Initial Plan
Generating Fake GPS Trajectories

Author: George Haddon-Gill
Supervisor: Padraig Corcoran

Project Description

The project I will be creating will generate fake GPS trajectories between two points that can be travelled between by a pedestrian. It will be coded in Python. Many location-based services (LBS), such as 'Strava' or 'Find My Friends', require users to submit their GPS trajectory which may come from a smartphone or Sat Nav system.

This project has a security application, if a user is concerned about their privacy, they may wish to obfuscate or hide their true trajectory. This security objective can be achieved by submitting to the LBS a set of fake trajectories along with the true trajectory. If the LBS or a nefarious individual cannot differentiate between the true and fake trajectories, they cannot determine which is the actual true trajectory.

The aim of this project is to develop a system for generating fake GPS trajectories which cannot be distinguished from real GPS trajectories. These fake trajectories will have to be as realistic as possible and there are various ways this can be achieved. Modelling the GPS data to have accurate noise and imprecisions - as you would see in real GPS data due to signal blockage, atmospheric conditions, and geometry etc. - will be essential for making seemingly genuine trajectories.

The success of the fake trajectories will be measured using a machine learning algorithm that will be given a training data set of trajectories labelled as whether they are real or fake. This will then classify a given unlabelled trajectory as real or fake. The closer the machine learning success rate is to 50% will determine how indistinguishable the real and fake trajectories are. Ideally, it should be no more likely to guess which is real or fake than randomly guessing.

Aims & Objectives

- Generate a path that links two points via the street network.
- Create a GPS trajectory along the path.
    - Determine how frequently the points which make up the trajectory need to occur along this path to be realistic.
- Obtain a real GPS dataset.
- Add noise to the GPS trajectory points that simulates inaccuracy in GPS data.
    - Determine which noise models are most successful in creating realistic trajectories.
- Create a way of measuring the success of a solution. A machine learning algorithm that has been trained with a large amount of real and fake data.
- Examine features of the map that could cause an impact on the GPS noise – e.g. Tall buildings, forests, mountains, tunnels, etc.

Work Plan

In addition to the objectives being completed each week, there will also be a weekly meeting with my supervisor on Tuesdays. This work plan will be dynamic as the amount of time taken to complete objectives is currently unknown and so will be estimated.

Date	Week	Objectives
31/01/2022	1	**Work:** The main objective for this week is creating the initial plan, to achieve this more knowledge about the topic will need to be acquired. It will be dedicated to gathering as much reading material about the topic as possible but not yet examining the details too closely. **Expected Outcome:** Have a finished initial plan. Have a large albeit unrefined reading list.
07/02/2022	2	**Work:** Whilst research will be carried out concurrently in every stage of the project, this week will be the most research-intensive week. I aim to find out as much as possible about the topic and determine different directions I can take it in. I aim to find as many traits of real GPS data that I can add to the simulated GPS data to maximise its credibility. I will also research what python libraries I could use in the project. **Expected Outcome:** Have the introduction, background, and most of the approach section of the report completed. Have some relevant python libraries downloaded and have a better understanding of how they work. Be in a position where the system can begin being coded.
14/02/2022	3	**Work:** Implementation will begin this week. I intend to create a program which will generate a path that links two points via the road network, most likely this will be the shortest path. This path should be turned into a sequence of points that make up a trajectory. **Expected Outcome:** Have a program that can link two geographical points via the street network. Plot points along this path that will comprise a fake GPS trajectory.
21/02/2022	4	**Work:** This week I will collect a dataset of real GPS data, that I can analyse the features of. Once a dataset has been found I will generate the fake trajectories in the same geographical location so that the comparison is fair. I will add some noise to the data – this will need to be perfected over time. **Expected Outcome:** Have access to an appropriate dataset of GPS data. Have a program that is able to generate GPS trajectories with a function that can apply noise to that data (i.e. move points by a set amount), this noise function will have parameters that can be adjusted to vary the amount of noise etc.

28/02/2022	5	**Work:** Some time will be set aside this week to get the report up to date with what has been created so far. Machine learning libraries will be researched, and time will be spent learning how to use the chosen machine learning library. **Expected Outcome:** The implementation section of the report will be filled out to this stage in the program's development. I will have a machine learning library downloaded that I have learnt how to make a basic classification system in.
07/03/2022	6	**Work:** The machine learning algorithm will be applied to my implementation, and it will be trained with a real GPS dataset and the current basic noise model version of the fake GPS trajectories. **Expected Outcome:** A working machine learning model that can distinguish between real and the current basic fake trajectories. An evaluation of how well it can differentiate between real and fake trajectories.
14/03/2022	7	**Work:** Data will be collected from the machine learning algorithm for use later on when comparing which techniques generate the most realistic fake trajectories. A comparison between the most basic noise model and no noise model will be carried out and the results documented. **Expected Outcome:** The machine learning algorithm will be fully working. A comparison of different noise parameters and no noise model at all will be carried out and documented in the report under the evaluation section.
21/03/2022	8	**Work:** Research and compare various types of noise to apply to the data. Change parameters to try and home in on the most optimal solution using the machine learning algorithm to assess the success. **Expected Outcome:** Have a way to choose and model different types of noise in the program. Have an idea of the best noise model for the data and some optimal parameters.
28/03/2022	9	**Work:** Formally document the data gathered in the previous week. Find a way to determine if there are "hot spots" on the map where the GPS signal would be worse such as tall buildings or tunnels – and how much worse it would be. **Expected Outcome:** Using data analysis techniques such as graphs and tables, I will show which is the preferred noise model. Have a way to find points of interest on a map where signal quality will be worse.
04/04/2022	10	**Work:** Using this data make changes to the noise model of the data to factor in that signal strength is not linearly distributed over the whole map and certain areas will be worse than others for inaccuracy.

		Expected Outcome: Have a feature that can be turned on or off that incorporates these areas of worse signal into the noise model. Compare this graphically when using different noise models with this feature turned on or off and add this to the report.
11/04/2022	Easter	Easter will be reserved as contingency time for if setbacks occur. It will be used to get caught up to where the work should be according to the time plan at this point.
18/04/2022	Easter	
25/04/2022	Easter	
02/05/2022	11	**Work:** These final two weeks will be for perfecting and finalising the report and the statistical analysis of the data. I will reach conclusions on the success of my project and how different additions to the simulations had improved or worsened the reliability of the results. **Expected Outcome:** Conclusion, results and evaluation section of the report will all be finished. Future work and reflections will be added. Report finished.
09/05/2022	12	