Cluster driver's destinations

The knowledge of a driver's commonly visited destinations leads to a better fuel economy in hybrid vehicles. The destinations are extracted by grouping GPS points from the driver's parking events using cluster analysis.

Carl Levin from Lund University and Christopher Håkansson from Chalmers have in collaboration with Volvo Cars developed a clustering algorithm, where the tuning parameters of the algorithm are automatically set. This clustering algorithm is used to group end-points of each trip made by the vehicle with the goal to identify important destinations. By knowing these destinations and the number of times they have been visited it is possible to predict to which destination the driver is going next. This enables the vehicle to extract the most probable route to that destination and by examining the route's properties the motor control can be optimized. This is especially useful in hybrid electric vehicles. By optimizing the use of the electric energy the carbon dioxide emissions are decreased.

The developed clustering algorithm is based on well-known clustering techniques. The examined clustering techniques belong to the group density-based clustering algorithms. This due to their ability to identify clusters of arbitrary shape. Parking locations are differently designed and therefore the spreading of GPS points will be of arbitrary shape. The algorithms can also handle noise points, which are generated when the vehicle is unintendedly shut off. The number of destinations varies with the user of each car and therefore it is necessary to have a clustering algorithm where the number of clusters does not need to be specified.

The clustering algorithms come with parameter settings which need to be tuned to generate the best clustering. The optimal parameter setting may vary between different vehicles and vary between different regions visited by the car. It was decided to develop a clustering procedure where the parameters did not have to be specified, but were to be adaptively set based on an internal evaluation measure. Internal evaluation meaning that it is not evaluated by a human. This measure looks at the clustering of the GPS points and weighs the similarity between the points belonging to the same cluster against the dissimilarity to the points belonging to other clusters.

The clustering algorithms and the internal evaluation method need some kind of distance measure to see which data points that are similar to each other. The distances measures considered were the Euclidean distance (the bee line) and the driving distance between the stops. Since the GPS points are generated by vehicles in a road network the driving distance was expected to give good cluster results.

Sadly the developed method where the parameters of the algorithms were adaptive set did not outperform the case when the parameters were set by experts. It only performed on par. This mostly because the internal evaluation method is not ideal for this kind of data. But no internal measure which suited the GPS data better was found. The driving distance did not improved the results either. This due to the difficulty to calculate accurate road distances using existing road map databases.