

COMP6208 Advanced Machine Learning Research Report Brief

University of Southampton
School of Electronics and Computer Science
Friday 13th February, 2015

Team **Seahorse**

Alexander Ally (aa2g11)
Hendrik Appel (hja1g11)
Samir Moussa (sm28g11)

For automobile insurers, telematics represents a growing and valuable way to quantify driver risk. Instead of pricing decisions on vehicle and driver characteristics, telematics gives the opportunity to measure the quantity and quality of a driver's behaviour. This can lead to savings for safe or infrequent drivers, and transition the burden to policies that represent increased liability. AXA, a large insurance company, has provided a dataset of over 50,000 anonymised driver trips. The intent of this research project is to develop a machine learning algorithm to discover the signature of driving type. Does a driver drive long trips? Short trips? Highway trips? Back roads? Do they accelerate hard from stops? Do they take turns at high speed? The answers to these questions combine to form an aggregate profile that potentially makes each driver unique.

Provided is a directory containing a number of folders. Each folder represents a driver. Within each folder are 200 .csv files. Each file represents a driving trip. The trips are recordings of the car's position (in meters) every second as (x,y) coordinates. In order to protect the privacy of the drivers' location, the trips are centred to start at the origin (0,0) and randomly rotated. A small and random number of false trips (trips that are not driven by the driver of interest) are planted in each driver's folder. The challenge is to identify trips which are not from the driver of interest, based on their telematic features, predicting the probability that each trip was taken by the driver of interest.

For this project, Seahorse aim to come up with a "telematic fingerprint" capable of distinguishing when a trip was driven by a given driver. The features of this driver fingerprint could help assess risk and form a crucial piece of a larger telematics puzzle. The main difficulties are feature extraction and feature representation which will be tackled using various machine learning methods. The group will consider multiple architectures and approaches including clustering algorithms, SVM and deep learning to assess the effectiveness of different methods and provide an optimal solution.