DVT DRIVESTATS ALGORITHM

DriveStats

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

TEAM ANTZ

what does the dvt drivestats algorithm do?

Our algorithm makes use of a few very interesting properties of common statistics to give a driver a score for the quality of the their driving out of 10.

how does it work?

Using the data gathered from sensors built into all modern android devices we are able to determine a person's acceleration in the x, y, and z axes, as well as many other important facts. By taking samples from these sensors at regular intervals (currently we have found that 1/3 of a second works well), we are able to estimate the manner in which an individual has been driving.

ALGORITHM STEPS

Collect data from the Android's sensors.

We determine many bad things all previous users have averaged per second. And transform that data into a normal distribution.

Determine how many bad things the current user does per second.

One of the useful properties of the data we exploit, is that we are able to estimate a position on a normal distribution for the likelihood of the observed results in terms of the number of bad things per second.

Weightings are applied to the different readings from the sensors (because acceleration through a turn is worse than acceleration while moving forward, and acceleration upwards is even worse!)

Using this information, we are able to estimate a probability that the score of the current trip was a result of random chance when compared to the normal distribution of all other drivers.

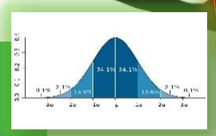
From this data we calculate a Z-score for the distance of the person's number of bad things from the normal distribution of the number of bad things.

Using our z-score and by integrating the normal distribution function, we are able to estimate the area to the left of the normal distribution's score.

Because the area under a normal distribution always equals 1, we can simply multiply the area by 10 to achieve a score out of 10 for the quality of that person's driving.

$$z = \frac{X - \bar{X}}{S}$$

where z is the standard score,
S = the standard deviation of a sample,
X = each value in the data set,
$\bar{X}$ = mean of all values in the data set.



$$f(x) = \frac{1}{\delta\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\delta}\right)}$$

WHAT PROBLEMS DID WE OVERCOME?

1 what data should we gather?

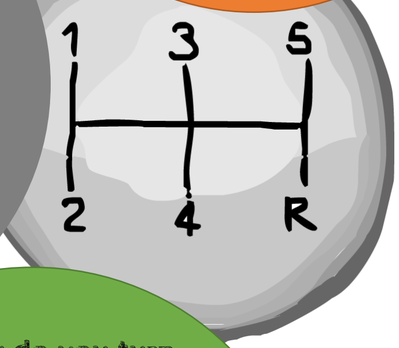
We decided that the most meaningful data we could collect would revolve around the acceleration, speed, and location of the car

2 how do we determine what constitutes bad driving?

We decided to use online data about driving to determine what acceleration is considered bad driving. The scores we arrived at were:
Forward Acceleration: $4.2ms^{-2}$
Cornering Acceleration: $3ms^{-2}$
Vertical Acceleration: $2ms^{-2}$

3 how do you determine a value for the quality of someone s driving when there is no objective measure of driving?

We used the data of our user and compared it to a normal distribution to determine how the number of poor data readings at any given instant compared.



4 how do you turn randomly collected data into a normal distribution?

Fortunately, follows a layout called a Poisson distribution; because it expresses the probability of a given number of events occurring in a fixed interval of time and/or space if these events occur with a known average rate. Because of several properties of our data we observe out data closely approximating a normal distribution.

The DVT DriveStats Algorithm would not have been possible without the invaluable help, advice, and insight, provided by the two following individuals:
- Prenil Sewmohan
- Jolandi Swanepoel