

대용량 GPS 궤적 데이터를 위한 효율적인 클러스터링

김태용^o 박보국 박진관 조환규
부산대학교 전기전자컴퓨터공학과
{ktyong22, limited10, jinkwan, hgcho}@pusan.ac.kr

An Efficient Clustering Algorithm for Massive GPS Trajectory Data

Taeyong Kim^o, Bokuk Park, Jinkwan Park, Hwan-Gue Cho
Dept. of Electrical and Computer Engineering, Pusan National University

요 약

최근 스마트 폰 및 네비게이션의 보급의 발달로 인해 쉽게 GPS 데이터를 얻을 수 있으며, 이러한 GPS 데이터를 이용하여 많은 연구들이 수행 되고 있다. 그 중 하나는 자동으로 도로를 생성해주는 도로 생성 시스템이다. 이 시스템에서 핵심이 되는 작업은 대표궤적을 추출하는 작업이다. 대표궤적 추출을 수행 할 때에는 시작과 끝이 비슷한 궤적 데이터들의 집합을 전제로 하여 궤적을 추출 한다. 하지만 실제 대용량의 GPS 데이터 집합들은 임의의 형태로 아주 다양하게 존재한다. 따라서 대표궤적을 추출하기에 앞서 전 처리과정으로 궤적 클러스터링 작업이 필요하다. 본 논문에서는 이러한 문제를 해결하기 위해 하나의 영역을 일정한 격자로 분할하고, 이산 프레셰 거리 평균을 이용함으로써, 임의로 형성된 궤적 데이터들을 클러스터링 방법을 제시한다.

1. 연구동기

최근 GPS수신 단말기 보급(스마트 폰, 네비게이션)의 증가에 따라 GPS 데이터를 쉽게 얻을 수 있고, 이에 따라 GPS 데이터를 이용한 연구가 활발히 진행되고 있다 [1,2]. 그 중 하나는 도로 생성 시스템으로 이 시스템은 사용자로부터 받은 GPS 데이터를 이용하여 해당 도로를 자동으로 찾아주는 시스템이다[3,4].

아래 그림-1은 강남지역을 대상으로 택시 GPS 데이터를 받아 지도에 나타낸 그림이다. 빨간색 곡선들은 실제 택시들이 지나간 GPS 데이터이며, 택시 궤적은 총 3000 개이다. 이러한 GPS 데이터들을 이용하여 자동으로 도로를 찾아주면 현재 사용되고 있는 인공위성 촬영이나, 현장실사의 비용과 노력을 줄여 줄 수 있으므로, 기대하는 효과가 아주 크다.

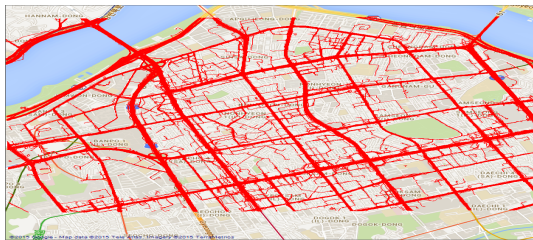


그림-1. 강남구 지역에서 수집된 택시 GPS 데이터. 빨간색 선은 총 3000대의 택시의 궤적을 나타내고 있다.

이 시스템에서 가장 중요한 작업은 유사한 궤적간의 대표 궤적을 추출하는 것이다. 이미 대표궤적을 추출하는 연구는 많이 진행되어 왔다[5,6]. 이 연구는 여러 개의 궤적데이터들이 유사하다고 가정하고 시작한다. 유사한

궤적 데이터들이란 궤적의 시작 지점과 끝 지점이 거의 비슷한 데이터들을 말한다.

아래 그림-2는 유사한 궤적에서 도로 중심선을 추출한 그림이다. 그림-3은 그림-1의 올림픽대로를 지나는 궤적들 중 시작과 끝이 유사한 궤적들만 모아 지도에 나타내었다.

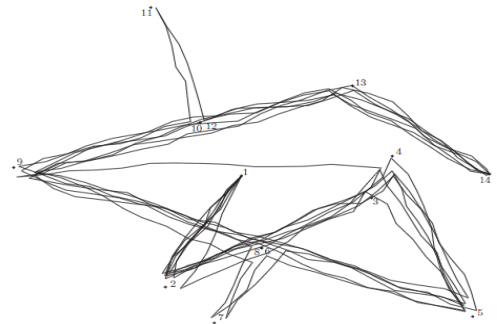


그림-2. 유사한 궤적에서 도로 중심선 추출하기[6].

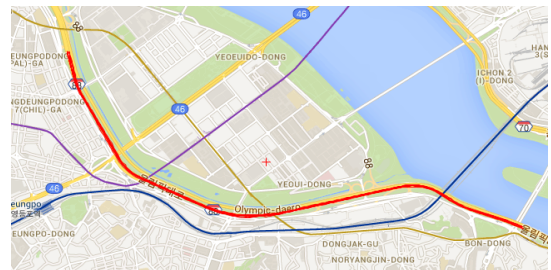


그림-3. 이상적인 궤적 GPS 데이터. 빨간색 선은 궤적들의 모음을 나타낸 것이며, 궤적들의 시작과 끝은 거의 유사하다.