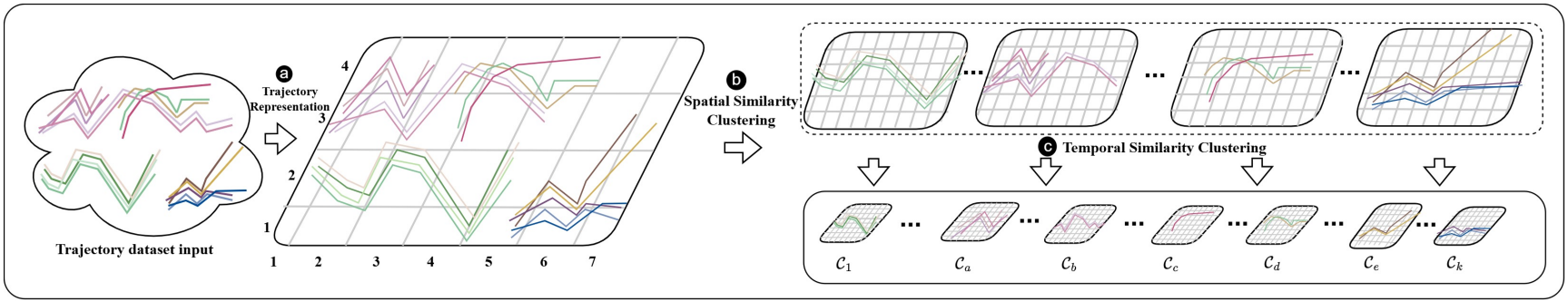
Large-Scale Trajectory Clustering with LSH

Cheng TIAN, Ziqiang YU, Xiaohui YU, Yang LIU, Wei LIU, Chaofan WANG

Frontiers of Computer Science, DOI: 10.1007/s11704-025-50766-6

Problems & Ideas

- Problems of conventional trajectory clustering approaches:
  - The high computational cost of calculating pairwise distances between numerous trajectory points.
  - Unnecessary exhaustive similarity computations between trajectories that do not belong to the same cluster.
- **Ideas:** Encoding trajectories as sequences of uniform grid cells, then performs two-phase Locality-Sensitive Hashing-based clustering by jointly considering spatial similarity (shared cells) and temporal similarity (traversal order), achieving linear-time complexity.



Since conventional LSH-based clustering fails to account for the temporal similarity of trajectories passing through common grid cells, we design an encoding scheme based on the traversal order of shared cells. By applying LSH to the encoded trajectories, our method enables comprehensive consideration of both spatial and temporal similarity in trajectory clustering.

Main Contributions

- Contributions:
  - We propose a Locality-sensitive hashing-based Trajectory Clustering (LTC) algorithm that achieves efficient large-scale trajectory processing without compromising accuracy;
  - Our trajectory encoding scheme represents movements as uniform grid cell sets, preserving spatiotemporal features. This reduces clustering to Jaccard-based set similarity computation, accelerated via locality-sensitive hashing;
  - Extensive experiments on four real-world datasets show that HPTC is more than twice as fast as most baselines.

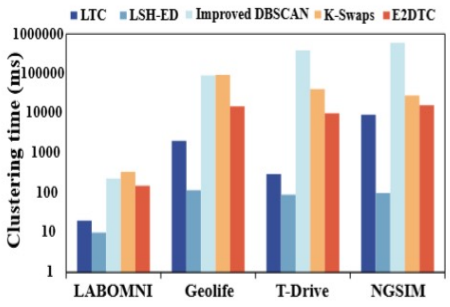


Fig.1

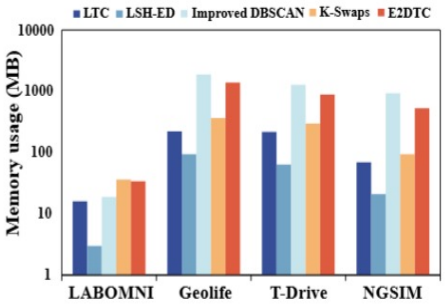


Fig.2

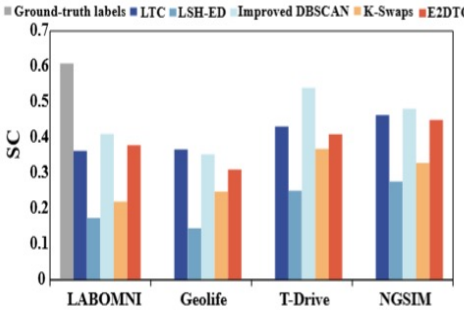


Fig.3

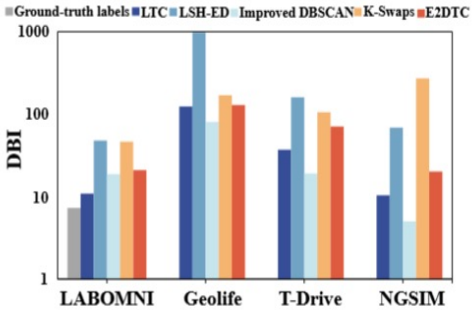


Fig.4

Comparative results of clustering efficiency and accuracy across four real-world trajectory datasets. Fig.1: Computation time comparison; Fig.2: Memory consumption comparison; Figs.3-4: Accuracy comparison under two evaluation metrics.