

군집 노이즈 필터링 기반 이동 궤적 정제 기법

김소정*, 이석훈**

Clustered Noise Filtering Based Trajectory Refinement Method

Sojeong Kim*, Sukhoon Lee**

이 논문은 2019년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임.
(NRF-2017R1C1B2012065).

요 약

위치 기반 서비스들은 대부분 스마트폰 및 웨어러블 디바이스의 GPS 좌표를 이용한다. 그러나 GPS 좌표들은 센서의 성능 및 오류로 인해 노이즈가 발생하며 이를 제거하기 위하여 다양한 필터들이 연구되었다. 이러한 필터들은 이동 궤적 정제를 위해 사용되지만 여전히 군집된 지역이 존재하는 등 노이즈가 존재한다. 따라서 본 논문에서는 군집 노이즈 필터링에 기반한 이동 궤적 정제 기법을 제안한다. 이를 위하여 사용자의 이동 궤적 중 머무르는 지점을 Sequenced DBSCAN을 이용하여 파악하고 이를 제거하는 군집 노이즈 제거 알고리즘을 제안한다. 또한 본 논문은 기존 필터와 제안 알고리즘을 조합하여 최적화된 필터 적용 순서를 찾는다. 실험을 통하여 100개의 이동 궤적을 12개의 필터로 이루어진 11개의 조합으로 정제한다. 그 결과 제안 기법은 효율적이고 정확도가 높은 최적의 조합으로 분석된다.

Abstract

The location-based services mainly use GPS coordinates for smartphones and wearable devices. However, GPS coordinates generate noise due to sensor performance and errors and various filters have been studied to removed them. These filters are used to refine the trajectory but still require a manual process to extract the clear trajectory. Therefore, this paper proposes trajectory refinement based on clustered noise reduction algorithm. To this end, we propose a cluster noise reduction algorithm that identifies and removes the points of stay of the user's trajectory using a Sequenced DBSCAN. In addition, this paper combines the existing filter and the proposed algorithm to find the optimized filter application order. Through experiments, we refine 100 trajectories with 11 combinations of 12 filters. As the result, the proposed method is efficient and accurate, so it is analyzed to be the best optimized combination.

Keywords

trajectory refinement, noise reduction algorithm, noise filtering, DBSCAN, optimized filtering

* 군산대학교 소프트웨어융합공학과 학사과정
- ORCID: <http://orcid.org/0000-0001-7441-8985>
** 군산대학교 소프트웨어융합공학과(교신저자)
- ORCID: <http://orcid.org/0000-0002-3390-5602>

• Received: Nov. 18, 2019, Revised: Dec. 26, 2019, Accepted: Dec. 29, 2019
• Corresponding Author: Sukhoon Lee
Dept. of Software Convergence Engineering, Kunsan National University,
Korea
Tel.: +82-63-469-8912, Email: leha82@kunsan.ac.kr

1. 서 론

최근 경로 예측 기반에 광고 서비스 또는 경로 추천 서비스 등과 같은 위치 기반 서비스가 제공되고 있다[1]-[2]. 대부분의 경로 예측 기법들은 스마트폰 및 웨어러블 디바이스의 GPS 좌표를 이용한다. 특히 웨어러블 디바이스의 발전으로 인해 사용자의 위치 데이터 수집이 용이해졌다[3]-[5]. 그중 스마트워치와 같은 웨어러블 디바이스는 GPS 좌표, 3축 가속도 등과 같은 사용자의 라이프로그(Lifelog)를 수집할 수 있다[6]-[8].

한편 이러한 라이프로그는 센서의 성능 및 오류로 인해 노이즈를 포함한다. 사용자가 건물 안으로 이동하여 일정 시간을 머무른 후 외부로 나오는 경우에 많은 노이즈가 발생한다[9]-[10]. 특히, 경로 예측과 같은 위치기반 서비스를 위해서는 사용자의 라이프로그로부터 명확한 이동 궤적을 추출하는 것이 필요하다. 이는 경로 학습 시 더 정확한 결과를 도출할 수 있다.

이를 위하여 노이즈 제거를 위한 필터들이 연구되었으며[11]-[13], 이러한 연구들은 이동 궤적을 식별하고 노이즈를 제거하여 사용자의 이동 궤적을 정제한다. 하지만 이러한 필터들을 적용하더라도 여전히 많은 수의 노이즈가 존재한다. 이는 경로 학습의 정확도를 저하시키며, 이를 정제하기 위하여 수작업으로 노이즈를 제거해야 한다.

본 논문에서는 사용자의 이동 궤적을 효과적으로 정제하기 위해 군집 노이즈 필터링에 기반한 이동 궤적 정제 기법을 제안한다. 이를 위하여 사용자의 이동 궤적 중 머무르는 지점을 Sequenced DBSCAN을 이용하여 파악하고 이를 제거하는 군집 노이즈 제거 알고리즘을 제안한다. 또한, 기존 필터와 제안 알고리즘을 조합하여 최적화된 필터 적용 순서를 찾는다. 실험을 통하여 100개의 이동 궤적을 12개의 필터와 11개의 조합으로 적용하고 이를 다시 수작업을 통하여 정제한다. 실험 결과에 따라 최적화된 필터의 적용 순서를 도출한다.

따라서 본 연구에서 제안한 최적화된 이동 정제 기법은 기존에 존재하는 이동 궤적을 포함하는 데이터 셋들에 적용하여 명확한 이동 궤적을 추출할 수 있다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구를 소개하고 각 연구의 문제점을 제시한다. 3장에서는 제안하는 기법을 기술한다. 4장에서는 실험 결과 및 평가를 기술한다. 마지막으로 5장에서는 결론 및 향후 연구 방향을 서술한다.

II. 관련 연구

이 장에서는 이동 궤적 정제를 위한 연구들을 기술한다. 이를 위하여 먼저 GPS 좌표를 수집할 때 발생하는 노이즈 제거에 관련된 연구를 기술한다. 이후 머무는 지점을 식별하는 연구를 기술한다.

2.1 이동 궤적 추출 및 노이즈 필터링 연구

사용자의 경로를 학습하기 위해서는 수집된 GPS 좌표로부터 이동 궤적을 추출하고 노이즈를 제거하는 기법이 요구된다[11]-[13].

3-필터 기반 노이즈 제거 기법은 속도 필터, 각도 필터, 와이파이 필터를 이용하여 이동 궤적으로부터 노이즈를 제거한다[11].

또한 기존의 3-필터에 POI를 추출한 결과를 적용하여 효과적으로 노이즈를 제거한 연구도 진행되었다[12]. 하지만 이러한 연구들은 와이파이 필터를 이용하였으며, 와이파이 접속 정보를 이용한 필터는 본 논문에서 목적으로 하는 웨어러블 디바이스를 활용하는 환경과 맞지 않고, 와이파이에 접속하지 못하는 지역의 경우 이러한 필터를 사용하지 못한다는 문제를 지닌다.

한편, 경로 예측을 위한 CRPM(Continuous Route Pattern Mining) 알고리즘은 사용자의 이동경로 정제를 위하여 중복 필터, 속도 필터, 가속 필터, 총 거리 필터 그리고 각도 필터와 같은 5가지 필터를 적용한다[13].

그러나 이러한 연구들은 다양한 필터들을 적용하여 노이즈를 효과적으로 제거하지만, 머무르는 지점에 대한 노이즈를 제거하지 않는다.

2.2 군집 위치 식별 관련 연구

GPS 좌표로부터 머무르는 위치를 찾아내기 위한 연구들은 주로 군집화 기법을 이용한다.

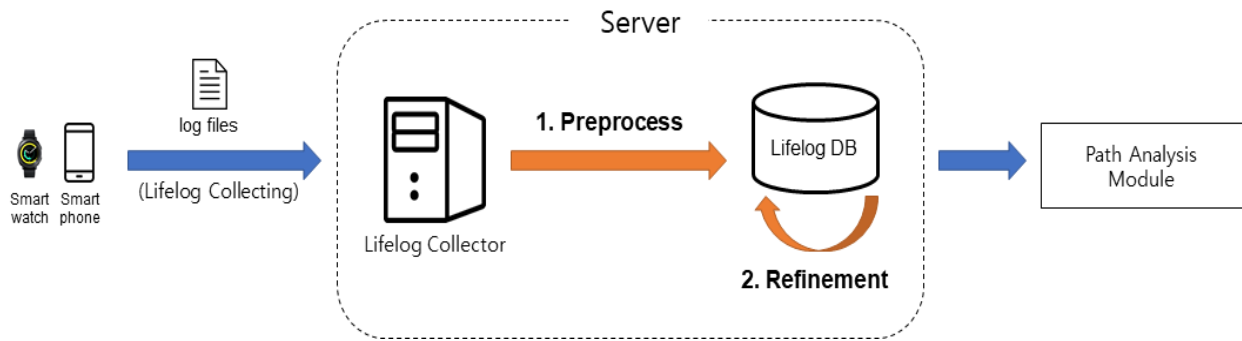


그림 1. 사용자의 이동 경로 분석을 위한 라이프로그 수집 및 전처리 과정
Fig. 1. Lifelog collection and preprocessing for user's path analysis

Ashbrook et al.[14]은 사용자의 이동 후의 위치를 예측을 위하여 K-means 알고리즘을 이용하여 클러스터링 한다. 하지만 이러한 위치 예측은 일정 시간 후 사용자의 위치만을 예측해 줄 뿐 이동 경로까지 알 수는 없다.

Kang et al.[15]은 사용자의 머무르는 지점을 찾아내기 위하여 K-means와 혼합 가우시안을 적용한 기법과 시간 기반 클러스터링을 적용한 기법과 같이 두 가지 클러스터링 방식을 활용한다.

밀도 기반 클러스터링 기법은 DBSCAN을 이용하여 머무는 지점을 파악한다[16]. 사용자의 방문 빈도에 따라 머무는 지점 탐지, 임계범위 설정, 관심 지점 군집화와 같은 3 단계를 통해 비슷한 유형의 이동 궤적들을 클러스터링 한다.

이러한 연구들은 이동 궤적 기반 군집화를 통하여 POI를 추출하고 사용자에게 의미 있는 위치를 찾는다. 하지만 이러한 군집된 위치를 찾는 기법들이 사용자의 이동 궤적을 정제하는데 사용되지는 않는다.

따라서 본 논문은 기존 연구들이 사용하는 이동 궤적 추출 및 노이즈 제거를 위한 필터를 활용하고 사용자의 군집된 위치를 찾아 제거함으로써 사용자가 이동하는 궤적만을 명확하게 추출하는 것을 목표로 한다.

III. 제안 기법

이 장은 군집 노이즈 제거 알고리즘에 기반한 이동 궤적 정제 기법을 제안한다. 이를 위하여 전처리

방법과 기본 필터를 설명하고, 군집 제거 필터를 제안한다. 이를 통해 필터 적용 순서를 기술한다.

3.1 전처리 방법

그림 1은 사용자의 이동 경로 분석을 위한 라이프로그 수집 및 전처리 과정을 보인다. 먼저, 스마트 위치 및 스마트폰을 이용하여 사용자의 GPS 위치 정보를 포함한 라이프로그를 수집한다. 수집된 라이프로그는 스마트폰에서 파일로 저장되며, 해당 로그 파일은 라이프로그 Collector로 전송된다.

서버로 전송된 이후 라이프로그는 두 단계의 전처리 과정을 거친다. 첫 번째 단계는 서버로 전송된 로그 파일을 파싱하여 데이터베이스에 저장한다. 이 과정에서 라이프로그들은 사용자가 이동한 궤적별로 아이디가 부여되어 저장된다. 두 번째 단계는 다양한 필터를 이용하여 명확한 이동 궤적을 추출하기 위해 노이즈를 제거한다.

이후 정제된 라이프로그는 명확한 이동 궤적이 요구되는 경로 학습이나 분석 분야에서 활용될 수 있다.

3.2 기본 필터

이동 궤적의 노이즈를 제거하기 위해 사용된 기본 필터는 시간 필터, 범위 필터, 속도 필터, 가까운 거리 필터, 먼 거리 필터, 각도 필터이다. 표 1은 기본 필터들의 설명이다.

시간 필터는 현재 위치와 이전 위치의 시간 차를

14 군집 노이즈 필터링 기반 이동 궤적 정제 기법

이용하여 새로운 이동 궤적 인지를 구분한다. 만약 이전 위치의 시간이 현재 위치의 시간보다 일정 시간 차이가 있다면 이는 사용자가 새로운 궤도를 따라 이동한다고 판단하여 다른 이동 궤적으로 식별한다. 이 연구에서는 시간 차를 10분으로 정의한다.

범위 필터는 위도와 경도의 일정 범위를 부여하여 해당 지역 안에서의 데이터만 추출한다. 이 연구에서는 한 대학과 그 주변 대학로를 범위로 설정한다.

속도 필터는 현재 위치와 이전 위치의 속도 차를 이용하여 노이즈를 판단한다. 제한 값은 사용자가 자동차를 타는 상황을 고려하여 정의하며, 이 연구에서는 속도 차를 120km/h로 정의한다.

가까운 거리 필터와 먼 거리 필터는 이전 위치와 현재 위치의 거리 차를 이용한다. 가까운 거리 필터는 거리 차가 5m 미만이면 현재 위치가 이전 위치에서 멀지 않은 것을 의미하며 여전히 남아있다고 가정하여 노이즈로 판단한다. 먼 거리 필터는 이전 위치와 현재 위치의 거리 차가 있다면 이는 사용자가 새로운 궤도를 따라 이동한다고 판단하여 다른 이동 궤적으로 식별한다. 이 연구에서는 거리 차를 100m 이상으로 정의한다.

각도 필터는 이전 위치, 현재 위치, 다음 위치를 연결한 선의 각도 차를 측정하여 일정 각도를 넘지 않을 경우 현재 위치를 노이즈로 판단한다. 이 연구는 각도 필터의 각도 차를 45°이하로 정의한다.

표 1. 기본 필터 설명

Table 1. Description of basic filters

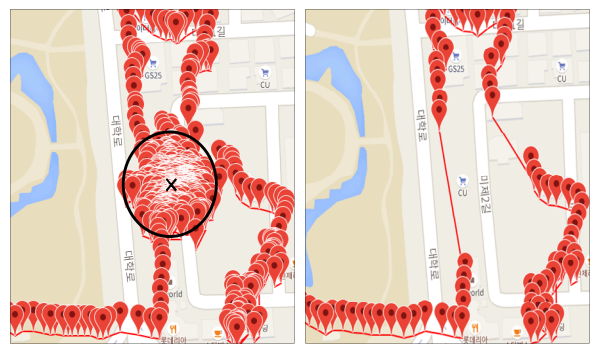
Filter	Description
Time	Create a new trajectory when duration is more than 10 minutes
Range	Delete logs outside of fixed boundary
Speed	Delete logs that velocity is more than 120 km/h
Short distance	Delete logs that distance is less than 5 meters
Long distance	Create new trajectory when distance is more than 100 meters
Angle	Delete logs that three points have an angle less than 45°

3.3 군집 노이즈 제거 기법

기본 필터를 거쳐 노이즈를 제거한다. 하지만 여전히 사용자가 머물러있을 때 발생하는 노이즈들로 인해 군집된 형태의 노이즈가 발생한다. 명확한 이동 궤적을 추출하기 위해서는 군집된 형태의 노이즈 제거가 필요하다.

따라서 본 논문에서는 사용자의 군집된 위치를 파악하기 위해 클러스터링 기법 중 하나인 DBSCAN (Density-based Spatial Clustering of Applications with Noise)[15]을 이용하여 이동 궤적에서 군집된 위치를 파악하고 이를 제거한다. 그림 2는 DBSCAN 기반 필터링의 예를 보인다. 그림 2(a)는 사용자의 이동 궤적 중 하나로 가운데 마킹된 부분이 머무르는 지점으로 판단된다. 그림 2(b)는 DBSCAN을 통하여 군집된 지점으로 파악된 점들을 삭제한 결과이다. 군집된 점들이 삭제됨으로써 이동 궤적이 단순하고 명료해진 것을 알 수 있다.

하지만 DBSCAN은 위치에 대한 정보만을 이용하기 때문에 왕복하는 이동 궤적에 대하여 머무르는 지점을 구분하지 못하고 모두 삭제되는 경우가 발생한다. 즉 그림 2(a)의 경우 마킹된 부분은 위로 올라가는 경로와 내려오는 경로가 모두 군집된 지점으로 식별되어 삭제되기 때문에 실제로 내려오는 경로만 봤을 때 사용자가 머무르는 지점이 없다고 하더라도 모두 삭제해버리는 문제를 지닌다. 따라서 본 논문은 경로가 겹쳐져 모두 삭제되는 문제를 해결하기 위하여 Sequenced DBSCAN을 제안한다.



(a) 적용 전 (b) 적용 후

그림 2. DBSCAN 기반 필터링의 예

Fig. 2. An example of DBSCAN-based filtering,

(a) Before the filtering, (b) After the filtering

Sequenced DBSCAN은 DBSCAN과 다르게 랜덤으로 한 점을 선택 후 선택된 점으로부터 슬라이딩 윈도우를 이용하여 이웃 반경을 설정한다. 또한 설정 파라미터 이웃 반경 내 최소 점 개수를 통해 밀집 영역을 결정한다. 표 2는 제안하는 Sequenced DBSCAN 알고리즘의 pseudo-code를 보인다.

표 2. Sequenced DBSCAN 알고리즘
Table 2. Sequenced DBSCAN algorithm

<pre> input : minPts, r, winSize output : C mark all points as unvisited; make N is empty; do randomly select an unvisited points p; mark p as visited; let L be the sequenced points from p having winSize; let M be the subset of L in r-neighborhood of p; If M has at least MinPts create a new cluster C, and add p to C; add all points of M into N; for each point p' in N If p' is unvisited mark p' as visited; let L be the sequenced points from p' having winSize; let M be the subset of L in r-neighborhood of p'; if M has at least MinPts add all points of M to N; end if if p' is not yet a member of any cluster add p' to C; end if end if end for output C; else mark p as noise; until no point is unvisited; </pre>

제안 알고리즘은 기존의 DBSCAN 알고리즘[16]에서 슬라이딩 윈도우를 적용하는 방식으로 개선한다. 입력 값은 설정 파라미터 최소 점 개수(minPts), 이웃 반경(r), 슬라이딩 윈도우 크기(winSize)이며, 알고리즘을 통해 생성된 클러스터(C)를 반환한다. 반환된 클러스터는 머무는 지점이라고 파악되므로 해당 클러스터들을 제거한다. 이 연구에서는 minPts를 15로 설정하고 이웃 반경을 20m로 설정하였다.

그림 3은 Sequenced DBSCAN 기반 필터링의 예를 보인다. 그림 3(a)는 Sequenced DBSCAN을 통하여 군집된 지점으로 파악된 점들을 삭제한 결과이다. 군집된 점들을 삭제할 때, 일정 반경 내의 모든 점들이 삭제되므로 이동 궤적에서 해당 반경 내의 결측값이 발생한다. 만약 군집된 지점의 반경이 넓다면 삭제되는 지점의 양이 많아지므로 사용자의 이동 궤적은 그만큼 끊기는 지점이 넓어진다.

따라서 본 논문은 이러한 결측값을 대체하기 위해 두 가지 방법을 이용한다. 첫 번째 결측값 대체 방법은 군집된 점들을 삭제할 때 군집 지점으로 안으로 들어간 직후의 점들과 군집 지점 밖으로 나가기 직전의 점들을 남겨둠으로 반경의 공백을 메운다. 이때 남겨두는 점들의 수는 사용자의 평균 이동 속도를 군집의 반경으로 나누는 것으로 구한다. 두 번째 결측값 대체 방법으로는 군집된 점들을 삭제할 때 일정 시간마다 점 한 개씩을 남겨 군집된 지점을 채운다. 그림 3(b)는 결측값을 대체한 결과를 보인다.



그림 3. Sequenced DBSCAN 기반 필터링의 예 (a) 결측값 발생; (b) 결측값 대체 결과

Fig. 3. An example of sequenced DBSCAN-based filtering, (a) Occurrence of missing value, (b) Imputation result

3.4 이동 궤적 정제 프로세스

명확한 이동 궤적의 정제를 위해서는 기존 필터 및 제안 필터의 적용 순서를 최적화해야 할 필요가 있다. 그림 4는 본 논문에서 제안하는 이동 궤적 정제 프로세스를 보인다. 제안 프로세스는 최적화된 필터 적용 순서를 의미한다.

필터 적용 순서는 1~3번을 통하여 우선 수집된 위치 정보들의 나열로부터 이동 궤적들을 구분한다. 이는 범위 필터로 정해진 범주 이외의 점들을 제거한 후, 시간 필터와 먼 거리 필터를 적용하여 현재 위치가 이전 위치와 시간적, 거리적 차이가 있을 경우 새로운 이동 궤적을 생성하도록 한다. 이후, 4~7 번을 통해 모든 이동 궤적들에 대하여 속도 필터, 짧은 거리 필터, 첫 번째 결측값 대체를 적용한 Sequenced DBSCAN, 각도 필터를 순서대로 적용한다.

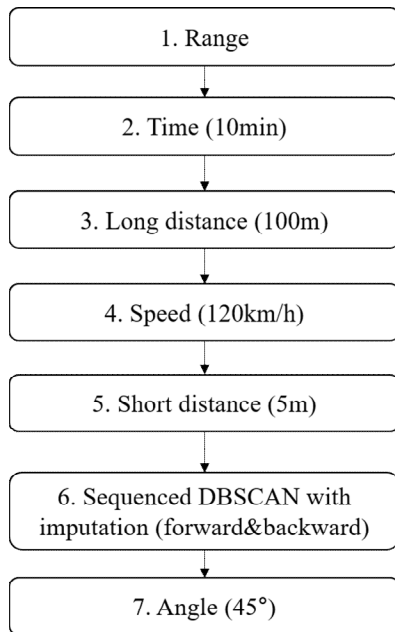


그림 4. 이동 궤적 정제 프로세스
Fig. 4. Trajectory refinement process

IV. 실험 및 평가

이 장에서는 제안하는 이동 궤적 정제 프로세스를 평가한다. 먼저 다양한 필터들의 조합과 제안하는 필터 적용 순서를 비교하기 위하여 실험 방법을 기술하고, 실험 결과의 분석을 통해 제안 기법의 우수성을 평가한다.

4.1 실험 방법

본 논문은 제안 프로세스의 우수성을 검증하기 위하여 기본 필터 및 제안 필터의 조합들 중 최적의 필터 조합을 찾는다. 비교를 위하여 기본 필터 및 제안 필터들의 조합들을 데이터셋에 적용하여 정제한 후, 수작업 정제를 통해 각 필터 조합의 효율성을 비교 평가한다.

이 실험에 사용하는 데이터셋 DSEM-Traj2018은 총 39명을 대상으로 170일 동안 GPS, 심박수, 3축 가속도 데이터를 수집한 라이프로그 데이터셋이다. 이 중 실험을 위하여 100개의 이동 궤적을 추출하여 사용한다.

이 실험에서 사용하는 필터는 총 12개로 그 종류는 표 3과 같다. ①번은 기본 필터 중 범위, 시간, 긴거리 필터를 순서대로 적용시킨 방법(3filter)이다. ②번은 기본 필터 중 범위, 시간, 긴거리, 속도, 짧은거리 필터를 순서대로 적용시킨 방법(5filter)이다. ③번은 앵글 필터, ④번은 DBSCAN, ⑦번은 Sequenced DBSCAN (Seq_DBSCAN)을 의미한다. ④번과 ⑦번은 각각 DBSCAN을 수행하면서 삭제되는 부분의 결측값이 발생하는데 ⑤ DBSCAN_f&b, ⑥ DBSCAN_30s, ⑧ Seq_DBSCAN_f&b, ⑨ DBSCAN_30s는 각각 3.3절에서 설명한 결측값 대체 기법을 DBSCAN과 Sequenced DBSCAN에 적용시킨 필터이다.

표 3. 필터의 종류
Table 3. Types of filter

Symbol	Filter	Description
①	3filters	range, time, long distance(100m)
②	5filters	range, time, long distance(100m), speed, short distance(5m)
③	Angle	angle
④	DBSCAN	DBSCAN
⑤	DBSCAN_f&b	DBSCAN with forward and backward imputation
⑥	DBSCAN_30s	DBSCAN with 30sec imputation
⑦	Seq_DBSCAN	Sequenced DBSCAN
⑧	Seq_DBSCAN_f&b	Sequenced DBSCAN with forward and backward imputation
⑨	Seq_DBSCAN_30s	Sequenced DBSCAN with 30sec imputation

표 4. 실험을 위한 필터 조합

Table 4. Combinations of filters for experiment

No	Combination	Description
1	①	3filters
2	②,③	5filters + angle
3	②,③,④	5filters + angle + DBSCAN
4	②,③,⑤	5filters + angle + DBSCAN_f&b
5	②,⑤,③	5filters + DBSCAN_f&b + angle
6*	②,⑧,③	5filters + Seq_DBSCAN_f&b + angle
7	②,⑦,③	5filters + Seq_DBSCAN + angle
8	②,⑥,③	5filters + DBSCAN_30s + angle
9	②,⑨,③	5filters + Seq_DBSCAN_30s + angle
10	②,③,⑧	5filters + angle + Seq_DBSCAN_f&b
11	②,④,③	5filters + DBSCAN + angle

본 논문에서 제안하는 필터의 적용 순서는 3.4절에서 제안하는 바와 같이 ②, ⑧, ③이다. 이 실험은 제안 필터인 첫 번째 결측값 대체 방법을 사용한 Sequenced DBSCAN (Seq_DBSCAN_f&b)의 우수성 검증과 제안 필터 적용 순서가 최적화된 프로세스임을 검증하기 위하여 다양한 필터들의 조합을 이용하여 비교 평가를 수행한다. 표 4는 이러한 필터들의 조합을 보이며, 본 논문은 제안하는 필터들의 조합은 6번과 같다.

이 실험은 11개의 필터 조합의 효율성을 평가하기 위하여 수작업 정제를 통한 각각의 데이터셋을 만드는 것으로 이루어진다. 수작업 정제는 사람이 수집된 이동 궤적들을 직접 보며 머무르는 지점이나 노이즈를 일일이 제거하는 방법을 의미한다. 따라서 이 실험은 수작업 정제를 거친 데이터셋을 일종의 정답셋으로 설정한다. 만약 어떤 데이터셋을 수작업으로 정제할 때 많은 작업양이 발생한다면, 그 데이터셋을 정답셋으로 만들기 위한 노력은 수작업의 작업양만큼 요구된다고 볼 수 있다. 또한, 동일한 데이터셋에 두 필터 조합을 각각 적용한다고 할 때, 필터를 거친 결과로 생성된 두 데이터셋을 다시 수작업 정제를 거칠 경우, 각각의 작업양은 해당 필터 조합이 얼마나 효율적으로 노이즈를 제거했는지를 판단할 수 있는 기준이 될 수 있다.

수작업 정제 시 작업 양의 측정을 위하여 자체 개발한 이동 궤적 정제 도구를 이용한다.

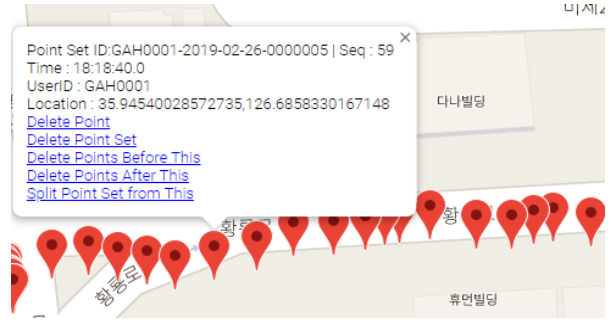


그림 5. 이동 궤적 정제 도구의 스크린샷

Fig. 5. A screenshot of a trajectory refinement tool

그림 5는 이동 궤적 정제 도구의 스크린샷을 보이며, 이 도구는 정제를 위한 행동을 할 때 발생하는 이벤트를 로그로 기록한다.

로그의 종류로는 delete_point, delete_before, delete_after, delete_set, split의 총 5가지로 구분된다. delete_point는 선택된 점을 지우는 행위로 하나의 점만 지워진다. delete_before는 선택된 점부터 이전 점들을 전부 지우는 행위, delete_after는 선택된 점부터 이후 점들을 전부 지우는 행위이며, delete_set은 현재의 이동 궤적을 전부 지우는 행위이다. split 경우는 현재 점을 기준으로 이전 점들과 이후 점들을 다른 이동 궤적으로 구분한다. 이러한 도구를 이용하여 정제 시 발생한 로그의 수와 로그 종류별로 지워진 점들의 개수를 집계한다. 또한, 정제 후 남은 이동 궤적 수도 집계한다.

4.2 실험 결과 및 평가

표 5는 27,266개의 점으로 이루어진 100개의 이동 궤적에 11개의 필터 조합을 적용시킨 후, 생성된 11개의 데이터셋을 이동 궤적 정제 도구를 이용하여 수작업 정제를 수행하였을 때의 로그 수, 총 삭제된 레코드 수, 로그 당 선택된 레코드 수, 로그 종류를 나타낸다. 7번 조합의 경우 로그 수가 179개, 총 삭제된 레코드 수 475개로 전체 필터 조합들 중 로그가 가장 적게 발생하였다. 또한, 6번, 11번 조합도 각각 로그 수가 221개, 220개이며 삭제된 레코드 수도 613, 645개로 다른 조합들에 비해 비교적 적은 수의 수작업 정제가 이루어진 것으로 확인되었다. 이는 7, 6, 11번의 조합들이 그만큼 더 효율성이 높다고 볼 수 있다.

18 군집 노이즈 필터링 기반 이동 궤적 정제 기법

표 5. 로그 종류별 삭제 된 레코드 수

Table 5. Number of deleted records by log

No.	Log	Total number of deleted records	Number of records selected per log	del_point log	del_before log	del_after log	delset_id log	split log
1	1080	18724	17.3	814	83	146	5	32
2	626	2851	4.6	459	48	100	1	18
3	247	875	3.5	143	31	50	2	21
4	303	1044	3.4	189	38	58	0	18
5	250	711	2.8	151	33	42	2	22
6	221	613	2.8	136	30	35	0	20
7	179	475	2.4	92	33	29	2	23
8	273	956	3.5	147	35	71	2	18
9	271	855	3.2	154	36	58	5	18
10	248	987	4.0	156	34	40	2	16
11	220	645	2.9	121	39	37	3	20

표 6. 정제 후 남은 개수

Table 6. Remaining number after refinement

No.	Number of points after refinement	Number of trajectory after refinement	Number of points per trajectory
1	8533	125	68.3
2	5776	117	49.4
3	5295	119	44.5
4	5659	118	48.0
5	5509	118	46.7
6	5556	119	46.7
7	5104	119	42.9
8	5172	116	44.6
9	5095	113	45.1
10	5549	114	48.7
11	5124	117	44.0

표 6은 수작업 정제 후 남은 점 개수, 남은 궤적 수, 궤적 당 점 개수를 보인다. 앞서 높은 효율을 보였던 6번, 7번, 11번 조합을 보았을 때 각각 5556개, 5104개, 5124개의 점이 남았고, 하나의 궤적 당 각각 46.689개, 42.891개, 43.795개의 점이 남은 것으로 측정되었다. 수작업 정제 후 남은 점의 개수는 최종적으로 경로 학습에 이용될 데이터의 양을 의미하며, 남은 점의 개수가 적을수록 해당 필터들의 조합 및 수작업 정제가 더 많은 데이터를 삭제했다는 것을 의미한다.

적용된 필터 조합의 우수성을 평가하기 위해서는 단순히 수작업의 로그 수와 함께 수작업 정제 이후

남아있는 점들의 개수도 함께 확인해 보아야 한다. 만약 어떤 필터 조합이 노이즈와 관계없이 점들을 삭제한다고 한다면, 많이 삭제할수록 노이즈가 삭제되는 양도 증가할 것이다. 이는 수작업 정제 시 노이즈의 삭제 양이 많으므로 효율적인 필터 조합으로 보일 수 있다. 하지만 노이즈가 아닌 명백한 이동 궤적에 포함되는 점들도 함께 삭제되기 때문에, 필터 조합의 평가는 수작업 정제 후 최종적으로 남아있는 데이터의 양도 함께 고려해야 한다.

표 7은 실험 결과를 토대로 기존 필터와 제안하는 필터간의 정성 평가를 보인다. 비교를 위하여 2, 11, 7, 6번 필터 조합을 선정하였다. 이 필터 조합들은 동일한 기본 필터 조합의 5번째 순서에 DBSCAN, Seq_DBSCAN, Seq_DBSCAN_f&b 기법을 각각 삽입한 조합들이다. 이를 통하여 기존의 기본 필터, DBSCAN 기법과 제안 알고리즘인 Seq_DBSCAN, Seq_DBSCAN_f&b 기법들의 상대적인 성능 차이를 명확히 확인할 수 있다.

표 7. 기존 필터와 제안 필터들 간 정성 평가

Table 7. Qualitative evaluation among basic filters and proposed filter

Combination filter	efficiency	precision
(2) basic filters	low	high
(11) DBSCAN	high	low
(7) Seq_DBSCAN	very high	low
(6) Seq_DBSCAN_f&b	high	high

평가 결과 기본 필터들만을 이용할 경우 수작업 정제 시 다른 기법들에 비해 여전히 많은 노이즈가 존재하여 필터의 효율성이 떨어지는 것을 볼 수 있다. 또한, DBSCAN과 Seq_DBSCAN은 각각 노이즈를 제거하는 수가 많아 높은 효율성을 보이며, 특히 Seq_DBSCAN은 가장 적은 로그 수를 지니므로 매우 높은 효율성을 보인다.

하지만, 노이즈 제거의 정확성으로 보았을 때는 Seq_DBSCAN은 다른 필터들에 비해 노이즈가 아닌 점들을 덜 지우는 것을 알 수 있다. 이는 반대로 DBSCAN이나 Seq_DBSCAN이 군집된 점들의 제거 시 필요 이상으로 지운다는 것을 의미하며, 결국 값 대체 기법이 효과적으로 작용한다고 볼 수 있다.

결론적으로, 제안 기법인 Seq_DBSCAN을 적용한 6번 필터 조합이 다른 필터들에 비해 상대적으로 필터의 효율성과 노이즈 제거의 정확도가 높아 최적의 성능을 지님을 알 수 있다.

V. 결 론

본 논문은 명확한 이동 궤적을 추출하기 위해 군집된 지점을 식별하고 이를 제거하는 군집 노이즈 제거 알고리즘인 Sequenced DBSCAN을 제안하였다. 또한 기존 필터와 제안 알고리즘을 조합하여 최적화된 필터 적용 순서를 제안하였다.

평가를 위하여 12개의 필터로 이루어진 11개의 조합을 100개의 이동 궤적에 적용시키고 정제된 결과를 다시 한 번 수작업으로 정제 하는 실험을 진행하였다. 그 결과 제안 프로세스는 다른 필터 조합들에 비해 필터의 효율성과 노이즈 제거의 정확도가 우수함으로 평가됨에 따라 가장 최적화된 조합으로 평가되었다.

향후에는 모바일 혹은 웨어러블 디바이스를 사용하여 사용자의 이동 궤적과 함께 측정될 수 있는 걸음 수, 활동량, 심박수 등과 같은 라이프로그를 수집하고 데이터 정제 시 각 요소들을 연계할 수 있는 방법에 대한 연구를 진행할 예정이다.

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] S. J. Kim and C. Y. Jung, "A Study on Lost Child Prevention Service Using LBS and Map Information", The Journal of Institute of Internet, Broadcasting and Communication, Vol. 17, No. 6, pp. 181-186, Dec. 2017.
- [3] S. Waltzer, "Global Smartwatch Revenue and ASP by Vendor by Price Tier : Q2 2017", Report, Strategy Analytics, Aug. 2017.
- [4] R. Rawassizadeh, M. Tomitsch, K. Wac, and A. M. Tjoa, "UbiqLog: A generic mobile phone-based lifelog framework", Personal and Ubiquitous Computing, Vol. 17, No. 4, pp. 621-637, Apr. 2012.
- [5] B. Jeong, "Current Status and Prospects of the Wearable Devices Market", Telecommunications Broadcasting Policy Periodicals, Vol. 30, No 20, pp. 1-7, Nov. 2018.
- [6] S. Nam, S. J. Kim, and S. Lee, "DSEM-Traj2018: Building Lifelog Dataset using Smartwatch for Path Prediction", World IT Congress, Feb. 2019.
- [7] J. Ji, S. Kwon, M. G. kim, and S. Lee, "Activity Classification Method for User Movement using Wearable Device", in Proc. the KIIT Fall Conference 2017, Daejeon, South Korea, pp. 318-319, Dec. 2017.
- [8] H. J. Sim, "BigData and Life-logging: Current Status and Outlook", http://www.kisdi.re.kr/kisdi/fp/kr/board/selectSingleBoard.do?cmd=selectSingleBoard&boardId=GPK_ICT_TREND&seq=29574&reStep=3999 &ctx=2014, Mar. 2014. [accessed: May 14, 2019]
- [9] 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.

- [10] H. Do and J. Oh, "A Study on smartphone indoor navigation technology using Extended Kalman filter", The Journal of Institute of Internet, Broadcasting and Communication, Vol. 18, No. 6, pp. 1-9, Dec. 2018.
- [11] S. Yang, H. Baek, J. Kim, M. Jeon, and Y. Park, "GPS Coordinates based on Trajectory Extraction Approaches for the Learning Personal Paths", Korea Information Processing Society 2011 fall Conference, Vol. 38, No. 2, pp. 1-4, Nov. 2011.
- [12] J. M. Kim and Y. T. Park, "GPS-based Noise Reduction Approach for Trajectory Recognition", Journal of KIISE : Software and Applications, Vol. 39, No. 4, pp.328-337, Apr. 2012.
- [13] L. Chen, M. Lv, Q. Ye, G. Chen, and J. Woodward, "A personal route prediction system based on trajectory data mining", Information Sciences Vol. 181-187, pp. 1264-1284, Apr. 2011.
- [14] D. Ashbrook and T. Straner, "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.
- [15] J. Kang, W. Welbourne, B. Stewart, and G. Borriello, "Extracting Places from Traces of Locations", ACM Mobile Computing and Communication Review, Vol. 9, No. 3, pp. 58-68, Jul. 2005.
- [16] J. M. Kim and Y. T. Park, "POI Detection and Route Identification for Building Route Models for Smartphone Users", Journal of KIISE : Software and Applications, Vol. 40, No. 12, pp. 799-808, Dec. 2013.
- [17] Jiawei Han, Micheline Kamber, and Jian Pei, "Data Mining (3rd Ed.) - Concepts and Techniques", acornpub, pp. 603-607, Apr. 2015.

저자소개

김 소 정 (Sojeong Kim)



2016년 3월 ~ 현재 : 군산대학교
소프트웨어융합공학과 학부생
관심분야 : 데이터 분석, 빅데이터

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
조교수
관심분야 : 사물인터넷, 메타데이터, 센서 레지스트리,
시맨틱 웹, 경로 예측