Problem Statement and Goals

Smita Singh, Abeer Alyasiri, Niyatha Rangarajan,
Moksha Srinivasan, Nicholas Lobo, Longwei Ye

September 25, 2022

Table 1: Revision History

Date	**Developer(s)**	**Change**
Sept. 23, 2022	Nicholas	Problem Statement
Sept. 25, 2022	Longwei Ye	Problem Statement
Sept. 23, 2022	Moksha Srinivasan	Goals, Stretch Goals
Sept. 23, 2022	Niyatha Rangarajan	Added definitions and edited sections for better understanding.
Nov. 23, 2022	Abeer Alyasiri	Incorporated rubric feedback
Apr. 5, 2023	Moksha Srinivasan	Final Documentation Update

1 Definitions

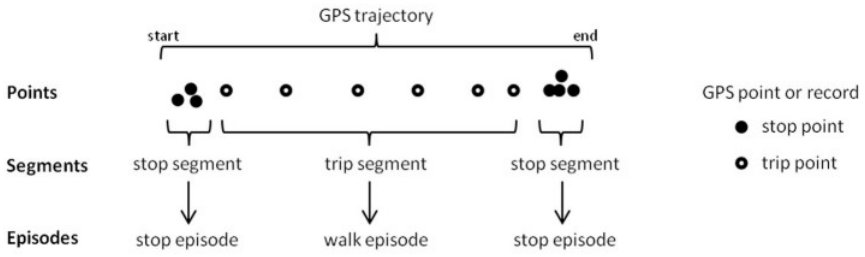

Fig. 1. GPS trajectory subdivided into points, segments, and episodes.

term	definition
Activity Locations (ALs)	Trip stops for amenities
ArcGIS	a geographic information system software for manipulating and visualization data.
Geographic Information System (GIS)	A computer system that analyzes and displays geographically referenced information
Geographic Positioning System (GPS)	A global navigation satellite system that provides location, velocity and time information.
Graphical User Interface (GUI)	A form of user interface that allows users to interact with electronic devices through graphical icons
Mode	The method of travel during a travel episode (walk, drive)

1.1 Episodes

In the context of activity analysis, a person's activities within a trace can be subdivided into episodes, which are differentiated based on type, an activity episode can be a stationary episode (stop episode) or a travel episode (e.g., car episode or walk episode). [1]

2 Problem Statement

2.1 Problem

With the large usage of global positioning systems, software has been created to help researchers develop tools and methods based off existing GPS data. One such software is the GERT toolbox made for ARCGIS Pro, that can match GPS

traces to transportation networks. This software suffers from specific data requirements that limit usability and functionality. Hence, it can not be extended to larger GPS data sets with newer data manipulation techniques. Additionally, it requires an expensive ARCGIS license with a monthly subscription fee. To overcome these weaknesses the GERT toolbox must be re-engineered with a focus on transferability, modularity, and scalability as well as being open source without using any proprietary software.

2.2 Inputs and Outputs

Input	Output
A dataset of latitude and longitude positions and times based on an entity's travel over the course of a certain period of time	Identify episodes of stop and travel based on the location behaviour.
	Identify the traveler's method of transportation that they are using during each episode.
	Estimate the possible route a traveler took based on mode and map-matched routes.

2.3 Stakeholders

Potential stakeholders in this project include any researchers and/or companies who are interested in matching GPS data to transportation networks in the context of travel episodes and route estimation analysis. Specifically, the stakeholders include the project's supervisor, Dr. Paez, the capstone professor, Dr. Smith, and marking teacher assistants.

2.4 Environment

yoGERT is an open-source toolbox that is able to run on personal laptops and desktops that uses Linux, Windows, and MacOS operating system with Python pre-installed.

3 Goals

Goal	Importance
The toolbox will be completely open-source without reliance on the proprietary ArcGIS or subsidiary software.	An open-source toolbox will allow a greater audience to process and use the abundance of GPS data that is collected every day. Additionally, it avoids dependency on expensive licensing required for proprietary ArcGIS Pro software.
The toolbox will quickly normalize and process common GPS file formats into compatible data types.	Similar to its predecessor, this toolbox aims to be highly flexible, allowing for a variety of input formats. Many existing tools are not input agnostic and as such cannot process a variety of data.
The toolbox supports fast and accurate route choice estimations.	Route choice estimations are models of possible routes a traveler may take based on their intermediary GPS information. These models can be used for a variety of applications. Examples include; forecasting traveler behaviour and predicting future traffic conditions on transportation networks.
The toolbox will be able to extract travel episodes, identify stops, intermediary trip information, potential activity locations (ALs), and assign trip purposes to episodes.	Many existing tools are able to extract travel episodes and assign trip purposes. By extending the functionality to include stop identification, ALs, and parse intermediary trip information, the toolbox provides much requested functionality for researchers.
~~The toolbox will be able to process 47.3 million points of GPS data within at most 6000 seconds.~~	~~Similar to its predecessor, this toolbox aims to be highly scalable, with the ability to process large amounts of data in a short amount of time. This makes it more viable to use GPS data for applications that require regular processing of data.~~

4 Stretch Goals

Goal	Importance
The toolbox will contain functionality to teach students about data processing methods, and develop intuition for how episodes are categorized.	This will allow the toolbox to be both a data processor as well as a learning tool for those who are new to GPS data analysis.
The toolbox will provide an interactive GUI with map overlays for episodes and corresponding information.	The toolbox will not only provide graphs based on inputs and desired outputs, but users will also be able to understand how the density of data points change for different episodes.

5 Bibliography

- [1] R. Dalumpines and D. M. Scott, "GIS-based episode Reconstruction Toolkit (Gert): A transferable, modular, and scalable framework for automated extraction of activity episodes from GPS Data," Travel Behaviour and Society, vol. 11, pp. 121–130, 2018.