Semantic Enrichment of GPS Trajectories

Victor de Graaff, Maurice van Keulen, Rolf de By
University of Twente

Semantic annotation of GPS trajectories helps us to recognize the interests of the creator of the GPS trajectories. Automating this trajectory annotation circumvents the requirement of additional user input. To annotate the GPS traces automatically, two types of automated input are required: 1) a collection of possible annotations, and 2) a collection of GPS trajectories to annotate.

The first type of input can be a set of *points of interest* (POIs), activities, weather types, etc. This collection is to be provided by an application developer, and can originate from the web, an external knowledge base, or an existing database, for example.

The type of annotation that we are interested in, is annotation with visited locations, in order to create a user profile at a later stage. We have collected POIs by scraping the web, using a self-configuring data harvester. This harvester is based on workflows, enabling us to add or remove certain steps for different goals of harvesting.

The result of our harvesting approach consists of a set of 27,384 POIs, origining from the Dutch Yellow Pages [1], and contains an address and a geographical point representation for each POI. These point representations cannot be used to overlay the GPS trajectories directly, and therefore need to be converted into a polygon before providing useful input for the annotation process.

Several different approaches to this problem can be thought of, including Voronoi diagrams, nearest-neighbors, and geocoding the addresses of the assumed neighbors. For each of the POI *footprint size estimation* approaches, the output consists of two parts: 1) a *polygon* representing the estimated parcel, and 2) an *uncertainty function* based on the distance to the center of the polygon. These approaches are being validated with cadastral data for the region of Enschede, The Netherlands, and the result of the best approach is used as the input for the GPS trajectory enrichment.

The other type of input for the enrichment process is GPS trajectories. This data is generally not smooth, containing outliers, and interruptions of the data stream. Analyzing these imperfections however, may provide valuable information on users entering a rural area, or buildings, respectively.

Combining the results of the footprint size estimation with the analyzed GPS trajectory then provides us with *uncertain annotated GPS traces*.

References

[1] "http://www.goudengids.nl."