Auto-Param DBSCAN & Human Activity Range Inference from Map Data

Abstract

An improved DBSCAN pipeline that selects parameters automatically (no manual /minPts) and applies clustering on map anchored location data to infer a person's activity range.

Problem

Standard DBSCAN requires hand tuned and minPts; values shift across densities and sampling rates.

Method

Auto-Parameter Selection

estimation via k distance curve elbow detected with a robust segmented regression; fall back to quantile of kNN distances when elbow is ambiguous.

minPts initialised using 2·d rule (d = dimensionality), then adjusted by local density stability check (silhouette on core points).

Sanity guard: cross validated cluster persistence under small jitters.

Map Data & Preprocessing

Input: GPS traces / check ins projected to EPSG:3857; snapped to road/POI graph where applicable.

Outlier trimming with MAD; temporal thinning to reduce oversampling bias.

Application

Activity Range

Run auto param DBSCAN to obtain core clusters; compute convex/alpha shapes per cluster.

Merge home/work/amenity clusters via temporal patterns; output daily/weekly activity envelope

polygons.

Metrics: coverage area, typical radius, and time weighted centre of activity.

Results (Illustrative)

Stable estimates across mixed density tiles; cluster persistence > 0.9 under ±10% jitter.

Activity envelopes track commuting corridors and weekend hotspots in sample maps.

Limitations & Ethics

Requires consent and strict privacy controls; outputs must be de identified.

Sampling gaps can fragment clusters; provide uncertainty bands on envelopes.

Future Work

Graph aware DBSCAN variants; streaming updates; interactive visualization layers