Convolutional Neural Network Classification of Telematics Car Driving Data

Guangyuan Gao [1] and Mario V. Wüthrich [2,*]

[1] Center for Applied Statistics and School of Statistics, Renmin University of China, Beijing 100872, China; guangyuan.gao@ruc.edu.cn

[2] RiskLab, Department of Mathematics, ETH Zurich, 8092 Zürich, Switzerland

[*] Correspondence: mario.wuethrich@math.ethz.ch

Abstract: The aim of this project is to analyze high-frequency GPS location data (second per second) of individual car drivers (and trips). We extract feature information about speeds, acceleration, deceleration, and changes of direction from this high-frequency GPS location data. Time series of this feature information allow us to appropriately allocate individual car driving trips to selected drivers using convolutional neural networks.

---

1. Introduction

General insurance companies have started to collect telematics car driving data which comprises high-frequency GPS location data (second per second), road conditions, day time of trips, etc. Telematics car driving data allows insurance companies to obtain driver-individual information, for instance, about driving distances, duration of trips, speed, force of acceleration and deceleration or changes of direction. Several papers have started to investigate risk factors from telematics car driving data with the common goal of obtaining more accurate risk profiles of individual car drivers. Ayuso et al. (2016a) and Ayuso et al. (2016b) study the risk exposures with respect to driving distances. They use the time elapsed until the first accident occurs as response variable to establish a survival model. Lemaire et al. (2016) study the predictive power of annual driving mileage for the number of claims. Boucher et al. (2017) refine the previous study by establishing a generalized additive model to analyze the exposures of both the driven distance and the duration. Paefgen et al. (2014) and Verbelen et al. (2018) further partition the total driving distance by road type and time slots as explanatory variables for claims frequency modeling. The above papers do not consider the telematics car driving data in time series structure, but they extract summary statistics from this telematics car driving data. More closely related to our work is Weidner et al. (2016) who analyze individual driving maneuvers using the discrete Fourier transform.

Another closely related stream of literature studies driving cycles. Driving cycles consolidate telematics car driving data to get speed-time profiles of driving behaviors. This stream of research aims at developing typical driving patterns which help us to understand vehicular emissions, energy consumption and impacts on traffic, we refer to Esteves-Booth et al. (2001), Hung et al. (2007), Wang et al. (2008), Kamble et al. (2009) and Ho et al. (2014). These papers study driving cycles in different cities of the world; since each city has unique driving characteristics (road topography, volume of traffic, vehicle fleet composition, etc.) these driving cycles differ between cities. In our previous work Gao and Wüthrich (2018), individual car driving trips have been aggregated to so-called speed-acceleration (v-a) heatmaps. In doing so, the law of large numbers is applied which allows us to reduce the sheer amount of telematics data, while still capturing crucial driver-individual information.

These *v-a* heatmaps capture the same information as the normalized speed-acceleration matrix described in Section 3.3 of Kamble et al. (2009) and illustrated in Figures 5 and 6 of Hung et al. (2007). Of course, this aggregation to *v-a* heatmaps leads to a loss of information, for instance, the time series structure of individual trips gets lost. This loss of information may be unsatisfactory. Therefore, we tackle the problem from a different angle in this work.

The aim here is to label individual car driving trips by directly analyzing speed, acceleration/braking, and changes of direction time series of individual trips. The benefit of this method is that we can consider and classify each trip individually, according to its characteristics and time series structure, and only the resulting labels need to be transmitted, say, to the insurance company (not the entire high-frequency GPS location data). Please note that many black box devices in cars only submit summary statistics; however, these summary statistics are often of a rather trivial nature like a maximal acceleration threshold has been exceeded in this trip, etc.

Let us give a simple consideration about the amount of data that is collected by high-frequency telematics data (second by second). On average we may collect roughly 100 KB of telematics data per driver and per day. This amounts for a typical driver to roughly 40 MB of telematics data per year. Thus, for a considerably small insurance portfolio of 100,000 drivers we collect roughly 4 TB of telematics data per year. This is a comparably large amount of data where, for instance, even identifying a given driver every day in this data may result in a non-trivial task. Therefore, one often aims at replacing this detailed (big) data by (explanatory) summary statistics.

In the present manuscript we select three drivers. For each of these three selected drivers we have roughly 250 individual trips, each of length 180 s and speeds in $[2, 50]$ km/h (urban speeds). The aim is to randomly select such a trip of 180 s and to allocate it to the right driver. By pure random allocation we would get roughly 33.3% of the allocations correct. We will train a deep convolutional neural network (ConvNet) for this labeling task, and we will see that these ConvNets allocate correctly more than 75% of the individual trips (out-of-sample). We believe that this is a very remarkable result because it is based on only 180 s of driving experience.

We use ConvNets for this classification task because they have proved to be very powerful in image and pattern recognition. Yann LeCun is often recognized for having invented ConvNets, but it is difficult to give a specific reference of LeCun because he developed ConvNets during his work at Bell Laboratories where he has been engaged since 1988. ConvNets have been applied very successfully for image and video recognition as well as for natural language processing. An important feature of ConvNets is that they have translation invariance properties which allow them to find similar structures at different places in images, see Zhang et al. (1988), Zhang et al. (1990) and Wiatowski and Bölcskei (2018). Other neural networks that are used for time series analysis are recurrent neural networks and long short-term memory networks. These may have the advantage that they can handle time series of different lengths, we do not consider these other networks in the present work.

Organization of this manuscript. In the next section we describe how the time series of speeds, acceleration/braking and changes of direction of each individual trip are constructed. Moreover, we provide descriptive statistics of these variables. In Section 3 we select three drivers, and we train a deep ConvNet to classify their trips. The resulting out-of-sample classification analysis is presented in Table 2. In Section 4 we demonstrate how well this ConvNet classification carries over to other triples of drivers. In Section 5 we classify the trips of all drivers according to the ConvNet found in Section 3, and we compare the results to the ones from Gao and Wüthrich (2018). Finally, in Section 6 we conclude, and we give an outlook.

2. Description of the Data

Before describing the data and its cleaning process we disclose the limitations of our data. The available data consists of GPS locations second per second of individual trips of $n = 416$ different car drivers (over a given time period). Unfortunately, we do not have more information than that, i.e., we neither have information about road conditions, day time of trips, car brand, size of car, etc.

Moreover, for confidentiality reasons, the GPS locations do not show the true coordinates, but they are initialized to always start in coordinate $(0,0)$. Of course, this limits our analysis considerably, nevertheless, the results that we obtain are still remarkable.

2.1. Pre-Processing Telematics GPS Data

We consider GPS location data second per second of $n = 416$ different car drivers. From this GPS location data, we calculate the average speed v_t, the average acceleration and braking a_t, and the average change in direction (angle) Δ_t every second t of each trip of each driver. Let (x_t, y_t) denote the GPS location (in meters[1]) every second t of a single trip of a given driver. From this GPS location data we calculate the average speed at time t (in m/s)

$$v_t = \sqrt{(x_t - x_{t-1})^2 + (y_t - y_{t-1})^2},$$

and the average acceleration at time t (in m/s^2)

$$a_t = v_t - v_{t-1}.$$

For a positive average speed $v_t > 0$ at time t, we define the direction of the heading by

$$\varphi_t = \text{atan2}\left(y_t - y_{t-1}, x_t - x_{t-1}\right) \ \in (-\pi, \pi],$$

where atan2 is a common modification of the arctan function that transforms Cartesian coordinates to polar coordinates such that the resulting polar angle is in $(-\pi, \pi]$. For positive speeds $v_s > 0$ at times $s = t - 1, t$ we can then consider the change in direction from $t - 1$ to t given by $\varphi_t - \varphi_{t-1} \in (-2\pi, 2\pi)$. For the change in direction (angle) at time t we define

$$\Delta_t = \left|\sin\left(\varphi_t - \varphi_{t-1}\right)\right|.$$

We choose absolute values of changes in angles because the subsequent analysis should not be influenced by the signs of the angles. Moreover, we choose the sine of the change in angle: the reason for this choice is that GPS location data has some imprecision[2] which manifests stronger at very low speeds, say, when the car is almost standing still. Taking the sine of the change in angle slightly dampens this effect, but requires $\varphi_t - \varphi_{t-1} \in (-2\pi, -3\pi/2] \cup [-\pi/2, \pi/2] \cup [3\pi/2, 2\pi)$ for identifiability reasons. Please note that the latter is usually fulfilled because changes of directions within one second cannot exceed $\pi/2$.

In Figure 1 we illustrate the first trips of the three drivers 57 (left), 206 (middle) and 238 (right). The lower line in blue color shows the speed pattern $(v_t)_t$ (in km/h), the upper line in red color shows that acceleration/braking $(a_t)_t$ (in m/s^2), and the middle line in black color shows the changes in angle $(\Delta_t)_t$. We note that changes in angle are bigger at lower speeds. From driver 238 we can very well see that these changes in angle often go along with braking first and then accelerating after changing the direction of the heading.

We remark that for the plots in Figure 1 and for the subsequent analysis we have applied further data cleaning. Maximal acceleration and braking a_t has been capped at ± 3 m/s^2. Weidner et al. (2017) state that normal acceleration goes up to 2.5 m/s^2, and extreme acceleration can go up to 6 m/s^2

---

[1]   The available GPS location data is given in meters and rounded to one decimal place.

[2]   The rounding of GPS locations to one decimal place provides a first source of imprecision which has a stronger influence on acceleration and changes in angle at speeds close to zero. A second source of imprecision may be that the GPS signal itself is not fully precise w.r.t. position and timing. Finally, it may happen that the GPS signal is not received at all, for instance, while driving through a tunnel. The latter can be identified more easily because it leads to missing values or accelerations beyond physical laws (if missing values are not marked). Our data does not have any missing values.

for vehicles driving straight ahead. Braking may be stronger and may vary up to $-8\,\text{m/s}^2$. We cap acceleration/braking at $\pm 3\,\text{m/s}^2$ for data quality reasons, because higher values may (in our case) also indicate an imprecise GPS signal (and we do not have sufficient observations beyond this capping; note that Hung et al. (2007) cap at $\pm 4\,\text{m/s}^2$). Our capping looks reasonable in view of the red graphs in Figure 1. Moreover, changes in angles are capped at $\pi/6$ per second which corresponds to a capping of Δ_t at $1/2$, and we set the change in angle equal to 0 for speeds below $3\,\text{m/s}$. The latter has mainly been done for data quality reasons; imprecise GPS location signals (rounding of GPS locations in meters at one decimal place) affect more negatively changes in angles at low speeds, also note that changes in angles are only defined for strictly positive speeds.
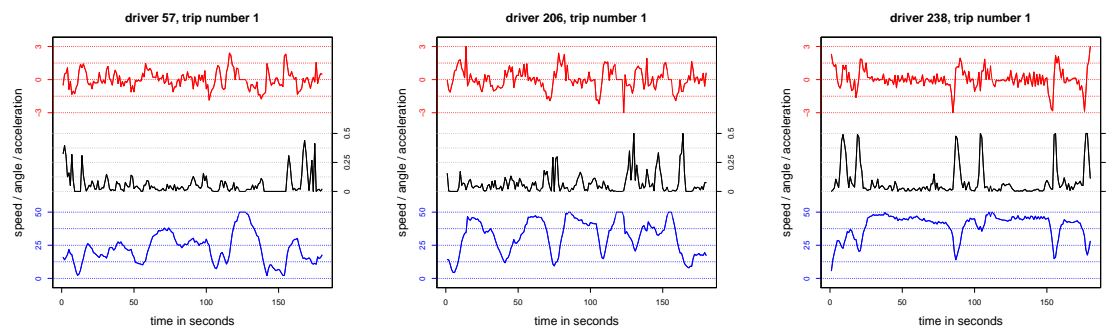

Figure 1. First trips of drivers 57 (**left**), 206 (**middle**) and 238 (**right**): the lower line in blue color shows the speeds $(v_t)_t$ (in km/h), the upper line in red color shows that acceleration/braking $(a_t)_t$ (in m/s^2), and the middle line in black color shows the changes in angle $(\Delta_t)_t$.

Finally, we aim at studying driving styles and not driving habits, i.e., we would not like to classify drivers with respect to being an urban driver or a long-distance highway driver, but we would rather like to understand whether we have an aggressive driver or a calm driver. For this reason, we only consider the parts of the trips that have speed v_t in $[2, 50]$ km/h (urban speeds). Therefore, we concatenate the trips correspondingly, which is done by the code illustrated in Listing 1.

Listing 1. R script for the concatenation at low speeds.

```
1  trip <- trip[which( (trip$v>=2*10/36) & (trip$v<=52*10/36) ), ]
2  trip[which( trip$v>50*10/36), 2:3] <- 0
3  trip$v <- pmin(trip$v, 50*10/36)
4  trip <- trip[1:180,]
```

On line 1 of Listing 1 we select the parts of the trips that have speeds in $[2, 52]$ km/h (note that $10/36$ transforms km/h into m/s). On line 2 we set both the acceleration/braking a_t and the change in angle Δ_t equal to zero for the speeds v_t in $[50, 52]$ km/h, on line 3 we cap the speeds at 50 km/h, and finally on line 4 we keep the first 180 s of these trips. The lower bound on the speeds in Listing 1 has the effect that we get rid of the waiting times in front of red traffic lights (and similarly for traffic jams), i.e., the idling mode. Please note that the idling mode can be significant in urban traffic, for instance, in Chinese cities it can reach up to 40% of the total driving time, see city of Shanghai in column $P_i(\%)$ of Table 7 of Wang et al. (2008). The upper bound provides a smooth concatenation between driving sections at urban speeds.

For our $n = 416$ drivers, this provides between 76 and 460 individual trips of length 180 s (in speed bucket $[2, 52]$ km/h), and on average we have 249 trips of 180 s per driver. The three selected drivers 57, 206 and 238 of Figure 1 have 205, 234 and 213 such individual trips at urban speeds of 180 s length, and the first trip of each driver is illustrated in Figure 1. In Section 3.2, below, we provide more summary statistics for these three selected drivers, for instance, the densities of the average speeds per trips are given in Figure 5, below.

2.2. Descriptive Statistics of Individual Trips

In this section, we introduce important empirical indicators that allow us to distinguish the different drivers. For the moment, we fix one driver, and we assume that we have S individual trips of length $T = 180$ s of this driver. Denote the speed, acceleration/braking and change in angle of trip $s = 1, \ldots, S$ of that driver by $(v_{s,t}, a_{s,t}, \Delta_{s,t})_{1 \leq t \leq T}$. We define the average speeds, acceleration/braking and change in angle by

$$\bar{v} = \frac{1}{S} \sum_{s=1}^{S} \bar{v}_s \qquad \text{and} \qquad \bar{v}_s = \frac{1}{T} \sum_{t=1}^{T} v_{s,t},$$

$$\bar{a} = \frac{1}{S} \sum_{s=1}^{S} \bar{a}_s \qquad \text{and} \qquad \bar{a}_s = \frac{1}{T} \sum_{t=1}^{T} a_{s,t}, \qquad (1)$$

$$\bar{\Delta} = \frac{1}{S} \sum_{s=1}^{S} \bar{\Delta}_s \qquad \text{and} \qquad \bar{\Delta}_s = \frac{1}{T} \sum_{t=1}^{T} \Delta_{s,t}.$$

In Figure 2 we plot these summary statistics $\bar{v}$, $\bar{a}$ and $\bar{\Delta}$ of all $n = 416$ drivers. We observe a positive linear dependence between average speed $\bar{v}$ and average acceleration/braking $\bar{a}$ (Figure 2, left). The linear dependencies of the average speed $\bar{v}$ or the average acceleration/braking $\bar{a}$ with the average change in angle $\bar{\Delta}$ are rather small (see green lines in Figure 2, middle and right). From these plots we would guess that drivers 57 and 206 (blue and red colors) are similar, but driver 238 (orange color) seems to be different from the other two drivers.
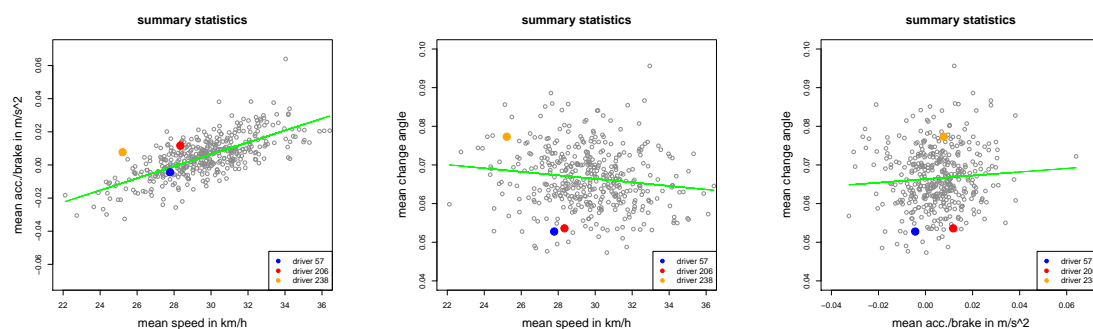

Figure 2. Summary statistics $\bar{v}$, $\bar{a}$ and $\bar{\Delta}$ of all $n = 416$ drivers.

The previous statistics are not of direct interest, but we are rather going to focus on the variations in these empirical means. We therefore consider the following average empirical standard deviations

$$\bar{\sigma}_v = \frac{1}{S} \sum_{s=1}^{S} \bar{\sigma}_{v,s} \qquad \text{and} \qquad \bar{\sigma}_{v,s} = \sqrt{\frac{1}{T-1} \sum_{t=1}^{T} \left(v_{s,t} - \bar{v}_s\right)^2},$$

$$\bar{\sigma}_a = \frac{1}{S} \sum_{s=1}^{S} \bar{\sigma}_{a,s} \qquad \text{and} \qquad \bar{\sigma}_{a,s} = \sqrt{\frac{1}{T-1} \sum_{t=1}^{T} \left(a_{s,t} - \bar{a}_s\right)^2}, \qquad (2)$$

$$\bar{\sigma}_\Delta = \frac{1}{S} \sum_{s=1}^{S} \bar{\sigma}_{\Delta,s} \qquad \text{and} \qquad \bar{\sigma}_{\Delta,s} = \sqrt{\frac{1}{T-1} \sum_{t=1}^{T} \left(\Delta_{s,t} - \bar{\Delta}_s\right)^2}.$$

These quantities determine the volatilities within the individual trips of the given drivers. In Figure 3 we plot these empirical standard deviations (2) against the empirical means (1) for all $n = 416$ drivers. We observe a strong linear relationship for the changes in angle, i.e., between $\bar{\Delta}$ and $\bar{\sigma}_\Delta$, see Figure 3 (right). For the other quantities, the relationship between the means (1) and

the variations (2) is less strong, and the empirical means $\overline{v}$ and $\overline{a}$ do not offer themselves as linear predictors for $\overline{\sigma}_v$ and $\overline{\sigma}_a$, respectively.
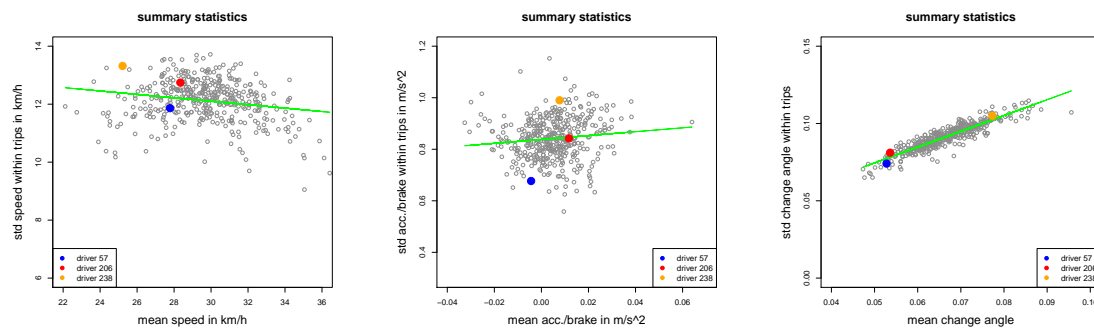

Figure 3. Empirical standard deviations (2) plotted against empirical means (1) for all $n = 416$ drivers.

Finally, in Figure 4 we provide the scatter plots of the average empirical standard deviations $\overline{\sigma}_v$, $\overline{\sigma}_a$ and $\overline{\sigma}_\Delta$ of the $n = 416$ drivers. We observe a strong linear relationship in particular between $\overline{\sigma}_v$ and $\overline{\sigma}_a$, and between $\overline{\sigma}_a$ and $\overline{\sigma}_\Delta$, respectively. In fact, in view of these last figures we claim that the drivers 57, 206 and 238 are clearly different. Remark that these three drivers were selected such that driver 57 has the 10th smallest value for $\overline{\sigma}_a$, driver 206 has the median value for $\overline{\sigma}_a$ and driver 238 has the 10th largest value for $\overline{\sigma}_a$ (among all considered drivers).
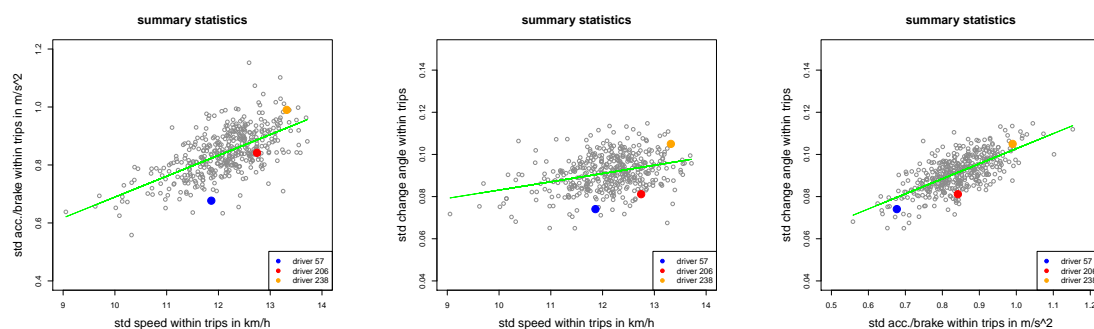

Figure 4. Summary statistics $\overline{\sigma}_v$, $\overline{\sigma}_a$ and $\overline{\sigma}_\Delta$ of all $n = 416$ drivers.

3. Classification of Individual Trips of Drivers 57, 206 and 238

One is tempted to use the summary statistics of the previous section as pricing information in car insurance, and in fact this is done in some insurance companies using the corresponding information from installed black box devices. The incentive of the present study is slightly different, namely, the goal is to classify individual trips. For instance, if we consider a randomly chosen trip of one of the three selected drivers 57, 206 and 238: can we allocate this trip to the right driver?

3.1. Remarks on the Chosen Trip Length and on the Selection of Three Drivers

Our goal is to allocate individual trips to the right drivers. In our analysis these individual trips have a total length of $T = 180$ s. This choice is a trade-off between the amount of available data and the minimal length of a trip. Each trip should have a minimal length to contain sufficient information to correctly perform our classification task. We use deep ConvNets for this classification task, and time series of 180 observations are the minimal length such that deep ConvNets may still work appropriately. Thus, ideally, we would prefer considering longer time series, but the choice of a total length of 180 s provides us with (only) 249 trips per driver on average. From these, say, 249 trips we use 80% for model calibration and 20% for out-of-sample back-testing of the fitted model. These are comparably low numbers, and increasing the length of the driving time would further lower these

numbers. Remark that this trip length of 180 s is very similar to the mean duration of micro-trips as stated on page 120 of Wang et al. (2008).

Initially, we have started our classification analysis with two selected drivers. The results of this initial analysis have been so convincing, that we (slightly) challenge this initial set up by choosing three drivers in the presented analysis. We will see that the results are still very good for three drivers.

3.2. Empirical Analysis of Drivers 57, 206 and 238

We start by an empirical analysis of the individual trips of the three selected drivers. In Figure 5 we plot the individual trip statistics $\overline{v}_s$, $\overline{a}_s$, $\overline{\Delta}_s$, $\overline{\sigma}_{v,s}$, $\overline{\sigma}_{a,s}$ and $\overline{\sigma}_{\Delta,s}$ for $s = 1, \ldots, S$ of the three selected drivers 57 (blue), 206 (red) and 238 (orange). We observe that the supports of the mean statistics $\overline{v}_s$, $\overline{a}_s$ and $\overline{\Delta}_s$ are very similar between the three drivers, and therefore these summary statistics do not seem to be helpful to classify individual trips. The biggest differences among the three drivers are observed between the empirical standard deviations $\overline{\sigma}_{a,s}$ of the acceleration/braking, see Figure 5 (bottom, middle). Remark that this variable has also been used to select the three (different) drivers, see final statement of Section 2.2.
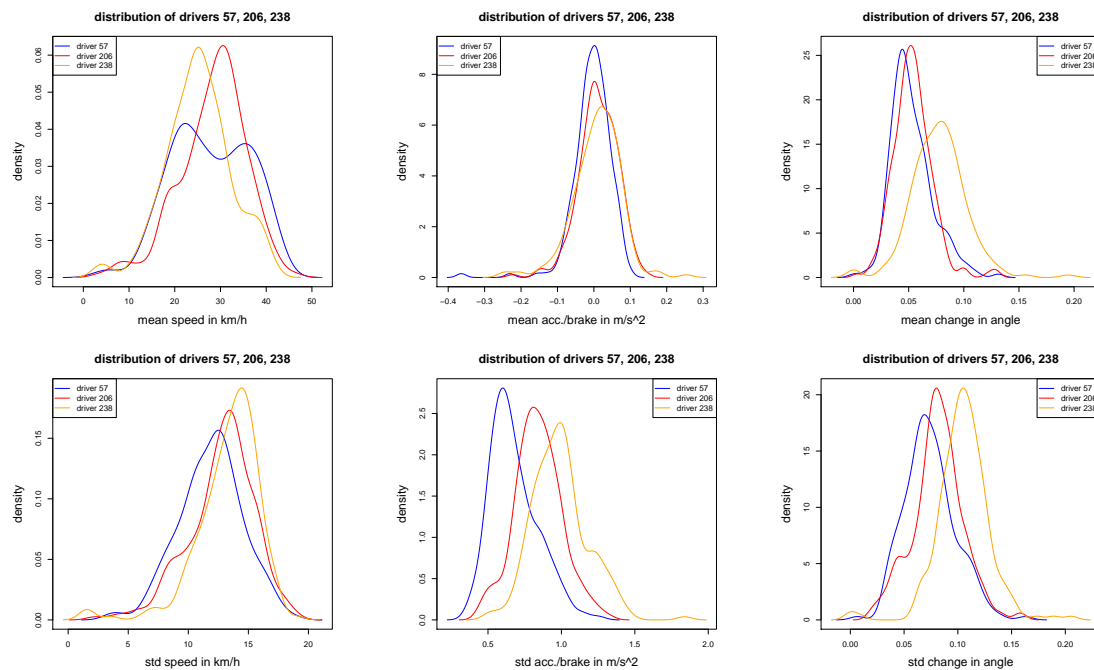

Figure 5. Densities of the individual trip statistics $\overline{v}_s$, $\overline{a}_s$ and $\overline{\Delta}_s$ (**first row**), and $\overline{\sigma}_{v,s}$, $\overline{\sigma}_{a,s}$ and $\overline{\sigma}_{\Delta,s}$ (**second row**) for $s = 1, \ldots, S$ of the three selected drivers 57, 206 and 238.

3.3. Convolutional Neural Networks Applied to Individual Trip Clustering

In this section, we use deep ConvNets to classify/allocate individual trips to the three selected drivers 57, 206 and 238. These individual trips $s = 1, \ldots, S$ are all the same length of 180 s, see Figure 1, and they have components in the following cuboid: $(v_{s,t}, a_{s,t}, \Delta_{s,t})' \in [2, 50]\,\text{km/h} \times [-3, 3]\,\text{m/s}^2 \times [0, 1/2]$. For neural network applications, we need to normalize these features to the same units. We therefore define the scaled features by

$$z_{s,t} = \left(\frac{2 v_{s,t}}{50} - 1, \ \frac{a_{s,t}}{3}, \ \frac{2 \Delta_{s,t}}{1/2} - 1 \right)' \ \in \ [-1, 1]^3.$$

Thus, an individual trip $s = 1, \ldots, S$ of a selected driver is characterized by $(z_{s,t})_{1 \le t \le T}$ for a total length of $T = 180$ s.

We use deep ConvNets to classify individual trips. Deep ConvNets have proved to be successful in image recognition because they can recognize similar structure at different locations. This ability is explained by the fact that deep ConvNets have the property of translation invariance (through using the same kernel at different locations) that allows them to detect these similar structures at different places in the images, see Wiatowski and Bölcskei (2018). We use a deep ConvNet having 3 convolutional layers. These 3 convolutional layers have 12, 10 and 8 filters, respectively. Between these convolutional layers we choose max-pooling layers of size 3, and the deep ConvNet is completed by a global-max-pooling layer and a dropout layer with dropout rate $p = 0.3$. The detailed architecture is provided in Listing 2.

Listing 2. R script for ConvNets in Keras.

```
1   model <- keras_model_sequential()
2   model %>%
3     layer_conv_1d(filters = 12, kernel_size = 5, activation='tanh',input_shape=c(180,3)) %>%
4     layer_max_pooling_1d(pool_size = 3) %>%
5     layer_conv_1d(filters = 10, kernel_size = 5, activation='tanh') %>%
6     layer_max_pooling_1d(pool_size = 3) %>%
7     layer_conv_1d(filters = 8, kernel_size = 5, activation='tanh') %>%
8     layer_global_max_pooling_1d() %>%
9     layer_dropout(rate = .3) %>%
10    layer_dense(units = 3, activation = 'softmax') %>%
11    compile(loss='categorical_crossentropy',optimizer=optimizer_rmsprop(),metrics='accuracy')
12  initializer_glorot_normal(seed=100)
13  model %>% fit(x.train, y.train, batch_size=nrow(x.train), epochs=300, validation_split=.1)
```

This ConvNet has 1237 parameters, for details see Listing 3. We have also tried various other ConvNet architectures, having 2 convolutional layers, having different numbersof filters, having different pooling layers, having other activation functions, etc. All these ConvNet architectures have had a very similar performance to the one chosen in Listing 2, thus, it seems that the fine-tuning of the explicit architecture is less important. There is only one parameter that has turned out to be very crucial, namely, the dropout rate of $p = 0.3$. Please note that the global-max-pooling layer reduces the last convolutional layer to 8 neurons (from the 8 filters) by discarding the time series structure, see lines 7–8 in Listing 2 and lines 13–14 in Listing 3. The dropout layer removes these 8 neurons independently from each other with a dropout probability $p = 0.3$. These random removals during training aim at guaranteeing that individual neurons cannot be over-trained to certain purposes. A lower dropout probability has led to over-fitting to the training data, and a higher dropout probability has led to a model that is not sufficiently sensitive to the data.

Listing 3. Structure of the deep ConvNet used.

```
1   Layer (type)                             Output Shape              Param #
2   =================================================================
3   conv1d_1 (Conv1D)                        (None, 176, 12)           192
4   -----------------------------------------------------------------
5   max_pooling1d_1 (MaxPooling1D)           (None, 58, 12)            0
6   -----------------------------------------------------------------
7   conv1d_2 (Conv1D)                        (None, 54, 10)            610
8   -----------------------------------------------------------------
9   max_pooling1d_2 (MaxPooling1D)           (None, 18, 10)            0
10  -----------------------------------------------------------------
11  conv1d_3 (Conv1D)                        (None, 14, 8)             408
12  -----------------------------------------------------------------
13  global_max_pooling1d_1 (GlobalMaxPooling1D)  (None, 8)             0
14  -----------------------------------------------------------------
15  dropout_1 (Dropout)                      (None, 8)                 0
16  -----------------------------------------------------------------
17  dense_1 (Dense)                          (None, 3)                 27
18  =================================================================
19  Total params: 1,237
20  Trainable params: 1,237
21  Non-trainable params: 0
```

The aim now is to fit this deep ConvNet to the three selected drivers. We therefore partition their data into a learning sample of 80% of the individual trips of each driver, and the remaining 20% of the trips are (only) used as a test set for out-of-sample back-testing of the fitted model. We choose a stratified partitioning for each driver where we use the average speeds $\overline{v}_s$ for stratifying. We then run the stochastic gradient descent algorithm on the learning sample using the `rmsprop` optimizer (root mean square propagation optimizer, see Section 8.5.2 in Goodfellow et al. (2016)). We run this algorithm for 300 epochs, and in Figure 6 we illustrate the fitting performance. The red dots show the training losses (based on 90% of the learning data), and the blue dots show the validation losses (based on 10% of the learning data). In the upper graph we see a decrease of the categorical cross-entropy loss (which is used as loss criterion in the gradient descent algorithm). The lower graph shows the resulting correct classification rates on training and validation data. From the validation data graphs, we also observe that after 300 epochs this ConvNet is sufficiently trained, and further training may lead to over-fitting to the learning data.
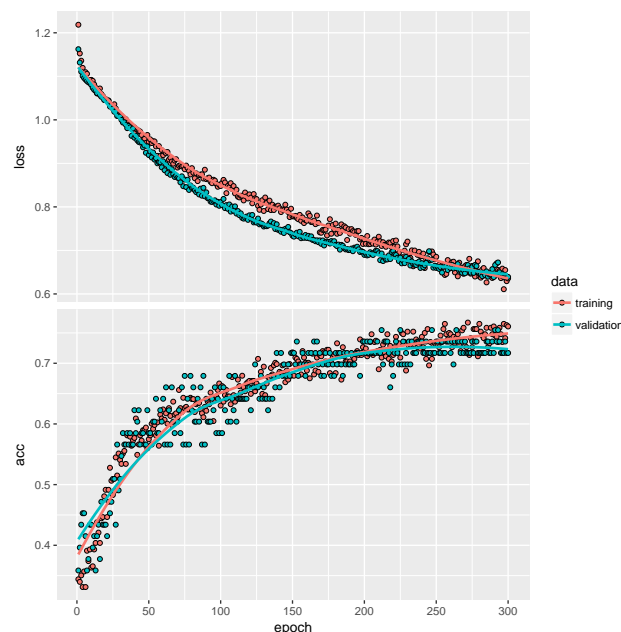

Figure 6. Stochastic gradient descent fitting of the ConvNet given in Listing 2 to the individual trips of the drivers 57, 206 and 238: the upper graph shows the categorical cross-entropy losses; the lower graph shows the correct classification rates for training (red) and validation (blue) data.

We then evaluate this model on the test data (which has not yet been seen by the model and the calibration algorithm). The corresponding results are presented in Table 1. We observe that more than 75% of the trips are correctly classified (allocated to the right drivers). This is very remarkable, because these results are far better than pure random allocation which would get 33.3% of the trips right! Moreover, we emphasize that this classification has been done on (only) 3 min of driving experience.

Table 1. Out-of-sample analysis of drivers 57, 206 and 238.

Variable	Driver 57	Driver 206	Driver 238
Total number of trips S	205	234	213
Size learning sample	163	187	171
Size test sample	42	47	42
Correctly classified test samples (in %)	78.6%	70.2%	85.7%
Misclassified test samples	9	14	6

3.4. Analysis of Misclassified Trips

In this section, we analyze the misclassified trips of the out-of-sample analysis given in Table 1. In Table 2 we give the resulting confusion matrix that shows the misclassified trips of the test sample (we provide absolute numbers because the sizes of the test samples are very similar between the three selected drivers). The trips of the middle driver 206 (in terms of $\overline{\sigma}_a$) seem to be more difficult than the ones of the other two drivers, and his misclassified trips are equally allocated to drivers 57 and 238. This may illustrate that his individual standard deviations of acceleration/braking $\overline{\sigma}_{a,s}$ are sandwiched between the ones of the other two drivers, see Figure 5 (2nd row, middle column). This motivates to analyze the characteristics (1) and (2) of the misclassified trips.

Table 2. Confusion matrix of the out-of-sample analysis of Table 1.

	True Labels		
	Driver 57	**Driver 206**	**Driver 238**
Predicted label 57	33	7	3
Predicted label 206	7	33	3
Predicted label 238	2	7	36

We start by considering the middle driver 206 (in red color) which seems to be the most difficult one of the three. In Figure 7 we illustrate all 47 test samples of driver 206: the 33 correctly classified with red dots, the 14 misclassified with blue and orange dots (according to the given wrong labels). From these plots we see that misclassification cannot be explained by the (simple) statistics of sample means and standard deviations $\overline{v}_s$, $\overline{a}_s$, $\overline{\Delta}_s$, $\overline{\sigma}_{v,s}$, $\overline{\sigma}_{a,s}$ and $\overline{\sigma}_{\Delta,s}$, respectively, because the misclassified trips do not seem to have a particular structure in these plots. The most explanatory factors may still be $\overline{\sigma}_{a,s}$ and $\overline{\sigma}_{\Delta,s}$ because misclassified trips tend to slightly cluster for these factors. The first conclusion from this is that the deep ConvNet really classifies beyond simple sample statistics by looking in more detail into the time series and structure of speed, acceleration/braking, and changes in angle of individual trips.
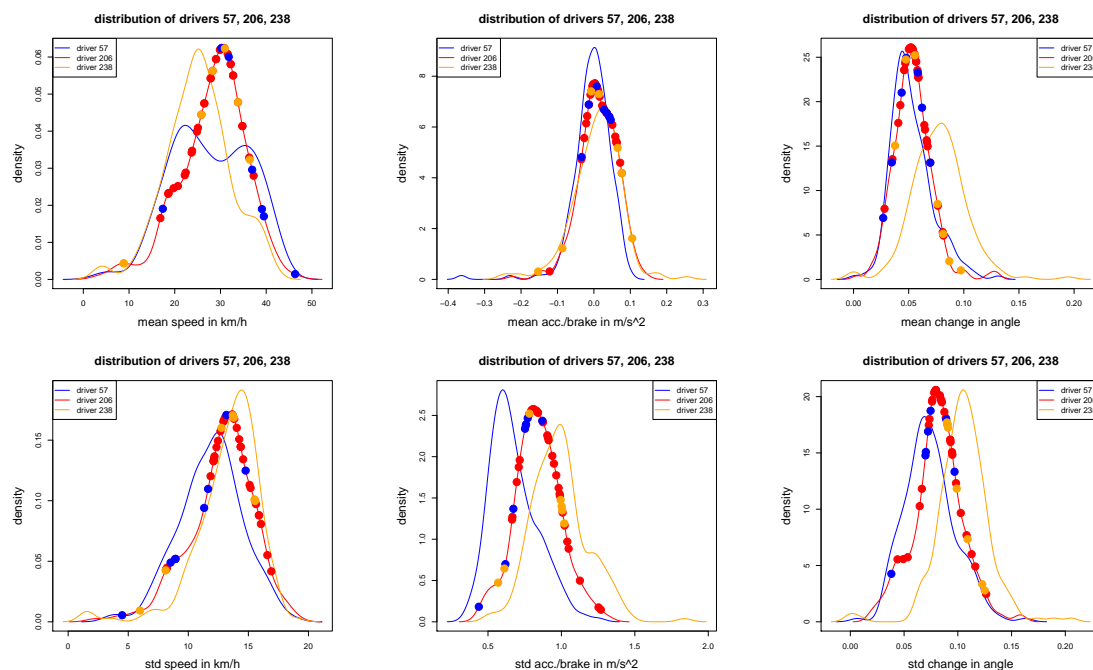

Figure 7. These figures show the same densities as in Figure 5: the dots illustrate the 47 test samples of driver 206: red dots show correctly classified trips, blue and orange dots show misclassified trips (according to the given wrong labels).

In Figure 8 we illustrate the 33 correctly classified trips (blue dots) and the 9 misclassified trips (red and orange dots) of driver 57, and in Figure 9 the 36 correctly classified trips (orange dots) and the 6 misclassified trips (blue and red dots) of driver 238. The conclusion remains the same, namely, that $\overline{\sigma}_{a,s}$ and $\overline{\sigma}_{\Delta,s}$ are probably the most important factors, but these factors solely cannot fully explain the classification and misclassification, respectively, i.e., the deep ConvNet uses other and more structure of the time series to perform the clustering.

Figure 8. Out-of-sample analysis of the 42 test samples of driver 57: correctly classified trips (blue dots) and misclassified trips (red and orange dots according to the given labels) of driver 57, illustrated in the density plots of $\overline{\sigma}_{v,s}$, $\overline{\sigma}_{a,s}$ and $\overline{\sigma}_{\Delta,s}$ for $s = 1, \ldots, S$.

Figure 9. Out-of-sample analysis of the 42 test samples of driver 238: correctly classified trips (orange dots) and misclassified trips (blue and red dots according to the given labels) of driver 238, illustrated in the density plots of $\overline{\sigma}_{v,s}$, $\overline{\sigma}_{a,s}$ and $\overline{\sigma}_{\Delta,s}$ for $s = 1, \ldots, S$.

4. Classification of Trips of Other Triples of Drivers

4.1. Individual Trips of Drivers 329, 174 and 227

In a second analysis we choose three drivers that have a similar behavior in the acceleration/braking pattern $\overline{\sigma}_{a,s}$, but have different behaviors in the change in angle statistics $\overline{\sigma}_{\Delta,s}$. The three selected drivers 329, 174 and 227 are illustrated in Figure 10. We observe that these three drivers have rather similar acceleration/braking statistics $\overline{\sigma}_{a,s}$, but they differ in their change in angle driving styles $\overline{\sigma}_{\Delta,s}$.

We then fit exactly the same deep ConvNet architecture as in the previous section to a stratified learning sample of 80% of the individual trips. The out-of-sample results on the test samples are presented in Table 3. Also, in this case we receive a remarkably good out-of-sample performance, and the deep ConvNet can distinguish the trips of these three drivers with an accuracy of roughly 73% (and though they have similar acceleration/braking statistics).
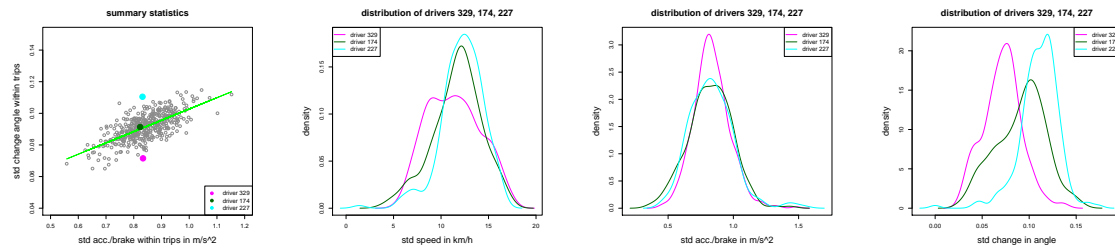
Figure 10. (**left**) Scatter plot $(\overline{\sigma}_a, \overline{\sigma}_\Delta)$ of all $n = 416$ drivers with drivers 329, 174 and 227 in magenta, green and cyan colors; (**middle** and **right**) density plots of $\overline{\sigma}_{v,s}$, $\overline{\sigma}_{a,s}$ and $\overline{\sigma}_{\Delta,s}$ for $s = 1, \ldots, S$ of the three drivers 329, 174 and 227.

Table 3. Out-of-sample analysis of drivers 329, 174 and 227.

Variable	Driver 329	Driver 174	Driver 227
Total number of trips S	446	232	214
Size learning sample	356	186	171
Size test sample	90	46	43
Correctly classified test samples (in %)	86.7%	54.3%	76.7%
Misclassified test samples	12	21	10

4.2. Individual Trips of Drivers 63, 30 and 163

In a third analysis we choose three drivers that have a similar behavior in all summary statistics. The three selected drivers 63, 30 and 163 are illustrated in Figure 11. We observe that these three drivers have rather similar density functions in all summary statistics of the individual trips.
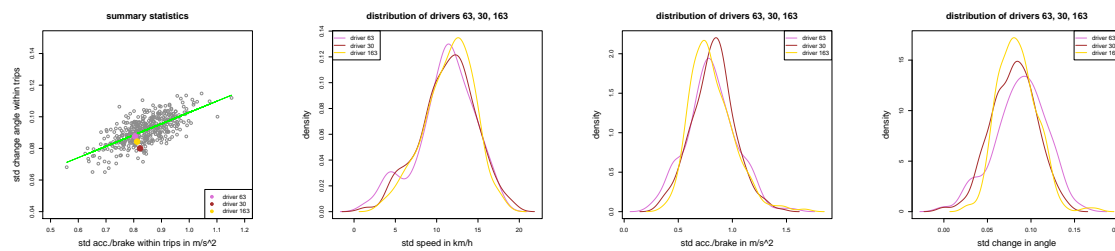

Figure 11. (**left**) Scatter plot $(\overline{\sigma}_a, \overline{\sigma}_\Delta)$ of all $n = 416$ drivers with drivers 63, 30 and 163 in orchid, brown and yellow colors; (**middle** and **right**) density plots of $\overline{\sigma}_{v,s}$, $\overline{\sigma}_{a,s}$ and $\overline{\sigma}_{\Delta,s}$ for $s = 1, \ldots, S$ of the three drivers 63, 30 and 163.

We then fit exactly the same deep ConvNet architecture as in the previous sections. The out-of-sample results are presented in Table 4. We observe that the deep ConvNet has much more difficulty to distinguish the trips of these three drivers. Nevertheless, the performance of this deep ConvNet is still better than random guessing with an out-of-sample accuracy of roughly 47%.

Table 4. Out-of-sample analysis of drivers 63, 30 and 163.

Variable	Driver 63	Driver 30	Driver 163
Total number of trips S	194	259	203
Size learning sample	155	207	163
Size test sample	39	52	40
Correctly classified test samples (in %)	35.9%	59.6%	45.0%
Misclassified test samples	25	21	22

4.3. Individual Trips of Drivers 349, 249 and 288

In our fourth analysis we aim to compare three drivers that have a similar behavior in speeds $\overline{\sigma}_{v,s}$ and changes in angles $\overline{\sigma}_{\Delta,s}$, but differ in the acceleration/braking pattern $\overline{\sigma}_{a,s}$. Please note that it is

more difficult to find such drivers because of the positive linear relationships illustrated in Figure 4 (left and right). We select the three drivers 349, 249 and 288, illustrated in Figure 12.
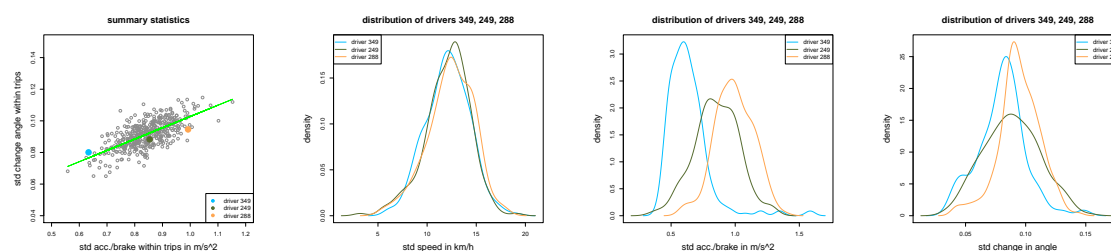

Figure 12. (**left**) Scatter plot $(\overline{\sigma}_a, \overline{\sigma}_\Delta)$ of all $n = 416$ drivers with drivers 349, 249 and 288 in light blue, dark green and orange colors; (**middle** and **right**) density plots of $\overline{\sigma}_{v,s}$, $\overline{\sigma}_{a,s}$ and $\overline{\sigma}_{\Delta,s}$ for $s = 1, \ldots, S$ of the three drivers 349, 249 and 288.

We then fit exactly the same deep ConvNet architecture as in the previous sections. The out-of-sample results are presented in Table 5. We observe that the deep ConvNet classification performs very well, and more than 75% of the trips are correctly classified (out-of-sample).

Table 5. Out-of-sample analysis of drivers 249, 349 and 288.

Variable	Driver 349	Driver 249	Driver 288
Total number of trips S	244	255	193
Size learning sample	195	203	155
Size test sample	49	52	38
Correctly classified test samples (in %)	83.7%	67.3%	78.9%
Misclassified test samples	8	17	8

5. Classification of The Trips of All Drivers

A major interest from an actuarial point of view is to extract feature information from telematics car driving data for car insurance pricing. Based on our previous analysis, insurers are tempted to select "archetypal" drivers based on propensity to claim, and then allocate other drivers (or their trips, respectively) to these archetypal drivers. This approach makes the analysis of telematics car driving data feasible (scalable) because the network only needs to be trained on a sub-sample of representative drivers, and the remaining drivers are allocated according to the resulting classification.

5.1. Convolutional Neural Network Approach

In Section 3 we have classified the trips of the three drivers 57, 206 and 238. These three drivers are rather different and we have calibrated a deep ConvNet that manages to achieve an out-of-sample accuracy of more than 75%, see the confusion matrix in Table 2. We may now interpret these three drivers as our archetypal drivers to classify all other drivers according to the deep ConvNet found in Section 3. Please note that in a real insurance context, the choice of the archetypal drivers should be based on propensity to claim (which unfortunately is not available for our data).

In Table 6 we provide the allocation of all 103,734 individual trips of all $n = 416$ drivers to the three selected drivers 57, 206 and 238 (using the fitted deep ConvNet found in Section 3). We observe a fairly equal distribution of these trips to the selected drivers.

Table 6. Allocation of all trips to the three drivers 57, 206 and 238, and allocation of the drivers to the three selected drivers according to the likelihood of the individual trip allocation.

Variable	Driver 57	Driver 206	Driver 238
Allocation of all 103,734 trips	35.0%	35.2%	29.7%
Allocation of the 416 drivers	29.8%	45.0%	25.2%

In Figure 13 we plot the relative allocation of the trips of each driver $i = 1, \ldots, 416$ to the three selected drivers 57 (blue), 206 (red) and 238 (orange). On the left-hand side, the drivers $i = 1, \ldots, 416$ are ordered according to their empirical standard deviations $\overline{\sigma}_v$ of the speeds, in the middle w.r.t. their acceleration/braking behavior $\overline{\sigma}_a$ and on the right-hand side according to their change in angle behavior $\overline{\sigma}_\Delta$. We observe that the main trigger for the allocation is their acceleration/braking behavior, the change in angle is slightly less important in this allocation, and the volatility of speed only shows a low interaction.

Figure 13. Relative allocation of the individual trips of drivers $i = 1, \ldots, 416$ to drivers 57 (blue), 206 (red) and 238 (orange): individual drivers are ordered w.r.t. $\overline{\sigma}_v$ (**left**), $\overline{\sigma}_a$ (**middle**) and $\overline{\sigma}_\Delta$ (**right**).

Moreover, we can also classify the drivers $i = 1, \ldots, 416$ to the cluster for which he has the highest number of allocations among his individual trips. This classification is provided on the last line of Table 6. It shows that the middle driver attracts 45% of all drivers (which is slightly more than the proportion of individual trips he attracts).

5.2. Principal Component Analysis of Telematics Heatmaps: Revisited

In this section, we revisit the telematics speed-acceleration (*v-a*) heatmaps studied in Gao and Wüthrich (2018) which correspond to the normalized speed-acceleration matrices considered in Section 3.3 of Kamble et al. (2009). Since these *v-a* heatmaps should not reflect whether we consider a city driver or a highway driver, we restrict our considerations in this section to the speed bucket $[5, 20]$ km/h; we also refer to Figure 5 in Hung et al. (2007) for different driving patterns for different types (and speeds) of trips. Thus, we consider the *v-a* rectangle $[5, 20]$ km/h $\times [-3, 3]$ m/s². This *v-a* rectangle is partitioned into $d \in \mathbb{N}$ congruent sub-rectangles $(R_j)_{1 \leq j \leq d}$. The *v-a* heatmap $\boldsymbol{x}_i = (x_{i,j})_{1 \leq j \leq d}$ of driver $i = 1, \ldots, n$ is then received by calculating the relative amount of time this driver spends in sub-rectangles $(R_j)_{1 \leq j \leq d}$ over all his individual trips. Thus, every driver i is characterized by a discrete probability distribution $\boldsymbol{x}_i$ on the $(d-1)$-unit simplex in $[0, 1]^d$, for details we refer to Section 3 of Gao and Wüthrich (2018).

In Figure 14 we present the *v-a* heatmaps $\boldsymbol{x}_i$ of the three selected drivers $i = 57, 206, 238$. We observe similarities in the general structure of these heatmaps; however, the level sets are slightly different between these three drivers which indicates that they have different driving styles. For instance, the width of the level sets on the *y*-axis indicates stronger acceleration and braking.
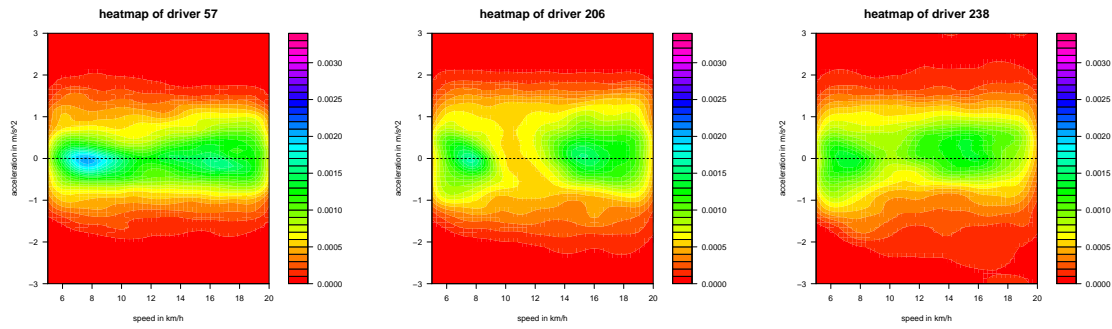
Figure 14. *v-a* heatmaps x_i of the three selected drivers $i = 57, 206, 238$.

In Gao and Wüthrich (2018), these *v-a* heatmaps have been analyzed by using a principal component analysis (PCA) over the *v-a* heatmaps of all drivers. We therefore denote the normalized design matrix received over all drivers $(x_i)_{1 \le i \le n}$ by $X \in \mathbb{R}^{n \times d}$. Using a singular value decomposition (SVD) we decompose this design matrix into

$$X = U \Lambda V',$$

where $U \in \mathbb{R}^{n \times d}$ is orthogonal ($U'U = \mathbb{1}$), $V \in \mathbb{R}^{d \times d}$ is orthogonal ($V'V = \mathbb{1}$), and $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_d)$ is a diagonal matrix containing the singular values $\lambda_1 \ge \ldots \ge \lambda_d \ge 0$. The principal components are then obtained by multiplying X from the right with V, that is, the principal components are given by the columns of the matrix

$$XV = U\Lambda = U\text{diag}(\lambda_1, \ldots, \lambda_d) \in \mathbb{R}^{n \times d}. \tag{3}$$

Note from (3) that the *k*-th principal component of driver *i* is obtained by multiplying his normalized *v-a* heatmap with the *k*-th column of *V* (the so-called *k*-th right-singular vector of *X*). In Figure 15 we present the first three right-singular vectors of *X* given by the first three columns of the orthogonal matrix *V*; these correspond to the biggest three singular values $\lambda_1 \ge \lambda_2 \ge \lambda_3$. The figure illustrates that the 1st right-singular vector mainly measures the *y*-width of the level sets and, henceforth, studies the magnitude of acceleration and braking. The 2nd right-singular vector analyses the differences between low and high speeds in speed bucket $[5, 20]$ km/h, and the 3rd right-singular vector measures the differences between acceleration and braking. These properties are used below to compare the PCA results to the deep ConvNet classification results.
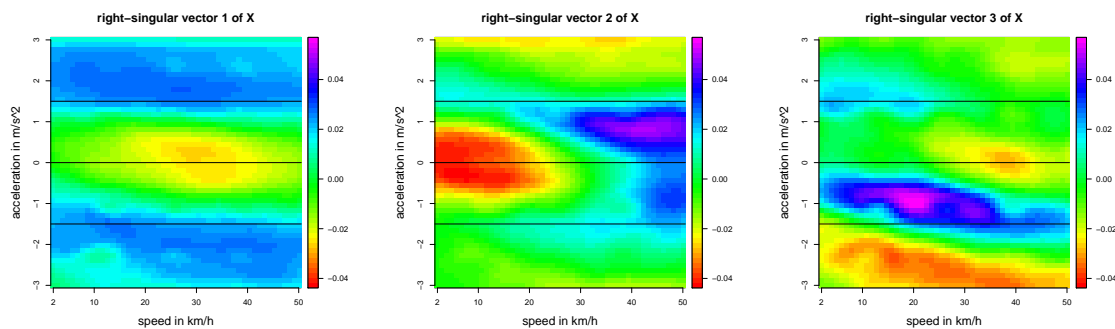

Figure 15. First three right-singular vectors of *X* given by the first three columns of the orthogonal matrix *V*.

Finally, in Figure 16 we provide the first three principal components of all drivers, received by the first three columns of the matrix XV in (3). The drivers are colored according to the allocation to the three selected drivers 57, 206, 238 received from the deep ConvNet, see Table 6. The three selected drivers are illustrated by slightly larger dots (with black borders). The first two principal components

(Figure 16, left) show a separation of blue dots from red and orange ones; the red and orange dots however are quite mixed (in fact, the first principal component is the main driver of the separation). Thus, a PCA based on the first two principal components leads to a slightly different clustering compared to the deep ConvNet approach (using drivers 57, 206 and 238 for determining the clustering). On the one hand, it is quite surprising that we receive similarities between the principal components and the blue-red-orange dots clustering of the ConvNet approach because the three drivers 57, 206 and 238 have just been selected because they have rather different sample statistics $\overline{\sigma}_a$, see Figure 4 (right). Moreover, the *v-a* heatmaps only look at speed and acceleration/braking, but neither do they consider these quantities as time series nor do they consider the changes in angles, as the deep ConvNet approach does. On the other hand, similarities can be explained by the fact, that the first principal component in the PCA explains acceleration/braking, see Figure 15 (left), and this has also been the crucial variable used to select the three drivers 57, 206 and 238.
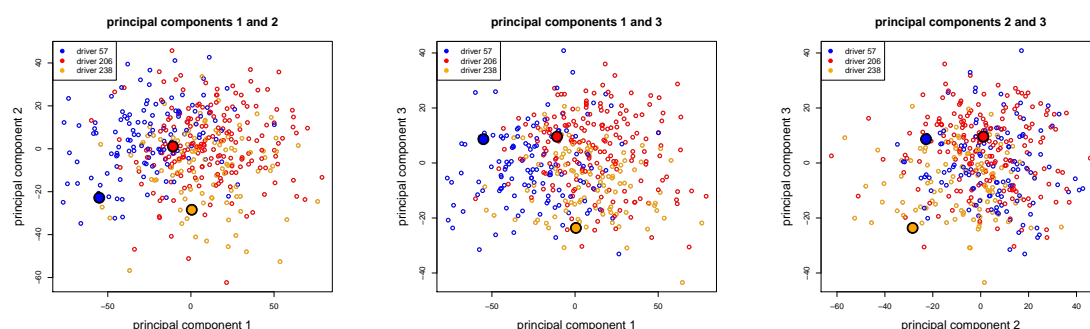

Figure 16. The first three principal components of the $n = 416$ drivers illustrated in two-dimensional scatter plots, colored according to the ConvNet classification of Table 6, the three selected drivers 57, 206 and 238 are illustrated by bigger dots.

6. Conclusions and Outlook

We have studied individual trips of different car drivers by considering their time series structures of speeds, acceleration/braking, and changes of angles. We have trained a deep ConvNet to classify individual trips to three selected car drivers. The findings are that deep ConvNets manage to perform this classification task very successfully, and that the classification goes beyond simple summary statistics. It is worth mentioning the following points:

- In our analysis we have been restricted by the fact that on average we (only) have 249 trips of length 180 s per driver. If more data is available, one could explore the minimal driving time required to receive reliable classification results.
- Remark that if a network is calibrated sufficiently well then it is not necessary to update the calibration on new data (unless the driving styles of the corresponding drivers change). Thus, this method is scalable because the network needs only be trained once (sufficiently well).
- We have performed our analysis on three drivers, increasing the number of drivers will decrease the performance of classification, or differently speaking, it will require more data to receive similar classification results.
- We did not have any information about road condition and car type. It would be interesting to see how this additional information changes our findings.
- It would be interesting to see whether this classification method would also be helpful to analyze the micro-trip concatenation in synthetic driving cycle construction, see Hung et al. (2007), for instance, it may help to decide whether a synthesized driving cycle is representative and indistinguishable from a real driving cycle.
- We have used deep ConvNets that have been rather successful in our analysis. A limitation of ConvNets is that the input of all observations needs to have the same dimension. Recurrent neural networks and long short-term memory networks are more flexible in this regard and one may explore these networks.

From an insurance and actuarial point of view, we are mainly interested to separate "bad" from "good" drivers. In a next step one may analyze which drivers and trips cause claims more likely. Therefore, deep ConvNets could be trained to classify such different types of drivers and trips, respectively. This then allows insurance companies to label individual trips according to their hazard level, similarly to the analysis in Section 5. This labeling can be used as feature and risk driver information in insurance pricing. Please note that this individual trip labeling also possesses the advantage that it can deal with the situation that the same car may be driven by different people, or that the same person drives in different conditions.

Author Contributions: The authors contributed equally to this work.

Conflicts of Interest: The authors declare no conflict of interest.

References

Ayuso, Mercedes, Montserrat Guillen, and Ana María Pérez-Marín. 2016a. Telematics and gender discrimination: Some usage-based evidence on whether men's risk of accidents differs from women's. *Risks* 4: 10. [CrossRef]

Ayuso, Mercedes, Montserrat Guillen, and Ana María Pérez-Marín. 2016b. Using GPS data to analyse the distance traveled to the first accident at fault in pay-as-you-drive insurance. *Transportation Research Part C: Emerging Technologies* 68: 160–67. [CrossRef]

Boucher, Jean-Philippe, Steven Côté, and Montserrat Guillen 2017. Exposure as duration and distance in telematics motor insurance using generalized additive models. *Risks* 5: 54. [CrossRef]

Esteves-Booth, A., Tariq Muneer, Howard Kirby, Jorge Kubie, and J. Hunter. 2001. The measurement of vehicular driving cycle within the city of Edinburgh. *Transportation Research Part D: Transport and Environment* 6: 209–20. [CrossRef]

Gao, Guangyuan, and Mario V. Wüthrich. 2018. Feature extraction from telematics car driving heatmaps. *European Actuarial Journal* 8: 383–406. [CrossRef]

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. Cambridge: MIT Press.

Ho, Sze-Hwee, Yiik-Diew Wong, and Victor Wei-Chung Chang. 2014. Developing Singapore driving cycle for passenger cars to estimate fuel consumption and vehicular emissions. *Atmospheric Environment* 97: 353–62. [CrossRef]

Hung, Wing Tat, Hingyan Tong, Chipang Lee, K. Ha, and L. Y. Pao. 2007. Development of practical driving cycle construction methodology: A case study in Hong Kong. *Transportation Research Part D: Transport and Environment* 12: 115–28. [CrossRef]

Kamble, Sanghpriya H., Tom V. Mathew, and Gaurav K. Sharma. 2009. Development of real-world driving cycle: Case study of Pune, India. *Transportation Research Part D: Transport and Environment* 14: 132–40. [CrossRef]

Lemaire, Jean, Sojung Carol Park, and Kili C. Wang. 2016. The use of annual mileage as a rating variable. *ASTIN Bulletin* 46: 39–69. [CrossRef]

Paefgen, Johannes, Thorsten Staake, and Elgar Fleisch. 2014. Multivariate exposure modeling of accident risk: Insights from pay-as-you-drive insurance data. *Transportation Research Part A: Policy and Practice* 61: 27–40. [CrossRef]

Verbelen, Roel, Katrien Antonio, and Gerda Claeskens. 2018. Unraveling the predictive power of telematics data in car insurance pricing. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 67: 1275–304. [CrossRef]

Wang, Qidong, Hong Huo, Kebin He, Zhiliang Yao, and Qiang Zhang. 2008. Characterization of vehicle driving patterns and development of driving cycles in Chinese cities. *Transportation Research Part D: Transport and Environment* 13: 289–97. [CrossRef]

Weidner, Wiltrud, Fabian W. G. Transchel, and Robert Weidner. 2016. Classification of scale-sensitive telematic observables for riskindividual pricing. *European Actuarial Journal* 6: 3–24. [CrossRef]

Weidner, Wiltrud, Fabian W. G. Transchel, and Robert Weidner. 2017. Telematic driving profile classification in car insurance pricing. *Annals of Actuarial Science* 11: 213–36. [CrossRef]

Wiatowski, Thomas, and Helmut Bölcskei. 2018. A mathematical theory of deep convolutional neural networks for feature extraction. *IEEE Transactions on Information Theory* 64: 1845–66. [CrossRef]

Zhang, Wei, Jun Tanida, Kazuyoshi Itoh, and Yoshiki Ichioka. 1988. Shift invariant pattern recognition neural network and its optical architecture. *Proceedings of the Annual Conference of the Japan Society of Applied Physics* 6p-M-14: 734.

Zhang, Wei, Kazuyoshi Itoh, Jun Tanida, and Yoshiki Ichioka. 1990. Parallel distributed processing model with local space-invariant interconnections and its optical architecture. *Applied Optics* 29: 4790–97. [CrossRef] [PubMed]