Mobility, Data Mining and Privacy

www.geopkdd.eu

The technologies of mobile communications and ubiquitous computing pervade our society, and wireless networks sense the movement of people and vehicles, generating large volumes of mobility data. Miniaturization, wearability, pervasiveness is producing traces of our mobile activity, with increasing positioning accuracy and semantic richness:

- Location data from mobile phones: GSM cell positions
- GPS tracks from mobile devices receiving geo-positions from satellites

This is a scenario of great opportunities and risks:

- on one side, mining this data can produce useful knowledge, supporting sustainable mobility and intelligent transportation systems;
- on the other side, individual privacy is at risk, as the mobility data contain sensitive personal information.

GeoPKDD objectives: How to discover useful knowledge about human movement behaviour from mobility data, while preserving the privacy of the people under observation? GeoPKDD has started a new exciting multidisciplinary research area, at this crossroads of mobility, data mining and privacy.

Breakthrough analytical methods for mining from massive trajectory datasets

T-Warehouse

Problems: How to explore the aggregated properties of moving objects along the spatio-temporal dimensions, such as the presence or speed of vehicles in the various locations at varying times.

Idea: A T-Warehouse is a spatio-temporal data cube representing various aggregated measures of the moving objects.

Method: Our T-OLAP engine allows exploratory analysis, drilling up and down the spatial and temporal dimensions.

A reasoning component
- to specify domain-driven ontologies
- to infer types of trajectories and patterns

Visual Analytics environment

Visualization of T-Patterns
- to support the navigation of the extracted patterns
- interactive clustering of very large trajectory datasets
- user-driven exploration and evaluation of the discovered T-clusters

Visual exploration of the T-Warehouse
- visual T-OLAP operations to browse aggregated measures, such as aggregated presence and speed of moving objects (triangle base = presence, triangle height = speed)

Mastering the Complexity of the GeoPKDD Process

Semantic-based query and reasoning system

Spatio-temporal query primitives
- to select and pre-process trajectory data w.r.t. geographic background knowledge
- trajectory reconstruction and anonymization primitives

Trajectory mining primitives
- to extract various mobility patterns and models
- a Data Mining Query Language, to allow the user drive the entire analytical process

Method: How to discover groups of objects with similar movement behaviour in massive trajectory datasets, such as a systematic home-work home commuting behaviour?

Idea: A T-cluster is a set of similar trajectories, according to a repertoire of trajectory similarity functions.

Method: our prototype-based T-clustering algorithm discovers clusters in trajectory data