Telematics Data Project

March 17, 2015

The goal of this project is to create an R package that can visualize and feature engineer telematics data for a set of car drivers. The package also provides the ability to create a predictive model that can uniquely identify a driver based on driving data.

The data for this project is sourced from Kaggle (http://www.kaggle.com/c/axa-driver-telematics-analysis) . The dataset provides geographic location data for about 2800 drivers. Each driver has 200 trips and each trip has the Cartesian coordinates of the driver for each second of the trip. i.e a trip of 400 seconds will have 400 data points with the x and y coordinates of the driver separated by 1 second intervals.

Project Goals (stated-originally)

The project will aim to accomplish the following:

1. Based on the X-Y coordinates create features for velocity, acceleration, breaks, and fast turns and speed limit breaks. Note that all the features listed are considered important for getting to a unique driver signature.
2. Provide a function that can visualize each driver trip. Specifically, visualize the following.
    a. The X-Y coordinate movement for each trip
    b. Plot Velocity of the driver with time
    c. Plot acceleration of the driver with time
    d. Plot breaks of the driver with time
    e. Plot quantile charts for velocity , acceleration and breaks
3. Create a Mastersummary.csv file that summarizes data from all the trips based on the features created.
4. Create predictive model that can
    a. segment drivers into various driver types
    b. Uniquely identify a driver signature. i.e Given a new trip identify if the trip was from a particular driver or not.

Project Accomplishments (after the submission)

1. Successfully created an R6 class structure for an Insurance Telematics Scenario.
2. Key accomplishments include
    a. Telematics : R6 Class
        i. This class represents data of all Drivers and has methods that operate across all drivers. Eg: segment drivers (k-means clustering), load drivers, create summary file etc.
        ii. This class provides the methods for feature engineers various features across all the driver trips. i.e calculates velocity , acceleration , breaks etc.
        iii. This class is multithreaded and can run in parallel to create the master dataset. The parallel package is used to do the same.

b. Driver : R6 Class
   i. This class represents a single driver and provides methods that are applicable on a specific driver
   ii. We pass the reference of the Telematics class to Driver class so that a driver can interact with the Telematics class.
   iii. All the trip plotting methods are provided in this class.
   iv. This class provides the ability to create a driver signature predictive model based on GBM.
c. telematicsUtil.R : Utility R File
   i. This R file provides a collection of methods used extensively in the package.
   ii. The meat behind velocity, acceleration, breaks and speed limit calculations are implemented in this file.
   iii. All the graphing methods are implemented in this file.
d. Documentation
   i. Documentation , demo's and tests are provided in the package structure
   ii. The table below lists the various locations of files related to documentation

Location	Description
Vignettes\project-summary.pdf	The Project Summary file based on the RNW file
Vignettes\examples\TelematicsExample.Rmd	R Markdown file with examples
Vignettes\examples\summary\Summary.pdf	This document
demo	R files containing all demos

Team Member

1. Niraj Juneja (SUNet ID: niraj9) – niraj9@stanford.edu
2. Shiro Miyao (SUNet ID: smiyao) – shiro-miyao@kcsg.com.sg

Target Audience and Use cases

The target audience for this package is insurance companies who want to price the insurance policy based on driving behavior. The driver signature created can be associated to a Profit and Loss statement and the policies can be subsequently priced. In addition, there is another use case for car manufactures to improve the safety features of a car based on the driving behavior of drivers.

Software and Languages

We used R.