

이산 Fréchet 거리 척도를 이용한 궤적 유사도 고속계산 휴리스틱

Fast Heuristics for Similarity of Trajectories Using Discrete Fréchet Distance Measure

요약

최근에 들어서는 GPS 수신기를 이용하여 어떤 물체의 위치 정보를 측정하는 것이 그렇게 드문 일이 아니다. 이렇게 측정된 위치 정보를 이어 다양한 궤적 정보를 얻을 수 있다. 그 중에서 차량의 궤적 정보는 실제 차량이 다닐 수 있는 경로를 나타내기 때문에 곧 전자 지도 생성 연구에 활발히 이용된다. 흔히 두 궤적 사이의 유사도를 계산할 때는 프레셰 거리(Fréchet distance)를 이용하는데, 궤적의 점 개수가 많아질수록 프레셰 거리를 이용하는 것은 연산 시간 상 어려움이 따른다. 본 논문에서는 이산 프레셰 거리를 통한 궤적 유사도 판별 시, 유사 궤적의 특성을 이용하여 향상된 판별법을 제안한다. 또한, 기존의 이산 프레셰 거리를 이용했을 때의 유사도 계산 결과와 비교하여 어느 정도의 활용 가능성을 갖는지 실험을 통해 알아본다.

1. 연구 동기

최근 스마트 기기를 비롯한 다양한 전자장치가 일상생활에서 널리 쓰이고 있다. GPS를 이용한 위치 정보 수집도 흔히 이용되는 서비스이다. 수집된 위치 정보를 순서대로 연결하면 점의 서열로 이루어진 다중곡선 궤적이 생성되는데, 이 궤적을 통해 전자 지도를 수정하거나 생성해내는 연구도 활발하다[1]. 경로가 유사한 궤적들을 모아 대표 궤적을 찾는 것인데, 이 문제는 두 곡선 사이의 유사성을 찾는 문제로 볼 수 있다.

곡선 간의 비교는 데이터의 유사관계뿐만 아니라 규칙성을 찾는 데도 활용될 수 있어 로봇공학[2], 생물정보학[3] 등 여러 분야에서 관련 연구가 진행되어오고 있다. 흔히 사용되는 곡선 유사도를 측정하는 방법으로 프레셰 거리가 있다[4]. 프레셰 거리는 두 곡선 사이의 거리를 나타낸 값인데, 이 척도는 궤적 사이의 유사도를 측정하는 경우에 활용할 수 있으나 높은 계산 복잡도 때문에 점이 많은 궤적 탐색에는 부적절하다는 문제가 있다.

본 논문에서는 두 궤적 사이의 유사도를 측정하는 방법인 프레셰 거리를 소개하고, 이를 개선하여 계산 복잡도를 낮춘 유사 궤적 탐색 방법을 소개한다. 또한 이 방법의 효용성을 실험을 통해 확인해본다.

2. 기존 방법

다중선분(polyline)은 여러 선분이 연결된 곡선이다. 다중선분 T 에 대해서 $|T|$ 는 이 다중선분을 구성하는 점의 개수를 나타낸다. 프레셰 거리는 이러한 두 다중곡선 사이를 이을 때 필요한 최소한의 거리를 통해 두 다중곡선의 유사도를 나타낸다. 프레셰 거리에는 연속적인 방법과 이산적인 방법이 있는데, 연속적인 방법은 그림 1에서 보이는 것과 같다. 한 다중선분 위의 각 점에서 다른 다중선분으로 갈 수 있는 모든 거리 중 최단거리를 측정한다. 측정된 최단거리들 중에서 가장 작은 값이 프레셰 거리가 된다[5]. 그림에서는 붉은 선분의 길이가 프레셰 거리가 된다. 프레셰 거리는 두 다중선분 사이에서 생길 수 있는 모든 거리를 살펴보기 때문에 정확한 값을 나타내며, $|T_p| = m$, $|T_q| = n$ 인 두 다중선분 T_p , T_q

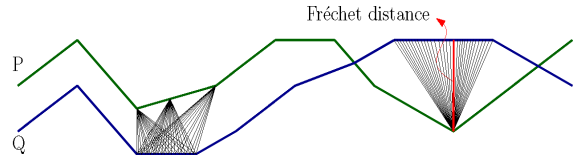


그림 1. 두 다중선분 P, Q에 대해 프레셰 거리가 측정되는 과정 및 선택된 프레셰 거리. 한 다중선분 위의 각 점에서 다른 다중선분으로 갈 수 있는 모든 거리 중 최단거리를 측정한다. 측정된 최단거리 중 가장 긴 거리가 프레셰 거리로 선택된다. 붉게 표시된 선분이 가장 길기 때문에 프레셰 거리가 된다.

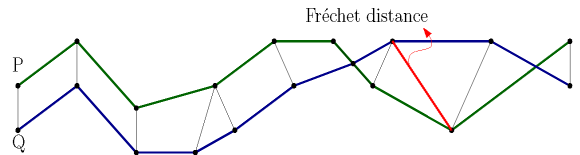


그림 2. 두 다중선분 P, Q에 대해 이산 프레셰 거리가 측정되는 과정 및 선택된 프레셰 거리. 한 다중선분 위의 각 점에서 다른 다중선분 위의 각 점까지의 거리 중 최단거리를 측정한다. 측정된 최단거리 중 가장 긴 거리가 프레셰 거리로 선택된다. 붉게 표시된 선분이 가장 길기 때문에 프레셰 거리가 된다.

에 대해서 계산 복잡도가 $O(mn \cdot \log mn)$ 이다. 때문에 아주 길거나 매우 조밀하게 샘플링 된 다중선분에 대해서는 시공간적으로 계산에 어려움이 생긴다.

이산 프레셰 거리는 프레셰 거리의 복잡한 계산을 줄여, T_p , T_q 에 대해서 $O(mn)$ 에 계산이 가능한 방법이다[6]. 그림 2에서 보이는 것처럼, 두 다중선분 사이의 모든 거리를 측정하지 않고, 다중선분이 갖는 각 점에서의 거리만 측정한다. 두 다중선분의 점을 하나씩 선택하여 쌍을 만들었을 때, 각 다중선분의 마지막 점까지 이 쌍의 길이가 최소가 되도록 유지하면서 다음 점으로 옮긴다. 탐색 과정에서 가장 길었던 쌍의 길이가 이산 프레셰 거리가 된다. 이산 프레셰 거리는 동적 프로그래밍을 이용하여 간단히 계산할 수 있다. 계산 의사 코드는 알고리즘 1과 같다[7]. 이산 프레셰 거리는 기존의 프레셰 거리보다는 계산 복잡도를 낮추었으나, 점이 많은 다중선분 또는 많은 다중선분들 사이의 유사도 계산에 적용하