대입하여 확률변수값을 반환받는다. 반환 받은 확률변수값은 시간 임계값 Δt로 지정된다.

본 논문에서는 GeoLife 데이터셋을 사용하여 실험을 진행하였다. GeoLife는 GeoLife 프로젝트(Microsoft Research Asia)에 의해서 182명의 user들의 GPS 데이터를 5년동안(from April 2007 to August 2012) 수집해놓은 GPS 궤적 데이터셋이다[8,9].



(그림 2) 모집단의 시간 분포

그림 2는 GeoLife 데이터셋의 유저 2의 2008년 10월 24일 궤적의 시간 간격이다. x축은 시간이며 y축은 빈도이다. 그림 2의 경우에는 1~5초 사이 간격으로 GPS가 수집되지만 GPS를 수집하는 기기의 설정상태에 따라서 신호를 수집하는 시간 간격이 각각 다르기 때문에 일괄적인 임계값을 지정하면 실내에서 머무른 지점이 아닌 곳이 SP로 지정될 수 있다. 그래서 각 데이터 별로 적절한 시간 임계값을 지정해야만 신뢰성 높은 SP가 추출될 수 있다. 하지만 모집단들의 시간 분포를 확인 후 적절한 임계값을 매번 지정할 수는 없기 때문에, 본 논문에서는 정규 누적분포함수의 역함수를 이용해서 시간 임계값 Δt를 지정한다. 모집단의 시간에 대한 정규 누적분포 그래프를 그려 보면 그림 3과 같다.



(그림 3) 유저별 각 날짜의 시간에 대한 정규 누적분포그래프

그림 3의 x축은 GPS간 시간 간격이고 y축은 누적 확률이다. 그림 2의 2번 유저에서 보는바와 같이 GPS를 수집하는 시간이 1~5초 사이에 이루어지는 것을 알 수 있듯이 1~5초 사이를 초과 하면 누적확률이 거의 대부분 1에 가까워지는 것을 볼 수 있다. 즉, GPS를 수집하는 기기의

설정된 시간을 초과 하면 누적확률이 대부분 1에 가까워지는 것이다. 본 논문에서는 누적 확률 0.9에 대한 역함수를 계산하여 반환된 확률변수값을 시간 임계값  $\Delta iT$ 로 지정한다. 임계값  $\Delta iT$ ,  $\Delta speed$ 를 위와 같은 방법으로 지정한 후, GPS 포인트간 시간과 속도가  $\Delta iT$ 를 초과하고  $\Delta speed$  미만인 GPS 포인트를 indoorSP로 지정한다.

### 3. 실험결과

GeoLife 데이터셋에서 2번 유저의 2008년 10월 24일 데이터를 대상으로 기존 SP 추출과 제안한 방법인 indoorSP를 추출하는 실험을 진행하였다. 기존 SP 추출 방법은 시간 임계값 30초, 거리 임계값 100m로 설정하였다. 제안 방법에서는 모집단의 시간에 대한 누적확률 90%(0.9)의 확률변수값인 434초를 시간 임계값  $\Delta iT$ 로 지정하고, 모집단의 평균 속도 4.96m/s를 속도 임계값  $\Delta speed$ 로 지정하였다. 실험 결과, 기존 SP 추출방법으로 생성된 SP는 36개가 추출되었고, 제안한 방법으로 생성된 SP는 13개가 추출되었다. 즉, 2번 유저는 2008년 10월 24일에 13곳의 실내에 머무른 것이다. 그림 4는 각각의 방법으로 생성된 SP를 구글지도에 매핑해 놓은 것이다.



(그림 4) 제안한 방법과 기존 SP 추출방법 비교

그림 4에서 빨간 점들은 기존 SP 추출방법을 통해 생성된 SP이며, 녹색 마크는 제안한 방법을 통해 생성된 SP이다. 실험 결과 제안한 indoorSP 알고리즘은 사용자실내에 머무른 지점만을 추출해 내는 것을 알 수 있다.

### 4. 결론

본 논문에서는 기존의 SP 추출 알고리즘을 보완할 수 있는 indoorSP를 제안하였다. 기존 SP 추출방식은 야외와 실내를 구분하지 않고 모두 SP로 지정하기 때문에 육안으로 구분하기 쉽지 않다. 이러한 부분을 보완하기 위해, GPS 시간 데이터 분포를 분석하여 임계값  $\Delta iT$ ,  $\Delta speed$ 를 지정한다. 지정된 임계값을 통해 GPS 포인트간 시간과 속도가  $\Delta iT$ 를 초과하고  $\Delta speed$  미만인 지점을 indoorSP로 추출한다. 기존 SP 추출 알고리즘들은 각각의 임계값을 모집단의 특성을 반영하지 않고 사용자가 임의로 설정하여 SP를 추출하기 때문에 신뢰성을 장담할 수 없다. 제안한 indoorSP 추출 알고리즘은 실내에 머무른 지점만을

추출해 낼 수 있고, GPS 시간 데이터 분포의 특성을 반영하여 임계값을 지정해줌으로서 보다 신뢰성 높은 SP를 추출할 수 있다.

### ACKNOWLEDGMENT

본 연구는 2015년도 정부 (교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업(NRF-2009-0093828)와 미래창조과학부 및 정보통신기술진흥센터의 ICT융합 고급인력과정지원사업(IITP-2015-H8601-15-1006)의 연구결과로 수행되었음.

### 참고문헌

- [1] 이향진, 최정화, 박영택, “모바일 사용자의 대용량 GPS 데이터로부터 시맨틱 관심지점 발견” 정보과학회논문지: 소프트웨어 및 응용, 39.3, 175-184, 2012.
- [2] GUIDOTTI, Riccardo, et al. “Retrieving Points of Interest from Human Systematic Movements”, Software Engineering and Formal Methods, Springer International Publishing, p. 294-308, 2014.
- [3] ZIGNANI Matteo and GAITO Sabrina, “Extracting human mobility patterns from GPS-based traces”, Wireless Days (WD), 2010 IFIP. IEEE, p. 1-5, 2010.
- [4] ZHENG, Yu, et al, “Mining interesting locations and travel sequences from GPS trajectories”, Proceedings of the 18th international conference on World wide web, ACM, p. 791-800, 2009.
- [5] XIAO, Xiangye, et al, “Finding similar users using category-based location history”, Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM, p. 442-445, 2010.
- [6] XIU-LI Zhao and WEI-XIANG Xu, “A clustering based approach for discovering interesting places in a single trajectory”, Intelligent Computation Technology and Automation, 2009. ICICTA'09, Second International Conference on. IEEE, p. 429-432, 2009.
- [7] 박진관, 이성로, 정민아, “실내·외 구분 및 위치의 정확성을 개선한 Stay Point 추출 기법”, 전자공학회논문지 제52권 6호, 95-104 (10 pages), 2015.
- [8] ZHENG Yu, XIE Xing and MA Wei-Ying, “GeoLife: A Collaborative Social Networking Service among User, Location and Trajectory”, IEEE Data Eng. Bull, 33.2: 32-39, 2010.
- [9] Geolife GPS trajectories. <http://research.microsoft.com/en-us/downloads/b16d359d-d164-469e-9fd4-daa38f2b2e13/>. Microsoft Research.