

Weekly report: May 30-June 5, 2015

Ivan Rios S.

June 5, 2015

1 Goals for the week

- Improve the trip detection algorithm

2 Activities

- Document and make the algorithm more readable
- Bug fixing in the algorithm
- Comparison of results
- Modification of GPS data cleaning algorithm to increase the number of points in a trajectory.

3 What I learnt

Initially, the script had several problems of structure and documentation of the logic. With Mikhail's help, I started by making the code cleaner and better explained. Then, we worked on simplifying the process followed by the algorithm. We cleaned 20% of the variables used in the algorithm that is currently being used. Up to this point, I found that:

- 20% more trips were detected. This is mostly based on the trips followed by professor Golab, professor Keshav and myself. Right now, there are several one way trips that with the new algorithm have become into 2 trips (back the origin of the trip).
- Based on the estimated length of each trip, the algorithm is detecting more accurately the total distances of each trip. In my own trips and professor Keshav's (fixed trips to university), there is a variation of 150 meters max but generally the total distance does not vary more than 50 meters.
- Professor Golab's trips have been challenging to detect. The results obtained before and now are very similar which suggests that there might

be problems with the data collected by the bike; the data is complete but there might be gaps of the trip when the data is not accurate which is generating these problems with the algorithm.

- With the modifications done to the algorithm, we have the following parameters with their default values in parenthesis):
 - Minimum time between consecutive points (10 seconds): the minimum amount of seconds between consecutive points that needed to pass in order to consider that point as part of the analysis (I propose deleting this parameters since I think the original idea was to reduce the number of data points processed by the server when a request was done in the webpage. Since it is an independent process now, I think it would be better to process every point)
 - Minimum charging current (10): this is the minimum value to consider that the bike is being charged.
 - Number of consecutive Magnetic Field value points saved for analysis (10): this is the number of data values saved in the algorithm in order to determine if there was movement by using the magnetic field.
 - Maximum magnetic field variation to be considered movement - higher limit (10): this is the value used to determine if there is movement based on the magnetic field data.
 - Minimum magnetic field variation to be considered movement - lower limit (5): this is the value used to determine if there is movement based on the magnetic field data (I propose eliminating this parameter as well; it is currently being used to determine if there is potential movement).
 - Minimum total distance (500 m): minimum total distance traveled to be considered a trip.
 - Maximum total distance (1000000 m): maximum total distance traveled to be considered a trip
 - Maximum distance between consecutive points (10000 m): maximum total distance allowed since the previous point in order to consider that point as part of the analysis
 - Minimum distance between consecutive points (18 m): minimum total distance allowed since the previous point in order to consider that point as part of the analysis (I recommend eliminating this parameter, again we should analyze every point)
 - Minimum distance between consecutive points to be considered significant movement (70 m): this is used as a secondary parameter to keep track of significant movement. I consider it secondary because the algorithm mainly focuses on the amount of time passed between consecutive points as an argument to decide to start or end a trip.

- Maximum time passed with not in a trip (300 s): This is the maximum number of seconds allowed to pass with no significant movement in trip. This means that if there is a trip going on, it will be ended if 300 seconds have passed with no significant movement.
- Currently, the code is still considering a great amount of variables so I think it would be required to delete some of them and simplify the process.
- Also, I modified the GPS data cleaning algorithm and it is now more sensitive to turns so the number of long straight distances (with long times between points) has been reduced.

4 Proposed goals for next week

- Publish the new algorithm in the “beta” web page.
- Determine the order in which parameters influence the results obtained from the the trip detection algorithm and make deeper analysis on appropriate values for them.
- Implement suggested changes.
- Research on data mining trip detection algorithms.