USAGE-BASED VEHICLE INSURANCE

Methodology for UBI predictions based on driving styles

Holon Technologies OÜ

Aadress: Laada st 20
Rakvere, Estonia

Phone +372 504 99 66
Email raul@holontech.eu

Our Question

We would like to know efficient methods how to use driving behavioristic data combined with accident data in order to make sensible predictions about possible damages.

Our main questions regarding the problem to which we seek solution are:

1. A method how to classify source data (which methodology to use).
2. A model how in a combination of different classes, to determine the probability of damage, damage frequency and the probable amount of the damage
    2.1 . how the premium should be changed as a consequence.
3. How to find trend and how to find the probability of the trend to continue.

We can provide the source data as a SQL database (route data and damage data). The data contains RAW data and various extracts: hour view, day view, week view, month view, period view.

If necessary, a high-capacity server will be delivered to make the work of scientist group convenient. Our team will support with a specialist in insurance, IT, legal and business.

Yours faithfully,

Holon Technologies team

_____

Data structure on routes and accidents

Below is the data table generated on hourly basis which gives overview of the data to be provided.

1. Telematic Data

1.	Date	01.01.2018
2.	Weekday	Monday
3.	Time range	1 (00:00-01:00)
4.	Average speed (km/h)	55
5.	Maximum speed (km/h)	75
6.	Idle time (minutes)	11
7.	Driving time (minutes)	10
8.	Distance (km)	100
9.	Number of stops (times)	3
10.	Number of routes (count)**	2
11.	Country	Estonia
12.	Location (GPS coordinates) *	If available
13.	Location (in words) *	Harju
14.	Sensor has been disconnected (times)	2
15.	Route distances: 1-5 km (times)	0
16.	Route distances: 6-10 km (times)	0
17.	Route distances: 11-20 km (times)	1
18.	Route distances: 21-30 km (times)	0
19.	Route distances: 31-40 km (times)	5
20.	Route distances: 41-50 km (times)	0
21.	Route distances: 51-75 km (times)	0
22.	Route distances: 76-100 km (times)	0
23.	Cornering 10-44 degrees (times)	4
24.	Cornering 45-120 degrees (times)	1
25.	Cornering more than 121+ degrees (times)	5

2. Accident Information

A1.	**Accident date/time**	**01 Jan 2018 / 15:15**
A2.	Accident place (from telematics)	Coordinates
A3.	Accident place (from accident database)	Coordinates
A4.	Drivers fault (yes/no)	1
A5.	Damage costs (EUR, bodily injury)	
A6.	Damage costs (EUR, pain and suffering)	
A7.	Damage costs (EUR, without annuity)	50
A8.	Damage costs (EUR, with annuity)	150
A9.	Accident classification	Speeding

3. Example Vehicle Information

We provide following information regarding each vehicle:

Car registration number	ABCD1234 **
Car type	Passenger car
Usage	Regular usage
GPS device type	5
Device mounting	Fixed, stable
Company registry code	12345678
Company field of activity	Transport services

There will be 1500 vehicles per every car type (light vehicle and truck) at least 3 years data.

Additionally: if needed, we will format/convert the data according specified requirements.

**The data is pseudonymized and we do not provide exact end-points, it's a matter of confidentiality (we left out coordinates 10 km from departures and from end-destination).*

***If the route is longer than 1h then route is there is marked only beginnings of routes*

About the project

Insurance companies benefit from UBI by identifying good drivers. **We wish ESGI to help us with developing methods and formulas for systematizing and analyzing telematics data necessary for UBI pricing.** We believe that well developed methodology for modelling the classes and prediction of class changes based on driving styles for different vehicle types enables the wide usage of UBI in the Baltics resulting with positive impact in general.

Usage based insurance enables vehicle insurers to improve pricing mechanisms based on actual driving data, gain better control of claims and differentiate their offers to current and prospective policyholders. Usage-based policies provide people with the same coverage that conventional vehicle insurance policies do but are often less expensive (up to 40-60%). When customer safe driving habits are demonstrated through telematics data, private person or company may save from insurance costs and driving behavior can be differentiated. A usage-based insurance client receives benefits such as reduced cost for vehicle insurance and driving style feedback.

With better understanding of customer's risk profile, insurers will be able to provide tailored and actual risk-priced products. Better understanding about the data provides many improvements. Most importantly, cost for insurance will become transparent, customer can in real time control of his/her actual driving behavior and through this benefit a better price for insurance compared to a customer without telematics device. It is also very important for insurers in customer retention aspect.

Problem solving

Below is the overview of our understanding on how to attend the matter in question. We see that scientists can provide valuable input for more appropriate methodology and adequate detailed view on the problem solving and developing a formula.

Data collection

Telematic data is collected using car-mounted equipment, with different sensors (GPS, gyroscope, etc.) which is collecting data and transmitting it via internet to the database of the company that collects the telematics data.

Data collection devices generates a row in the database, if the vehicle:

- passes 1 kilometer, or
- drives 1 minute, or
- driving direction changes by more than 10 degrees.

In reality, for example, on a winding road the data can be generated on almost every second (more than 10 degrees turn, intense route). Due to these rules, a very precise data set on the driving characteristics is generated. On the other hand, it is very important to measure if there is any effect on driving behavior in case vehicle does not make any stops during long journeys, night driving etc. (monotonous route) and how driving behavior in such way is measured with telematics device.

The driving data contains, as a minimum, information on distance travelled, speed, time, cornering. Depending on the model of the data collection device, additional data may be obtained via so-called eco-driving. Eco-driving data is the number of accelerations and braking, as well as the angle of entry into the bend and the angle of taking the bend.

Information for insurance companies

Typically, insurers determine the price of a premium by grouping information known about the client. For business clients, such information is field of activity, type of vehicles (e.g. truck, van, car), area of operation, expected mileage per year, previous claims experience (number of fleets, claims, earned exposures), etc. This means that insurers assess the risk and price using the amount of information available for best pricing. For example, if the field of activity is courier service, there is a statistically higher risk of damage (insurance experience) because vehicle drives usually on peak times at cities with high traffic density and more probability to get into traffic accident

By analyzing real driving data and the attached real claims experience data, **it is possible to find a correlation between the driving style and traffic accidents** and thus automate the determination of the risk and in future the cost of insurance premium. Such analysis has not yet been adopted in Estonia or in Baltics neither used in Finland, but insurers are interested to invest for future technologies which help better risk understanding.

Classification of data

In the course of analyzing driving data, it is necessary to classify the data. In addition, different weight should be given to different source data. For example, the weight of the average speed may be higher in the formula than the weight of the amount of braking.

The vehicle type is a very important parameter and it can be assumed that different vehicle types (van, car, truck, also vehicle make & models, fuel type, chassis possible) have different driving characteristics and also different logics in spreading into different classification.

Data can be classified according to various driving parameters:

For example, a mileage class where the I group includes cars with a mileage of less than 500km/month and the II group cars with mileage of up to 1000km/month, etc. The I group in the speed class includes cars with an average speed of less than 60 km/h and the II group cars with an average speed of over 60 km/h. In addition, those classes can in turn be combined into new form of classes. For example, cars that drive up to 500km/month and have average speed of less than 60km/h. are likely to drive mainly in urban areas. If GPS data is included there, based on the third-class data, it is possible to make a fairly accurate decision whether it is a vehicle driving mainly in urban areas or just a vehicle driving very carefully on a highway.

It is important to take into account route distances (0 – 5km, 6 – 10km, 11 – 20km, 21 – 30km, 31 – 40km, 41 – 50km, 51 – 75km, 76 – 100km, 101 + km)

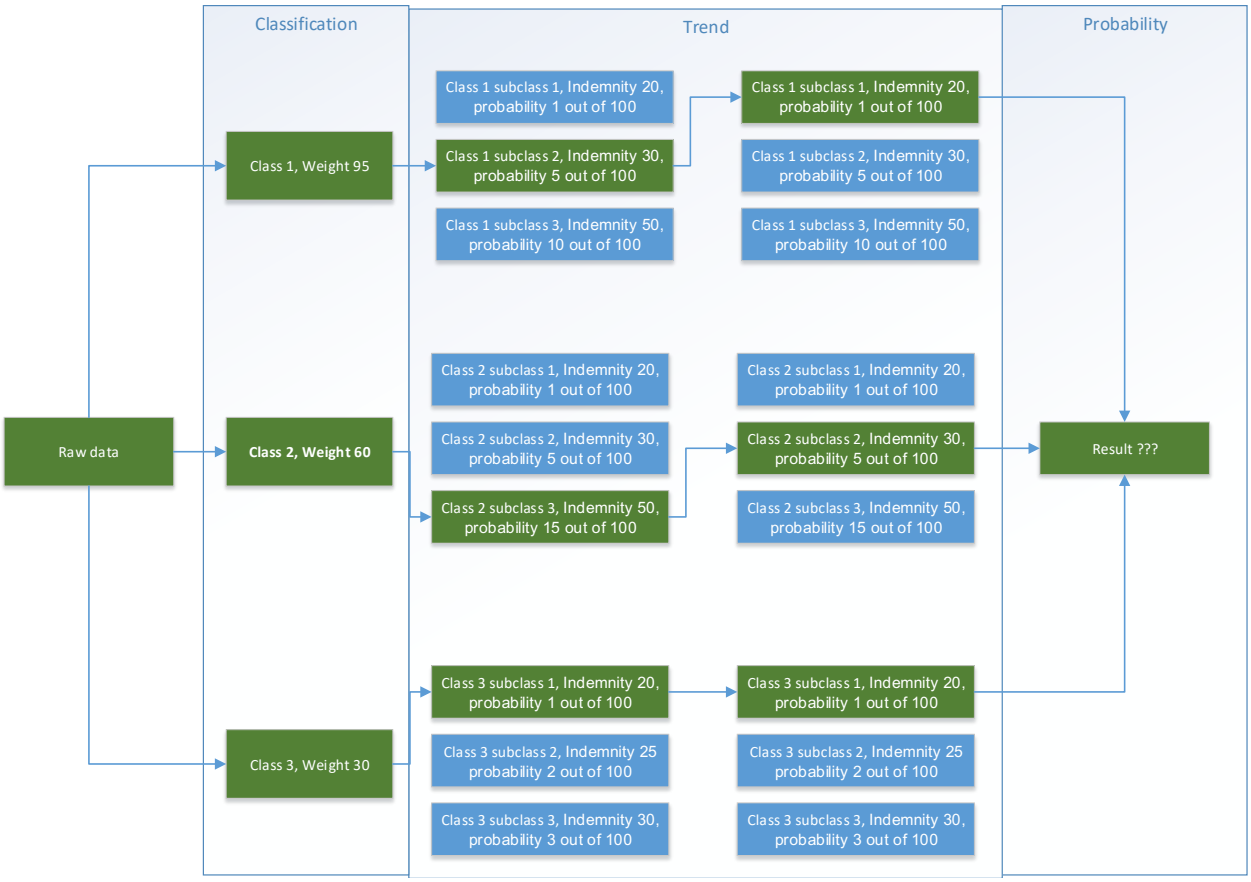
Classes should also have different weights. We assume that a class that involves mileage has a greater weight than a class that groups eco-driving data (acceleration, braking, and cornering). This is due to the fact that in case of different devices, eco-driving data is more or less reliable, while mileage, current speed and location are almost 100% correct.

As a result of the classification the driving characteristics of a particular vehicle are known for a certain period. What is also known is the trend – how the driving characteristics have changed over the period. For example, a vehicle that has previously exceeded the speed limit may start to behave more and more by rules in order to reduce the risk of damage and thereby get a better premium from insurance. With different classes and speed being just one if the parameters, it is necessary to know what happens upon classification of a specific vehicle if a similar trend continues in the future. For example, if a vehicle often starts over-speeding in the summer months, but in autumn starts driving within the speed limit again: what should be the classification of its speed if the insurance period is half a year (1 Apr – 1 Oct). How to make a prediction about future behavior? Which period can be predicted most reasonably or what is actually the reasonable period of data to be relied on?

It is apparently not possible to calculate the possible cost of a specific premium. However, since after classification we know the amount of damages standing out in different classes, that information can be used primarily to calculate the probability of damage in case of a particular driving style and, secondly, using that probability and taking into account the (average) amount of damage for such class, to predict the amount of probable damage as well as whether the amount of the premium should be increased or decreased and to what extent.

One vehicle may have more than one driver. A vehicle may also drive in several countries. For example, in Italy, the probability of getting into an accident is several times higher and the expected severity is supposedly also significantly higher.

One possible way to handle the data:



Additional questions

- How to use prediction errors in order to enhance the prediction itself?
- What kind of (open source) analytics tool(s) or software can be used for such analysis and calculation of a model in the future?
- What are the fatal incidents to harm a system and how to prevent this?
- How to avoid pricing errors?
- How to prove that the selected method is reliable?