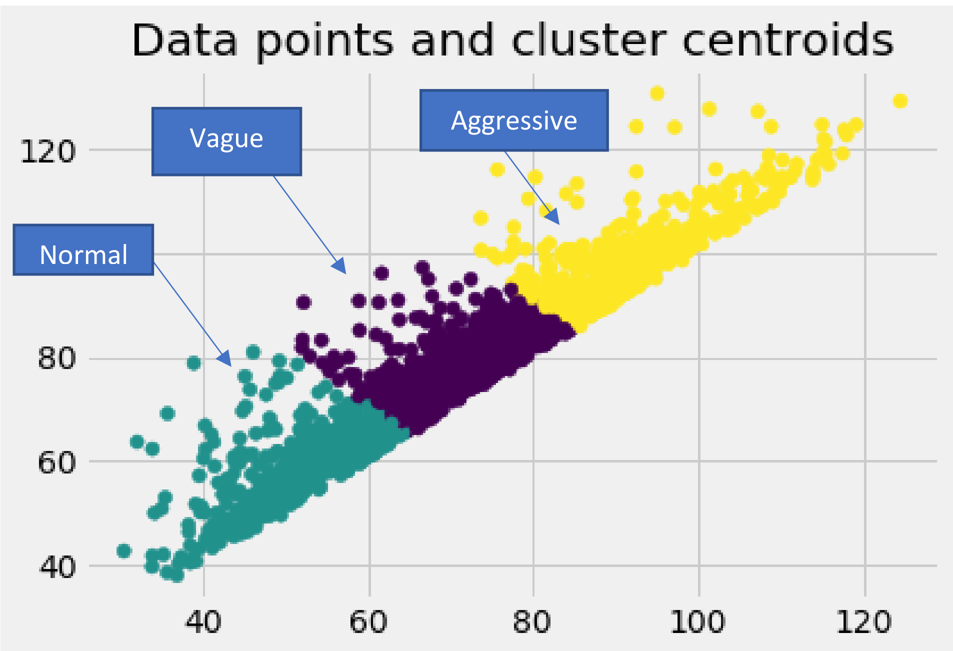
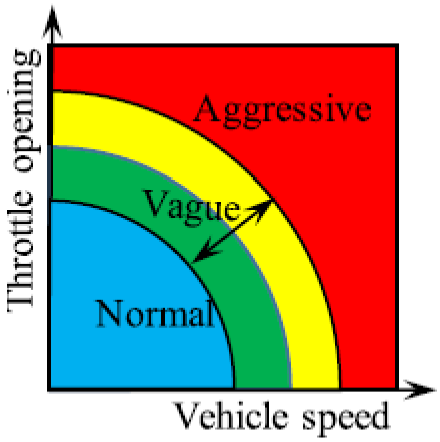
In this work we use a method called semi-supervised support vector machines (S3VM). S3VM are constructed using a mixture of labelled data (the training set) and unlabelled data (the working set). The objective is to assign class labels to the working set such that the "best" support vector machine (SVM) is constructed.Semi-supervised learning occurs when both training and working sets are nonempty. Semi-supervised machine learning is a class of machine learning that you have some labeled data sets and also an amount of unlabeled data. The semi-supervised learning approach is motivated from the fact that it is easier (and cheaper) to collect unlabeled training examplesThe usual support vector machine (SVM) approach estimates the entire classification function using the principle of statistical risk minimization (SRM). Theoretically, if there is adequate training data to estimate the function satisfactorily, then SRM will be sufficient. Supervised learning approaches are widely used for driving style classification; however, they often require a large amount of labeled training data, which is usually scarce in a real world setting.

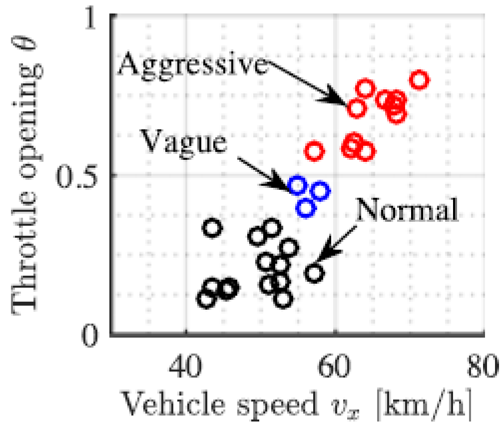
Some other approaches like Driving style survey analysis, SVM with Bayesian Filter or Hidden Markov Model can also be used but it still has the major drawback i.e. requirement of manually labelling lot of data and that it is very subjective , a driver on horizontal curve might seem aggressive to someone else and might not to other

In this case since Driving style is very subjective and the manual process of labelling the data is very time consuming. To address this problem,a semisupervised approach, a semisupervised support vector machine (S3VM), is employed to classify drivers into aggressive and normal styles based on a few labeled data points. First, a few data clusters are selected and manually labeled using a k-means clustering method. Then, a standard optimization tool is used to solve the nonconvex optimization problem. we also compare the S3VM method with a support vector machine method for classifying driving styles from different amounts of labeled data. Experiments show that the S3VM method can improve the classification accuracy by about 10% and reduce the labeling effort by using only a few labeled data clusters among huge amounts of unlabeled data.

I mainly focused on drivers' longitudinal driving behaviours on a horizontal curve road being as aggressive or normal driving First the Data is clustered using k-means , here I used 3 clusters, mainly because there'd be some drivers lying in vague driving style area ,other two are normal an aggressive







I have used linear kernel for both the SVM and S3VM (Gaussian Radial kernel might be used as well)

Data Used: All of the training and test data were collected
The device recorded a variety of parameters such as the throttle opening, braking force, vehicle position, steering wheel angle, longitudinal and lateral speed, yaw rate, roll angle, and acceleration at 50 Hz. A simple curve-negotiating task on a flat road was then chosen to reflect drivers' longitudinal driving The length of curvy road was 2247 m.

1) Feature Selection: the throttle opening could easily be used to distinguish driving styles. In this paper, we primarily focus on the longitudinal driving behavior when the vehicle is driving on a curvy road since it greatly influences fuel consumption .I selected the longitudinal speed V and the throttle opening, θ ∈ [0, 1](Normalized the data received) as the feature parameters to describe driving styles. Features are detailed as follows.

1) Vehicle Speed :high vehicle speed when driving on a curvy road, then the driver can be identified as an aggressive type. If a driver navigates the curvy road at a low speed, then the driver can be classified as a normal type.

2) Throttle Opening (θ): directly related to acceleration/deceleration. If prefer a larger longitudinal acceleration/deceleration ,tendancy to be more aggressive.

Acceleration is not taken since redundant (can be calculated from throttle and long. Speed)

2) Driver Participant: A total of 71 drivers data(vary)

3.)Data Preprocessing

In total, 400 datasets from 20 drivers were collected and each dataset consisted of about 6000 data points (2 min×60×50 Hz= 6000). The dataset of each individual driver was equally divided into four groups to get more data samples, and eventually, 1600 datasets were obtained.

n = 290 ,m = 1300, and r = 10.

to make labelling work easier, each dataset is clustered using the kmeans clustering method. For example, clustered points with a larger speed (e.g., vx = 80 km/h) and a smaller throttle opening (e.g., θ 0.2) could be directly labeled as aggressive. We only needed to label a few clustered data that were obviously subject to aggressive or normal driving styles with red and black circles as can be seen in

4.) Training and Testing Procedures: A moving window with width Td = 6 s [11] was then applied to all of the training data All feature data located within the specified window were clustered into a single point. These data consisted of the vehicle speed and throttle opening. All of test data were preprocessed in the same way as the training data.

SVM vs S3VM for No. of labelled datasets ∈ {9, 25, 45, 70, 145, 290}. the comparative accuracy of SVM using the labeled data and S3VM using both labeled and unlabeled data method. The vertical axis is the average classification accuracy μ and the horizontal axis is the amount of labeled training data on a log scale.