



City Research Online

City St George's, University of London

Citation: Zhu, R. & Wüthrich, M. V. (2021). Clustering driving styles via image processing. *Annals of Actuarial Science*, 15(2), pp. 276-290. doi: 10.1017/s1748499520000317

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/24969/>

Link to published version: <https://doi.org/10.1017/s1748499520000317>

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to [City Research Online policy](#).

Clustering driving styles via image processing

Rui Zhu^a, Mario V. Wüthrich^b

^a*Faculty of Actuarial Science and Insurance, Cass Business School, City, University of London, London EC1Y 8TZ, UK*

^b*ETH Zurich, RiskLab, Department of Mathematics, 8092 Zurich, Switzerland*

Abstract

It has become of key interest in the insurance industry to understand and extract information from telematics car driving data. Telematics car driving data of individual car drivers can be summarized in so-called speed-acceleration heatmaps. The aim of this study is to cluster such speed-acceleration heatmaps to different categories by analysing similarities and differences in these heatmaps. Making use of local smoothness properties, we propose to process these heatmaps as RGB images. Clustering can then be achieved by involving supervised information via a transfer learning approach using the pre-trained AlexNet to extract discriminative features. The K -means algorithm is then applied on these extracted discriminative features for clustering. The experiment results in an improvement of heatmap clustering compared to classical approaches.

Keywords: Telematics car driving data, driving styles, unsupervised learning, image processing, transfer learning, AlexNet

Email addresses: rui.zhu@city.ac.uk (Rui Zhu), mario.wuethrich@math.ethz.ch (Mario V. Wüthrich)

1. Introduction

Nowadays, telematics car driving data becomes vital to general insurance companies. Classical car insurance pricing is typically based on generalised linear models using covariate information like age of driver, gender of driver, type of car, price of car, power of engine, etc. This conventional covariate information is not directly related to driving styles and driving habits, but it is rather brought in as proxy information for missing information about driving styles and skills. Of course, this raises some issues because these proxies only describe typical representatives of covariate characteristics, and an individual driver might be quite different from a typical driver. Moreover, recently concerns have been raised about discrimination as certain protected variables are not allowed to serve as proxies, for instance, gender under European law is not allowed to be used as an explanatory variable in regression models ([European Commission, 2012](#)). In contrast, telematics car driving data is much closer to the ground truth of driving style and driving skills because it continuously registers driving behaviour and manoeuvres.

However, telematics car driving data poses big challenges itself, one being the massive amount of data that it creates and another one being the accuracy telematics data typically has. For these reasons, there is a vastly growing literature on telematics data that aims at making it useful for understanding and pricing car insurance policies. Needless to say that new car insurance products should also aim at improving driving styles by continuously giving feedback to the customers about their driving. We briefly review recent developments on telematics car driving data.

Some studies aim to identify indicators of driving risk which can help insurers to obtain better risk profiles for individual car drivers. Driving

27 distance is one factor that has been widely explored (Lemaire et al., 2016;
28 Boucher et al., 2017; Verbelen et al., 2018), other methods aim at evaluat-
29 ing driving risk based on extracting behavior variables from usage-based-
30 insurance (UBI) data that goes beyond driving distance (Bian et al., 2018;
31 Ayuso et al., 2016a,b; Denuit et al., 2019). Carfora et al. (2019) propose
32 an indicator of driver aggressiveness based on cluster analysis results. More
33 recently, generalised linear models are built based on the internet of vehicles
34 (IoV) data to identify risky drivers, see Sun et al. (2020). Another direc-
35 tion of research is to study driving cycles which are usually represented by
36 speed-time profiles. By studying such driving patterns in different cities,
37 one can evaluate energy and emissions in road transportation (Hung et al.,
38 2007; Kamble et al., 2009; Ho et al., 2014).

39 Since telematics car driving data and, in particular, GPS location data
40 second by second results in a massive amount of data, this data needs to be
41 compressed or summarized in a suitable way to make it useful for insurance
42 pricing. Of course, this aggregation should be done at a minimal loss of
43 information. One way of aggregation is to build so-called speed-acceleration
44 (v - a) heatmaps which is a two-dimensional summary statistics of a speed
45 versus acceleration pattern, see Wüthrich (2017). This approach can reduce
46 the large amount of telematics data while keeping key information of indi-
47 vidual driving patterns. The corresponding v - a heatmap is generated from
48 the telematics data for each individual driver. Figure 1 shows two examples
49 of v - a heatmaps in the (5, 20] (km/h) speed interval. The x axis shows speed
50 v in km/h while the y axis shows acceleration a in m/s^2 for an individual
51 driver. The v - a heatmap then gives the distribution of the time spent by
52 a driver at each (v, a) location. From Figure 1 it is obvious that the two
53 illustrated drivers have quite different speed-acceleration behaviour.

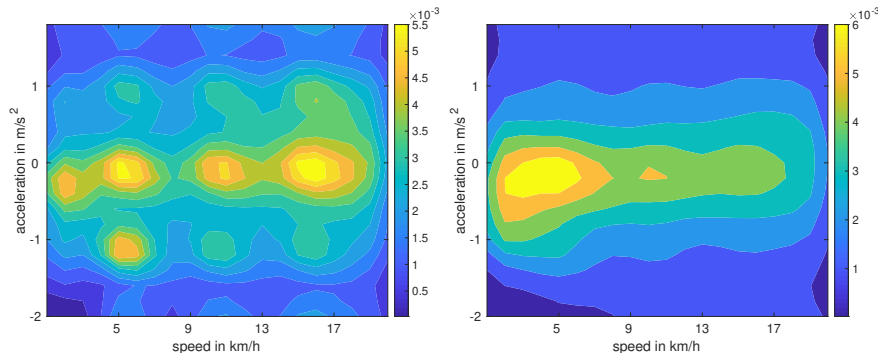


Figure 1: Two examples of v - a heatmaps.

54 Our goal here is to analyze different driving patterns based on these v - a
55 heatmaps. One direction is to study whether there are clusters of similar
56 heatmaps, so that we can cluster customers to different categories of driv-
57 ing styles. Given that the heatmaps are not labelled, this provides us with
58 a cluster analysis problem (Section 10.3 in [James et al. 2013](#)). [Wüthrich](#)
59 [\(2017\)](#) proposes to explore this direction by K -means clustering, that di-
60 vides data to K non-overlapping subgroups, and it is assumed that data
61 points within each subgroup are similar to each other. Thus, the car drivers
62 that are clustered to one subgroup by the K -means algorithm are believed to
63 share a similar driving style. In a further study, [Gao and Wüthrich \(2018\)](#)
64 extract low-dimensional features from v - a heatmaps that can be used in
65 regression models for car insurance pricing. Of course, at this stage, it is
66 not clear whether such a clustering provides any predictive power for car
67 frequency prediction. [Gao et al. \(2019\)](#) provide evidence on a small data
68 set that, indeed, clustering of v - a heatmaps can extract feature informa-
69 tion from telematics car driving data that has predictive power for claim
70 frequency prediction. However, their analysis is based on less than 2000
71 drivers, therefore, bigger portfolios and more analysis is needed to receive

72 more support for this approach. Weidner et al. (2017) also cluster driving
73 styles to evaluate driving behaviour. Different from the above approaches,
74 their study uses a hierarchical clustering method based on three variables,
75 vehicle velocity, acceleration and deceleration.

76 We note that there are two aspects that can be improved in the above
77 approaches. First, from the v - a heatmaps in Figure 1, we can observe
78 that within a small local area the values in each heatmap are close to each
79 other, which suggests a smoothness property or a spatial structure that can
80 be explored in the heatmap. This spatial structure has not been consid-
81 ered in Wüthrich (2017) and Gao and Wüthrich (2018), because the entire
82 heatmap has been stacked into a one-dimensional vector in these two studies.
83 Considering this spatial property may improve the clustering results. Sec-
84 ond, all heatmaps are unlabelled suggesting that this is a difficult clustering
85 task. Involving supervised information from other classification problems
86 may improve the clustering results.

87 In this paper, we propose to enhance the above two aspects via transfer
88 learning with the pre-trained AlexNet on heatmap images to extract dis-
89 criminative features that can bring supervised information to our clustering
90 task. First, we propose to process heatmaps as two-dimensional RGB images
91 rather than treating them as one-dimensional vectors to preserve the local
92 geometry. Machine learning algorithms in image processing have been well
93 developed by considering the local smoothness property of images. Thus,
94 our task becomes to cluster the heatmap RGB images rather than the one-
95 dimensional vectors of Wüthrich (2017). Second, the pre-trained models
96 in image classification tasks can be utilised to bring supervised informa-
97 tion to our clustering task. Here, we select the AlexNet model (Krizhevsky
98 et al., 2012) that is trained on the ImageNet database. From the pre-trained

99 AlexNet, we can extract discriminative features from the heatmaps that are
100 informative to distinguish between different image classes. More specifi-
101 cally, we feed the heatmap images to the pre-trained AlexNet and extract
102 discriminative features that can distinguish between different heatmap pat-
103 terns. These features are then used in the K -means algorithm for clustering.
104 By borrowing the discriminative or supervised information contained in the
105 pre-trained AlexNet, which has been trained on a different classification task,
106 we still expect that our clustering results are improved, i.e. similar images
107 can be clustered together. This is one example of transfer learning within
108 the machine learning community, which aims to transfer knowledge learned
109 from one specific task to a similar but different task (Pan and Yang, 2009).
110 Note that the feature extraction process proposed here is different from that
111 in Gao and Wüthrich (2018). This is because our feature extraction process
112 involves supervised information from ImageNet classification task while the
113 one in Gao and Wüthrich (2018) is purely unsupervised. We recognize that
114 there are many different ways to perform such classification tasks. AlexNet
115 used here is based on convolutional neural networks. These networks have
116 been designed to find common structure at different locations in images. Al-
117 ternatively, one may try, for instance, density-based clustering which allows
118 to discover clusters of arbitrary shapes.

119 The rest of the paper is organized as follows. Section 2 describes v - a
120 heatmap. Section 3 shows the details of K -means algorithm and AlexNet.
121 Section 4 compares the clustering results of driving styles on our data. Sec-
122 tion 5 presents some concluding remarks.

123 2. The v - a heatmap

124 To generate v - a heatmaps we follow the steps in (Gao and Wüthrich,
125 2018). We select speed range $(5, 20]$ km/h and acceleration range $[-2, 2]$ m/s².
126 We divide both the speed range $(5, 20]$ and the acceleration range $[-2, 2]$ to,
127 say, 20 equidistant intervals. Thus, we partition the two-dimensional space
128 of $(5, 20] \times [-2, 2]$ to 400 congruent subregions R_j , $j = 1, 2, \dots, 400$. Note
129 that we could choose the numbers of equidistant intervals differently, but
130 we select the fixed number of 20, here, to fix ideas and also because this
131 will be in line with our numerical analysis. Next, we record the relative
132 amount of time spent in each subregion R_j , x_{ij} , for driver i , $i = 1, 2, \dots, N$.
133 x_{ij} satisfies the following probability constraints: $x_{ij} \geq 0$ for all j and
134 $\sum_{j=1}^{400} x_{ij} = 1$. This allows us to draw a heatmap based on these data for
135 each individual driver. For driver i , the heatmap data is represented by a
136 vector $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{i400}]^T$ of probability weights, see Figure 1 for its
137 two-dimensional illustration.

138 3. Methodology

139 In this section, we first introduce the K -means clustering algorithm that
140 can be applied to cluster heatmaps to subgroups. Then, we discuss two
141 feature extraction approaches that can be applied before the K -means al-
142 gorithm, the unsupervised principal component analysis (PCA) and the su-
143 pervised pre-trained AlexNet. There are two advantages of applying feature
144 extraction beforehand. First, we usually extract fewer features from the orig-
145 inal data when the dimensionality is large, e.g. 400 variables to describe one
146 heatmap in our task, in order to reduce the redundant information contained
147 in the data. Second, the extracted features are usually good representations

148 of the original data and can provide the useful information for the clustering
149 task.

150 3.1. *K*-means clustering

151 *K*-means clustering (Section 10.3.1 in [James et al. 2013](#)) is a clustering
152 technique that aims to find non-overlapping *K* clusters such that the within-
153 cluster variation of all *K* clusters is minimized.

154 Given *N* car drivers $\{1, 2, \dots, N\}$, the within-cluster variation S_k of the
155 *k*th cluster, C_k , is defined as

$$S_k = \frac{1}{N_k} \sum_{i, i' \in C_k} (\mathbf{x}_i - \mathbf{x}_{i'})^T (\mathbf{x}_i - \mathbf{x}_{i'}), \quad (1)$$

156 where N_k denotes the number of drivers in the *k*th cluster with $\sum_{k=1}^K N_k =$
157 *N*. Note that here we use the squared Euclidean distance between drivers
158 to measure the within-cluster variation. Hence *K*-means clustering aims to
159 solve the following optimization problem:

$$\min_{C_1, C_2, \dots, C_K} \sum_{k=1}^K \sum_{i, i' \in C_k} (\mathbf{x}_i - \mathbf{x}_{i'})^T (\mathbf{x}_i - \mathbf{x}_{i'}), \quad (2)$$

160 such that C_1, C_2, \dots, C_K provides a partition of all drivers $\{1, 2, \dots, N\}$.

161 Given that there are K^N ways to divide *N* drivers to *K* subgroups, the
162 following algorithm is usually used to find an approximate solution (local
163 minimum) of (2) with less computational cost.

164 Step 1 Randomly assign each driver to one of the *K* groups as initialization
165 step.

166 Step 2 Calculate the cluster mean for each cluster.

167 Step 3 Assign each driver to the cluster with the closest cluster mean (w.r.t.
168 the squared Euclidean distance).

169 Step 4 Iterate steps 2 and 3 until the assignment does not change.

170 Note that this algorithm has monotonically decreasing total within-
171 cluster variation, and therefore converges to a local minimum of (2). When
172 using K -means clustering, we need to specify the number of clusters K ,
173 which acts as a hyper-parameter. An optimal selection can be done by var-
174 ious methods, such as the elbow method (James et al., 2013) that plots the
175 sum of within-cluster variations against K and selects K where an elbow
176 appears in the graph.

177 3.2. Feature extraction before applying K -means

178 In this section, we present feature extraction before applying the K -
179 means algorithm. These feature extraction techniques may be understood
180 as representation learning techniques, and we apply the K -means algorithm
181 only to the learned representations. Interestingly, the K -means algorithm
182 does not use any information about the spatial structure of the heatmaps
183 because all information is stacked into a one-dimensional vector \mathbf{x}_i , however,
184 the second method presented in this section reflects spatial information in
185 the learned representation and, thus, the K -means results will have an im-
186 plicit spatial component.

187 3.2.1. Principal component analysis

188 Principal component analysis (PCA) (Jolliffe, 1986) is a simple, yet,
189 effective way to extract features that contain the most variation information
190 in data.

191 Given N drivers, we have a data matrix $\mathbf{X} \in \mathbb{R}^{N \times 400}$ that contains all
192 information \mathbf{x}_i of the drivers $i = 1, 2, \dots, N$ on the rows of \mathbf{X} . To obtain the
193 first few principal components, we first subtract the column means from \mathbf{X}

194 to obtain the mean-centred \mathbf{X}^c . We then apply the reduced singular value
 195 decomposition (SVD) to \mathbf{X}^c :

$$\mathbf{X}^c = \mathbf{U}\mathbf{D}\mathbf{V}^T, \quad (3)$$

196 where $\mathbf{U} \in \mathbb{R}^{N \times q}$ and $\mathbf{V} \in \mathbb{R}^{400 \times q}$ are two matrices with columns of left and
 197 right singular vectors, $\mathbf{D} \in \mathbb{R}^{q \times q}$ is a diagonal matrix with singular values
 198 $d_1 \geq d_2 \geq \dots \geq d_q \geq 0$.

199 In PCA, the columns of \mathbf{V} are known as principal components (PC) and
 200 the rows of $\mathbf{T} = \mathbf{U}\mathbf{D}$ are known as PC scores. In practice, we usually select
 201 the first r ($r \leq q$) PCs that can explain most of the variation of the data,
 202 e.g. 75%, to provide a good representation of the original dataset. Note that
 203 PCA is a purely unsupervised dimension reduction method because we do
 204 not involve any label information during the whole process. Moreover, it
 205 does not use the geometric structure of the heatmaps.

206 3.2.2. Transfer learning with the pre-trained AlexNet

207 From the previous section, we can see that the heatmap for each indi-
 208 vidual driver is simply treated as a row vector in \mathbf{X} . This approach ignores
 209 the geometric structure of the heatmaps, i.e. that the values of a small local
 210 area in the heatmap are similar to each other. To make use of this prop-
 211 erty, we propose to treat the heatmaps as RGB images rather than single
 212 vectors \mathbf{x}_i . Another advantage of treating the heatmaps as RGB images is
 213 that there is a rich literature and many algorithms in well-developed areas of
 214 image processing, in order to improve the clustering of driving styles.

215 Instead of using the purely unsupervised PCA, we propose to extract fea-
 216 tures with supervised information for better clustering via transfer learning.
 217 Transfer learning has attracted quite some attention in the machine learning

218 community in recent years (Pan and Yang, 2009; Torrey and Shavlik, 2010;
219 Shin et al., 2016). It aims to transfer the knowledge learned from source
220 tasks to a similar but different target task. In our task, there is a lack of
221 supervised information for the heatmap images, i.e. we do not have labels
222 of driving styles for the heatmaps, which makes the clustering task difficult.
223 This is the typical problem in common in clustering tasks. We aim to solve
224 this problem by borrowing supervised information learned from other im-
225 age classification tasks. For example, we can utilise the deep convolutional
226 neural network, AlexNet (Krizhevsky et al., 2012), that is trained on the
227 ImageNet data (Deng et al., 2009) to classify images to 1000 classes. Hence,
228 the features extracted by AlexNet contain supervised information that is
229 useful to differentiate images from different classes. If we feed our heatmaps
230 to AlexNet, then the features extracted by AlexNet may also be good to
231 distinguish between heatmap images with different patterns, i.e. different
232 driving styles. More specifically, we transfer the supervised information
233 from the source task, classifying ImageNet images, to our target task, clas-
234 sifying heatmap images. By using these extracted features, we can expect
235 an improvement in the clustering results.

236 AlexNet is the most popular deep convolutional neural network devel-
237 oped in the past decade. AlexNet has eight learned network layers with five
238 convolutional layers and three fully-connected layers. The architecture of
239 AlexNet is shown in Figure 2, where the light blue cube shows the input
240 RGB image, the orange cubes show the five convolutional layers and the
241 black rectangles show the three fully-connected layers. In our task, we un-
242 derstand the $v-a$ heatmaps now as RGB images, and we feed these RGB
243 images into the pre-trained AlexNet. Note that RGB images are repre-
244 sented by three-dimensional arrays representing the red, green or blue color

245 channels.

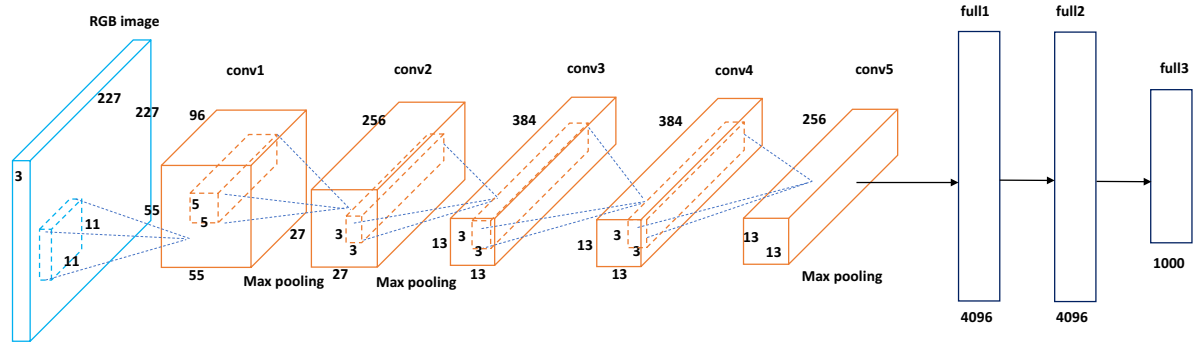


Figure 2: The architecture of AlexNet.

246 4. Data analysis

247 In the following data analysis, we compare the clustering performances
248 of (a) K -means, (b) K -means on PCA features and (c) K -means on AlexNet
249 features. We have performed this analysis on heatmaps coming from real
250 telematics car driving data and on simulated data. Our results did not differ
251 on the two data sets. Therefore, we have decided to present the results on
252 the simulated data, because this simulated data is publicly available which
253 allows one to replicate our results. We remark that the data generator for
254 the simulated data is based on bottleneck neural networks that have been
255 trained on real telematics car driving data, for more details we refer to [Gao](#)
256 [and Wüthrich \(2018\)](#).

257 *4.1. Simulated data*

258 The simulated heatmap data is obtained from the heatmap simulation
259 machine (Gao and Wüthrich, 2018)¹ with default parameter settings and
260 seeds. This simulation machine provides heatmaps of 2000 drivers. The
261 heatmap data is represented by a matrix $\mathbf{X} \in \mathbb{R}^{2000 \times 400}$.

262 *4.2. K-means clustering*

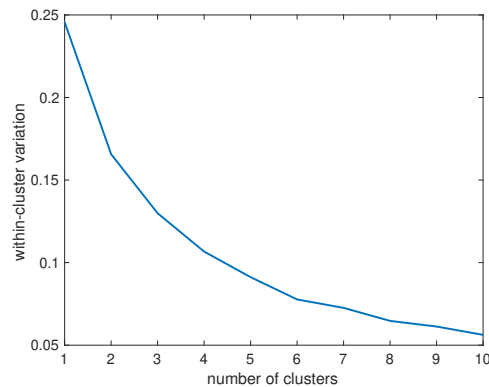


Figure 3: The scree plot of K -means.

263 We first show the results of applying K -means clustering to the heatmap
264 data directly. Figure 3 shows the scree plot of K -means clustering when we
265 use the original heatmap data \mathbf{X} as input. From this plot it is not obvious
266 which number of clusters we should choose as there is no clear elbow in the
267 picture. Based on Figure 3, we may need to set K to a large number, e.g.
268 larger than 10. However, we usually do not aim to set K to a very large
269 number because this may lead to over-fitting, and because for insurance
270 pricing we prefer categorical variables with only a few levels. When K is
271 set to the total number of drivers we receive the smallest within-cluster

¹<https://people.math.ethz.ch/~wueth/simulation.html>

272 variation of zero; however, no drivers are clustered in this case. This is why
273 we would like to see a scree plot with an elbow where the within-cluster
274 variation decreases quickly before the elbow while slowly after the elbow,
275 which gives us a natural selection criterion for K .

276 4.3. K -means clustering on PCA features

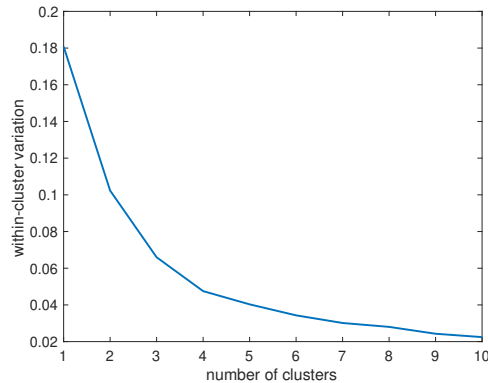


Figure 4: The scree plot of K -means on PCA features.

277 In this section, we show how the clustering results improve when we
278 extract features from the original data by PCA. The first two principal
279 components (PCs) are used, which explain 74% of the total variation in the
280 data. Thus, we represent \mathbf{x}_i by a two-dimensional vector, and we apply
281 K -means clustering on the two extracted PCA features.

282 Figure 4 shows the scree plot of K -means clustering on PCA features.
283 Compared to Figure 3 on the original data, there is a clear elbow shown
284 around $K = 4$ in Figure 4 (with PCA features). This suggests that $K = 4$
285 is a good choice for the number of clusters. I.e. this result gives us a natural
286 candidate for hyper-parameter K . Note that the PCA extraction reduces
287 the noise in the data because it focuses only on the most relevant PCs, and
288 the learned representations then allow for a more clear clustering picture.

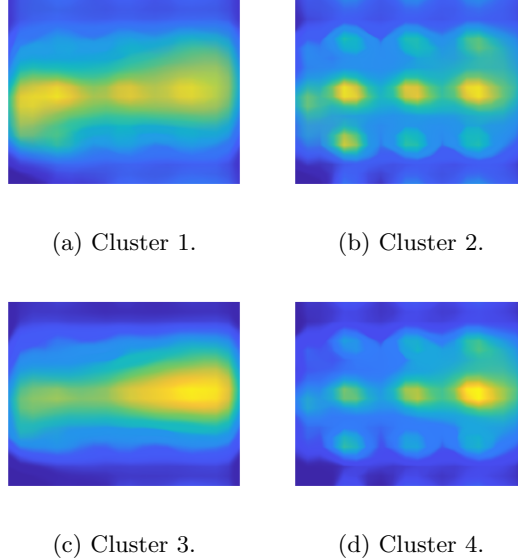


Figure 5: The cluster means of the four clusters identified by K -means on PCA features.

289 The cluster means, i.e. the average heatmap images of each cluster, are
 290 shown in Figure 5 when setting $K = 4$ (on PCA features). Figure 5 shows
 291 that different driving styles are presented in different clusters. For example,
 292 Cluster 2 shows a non-smooth driving style with a lot of time spent at high
 293 speeds and low speeds without any acceleration. The drivers in this cluster
 294 also tend to spend quite some time at low speeds and negative acceleration
 295 (braking). Cluster 4 shows a different non-smooth driving style where the
 296 drivers spend a large amount of time at high speeds and zero acceleration.
 297 Cluster 3 shows a smooth driving style. Cluster 1 seems to be a combination
 298 of both smooth and non-smooth driving styles, because the middle part
 299 of the mean image is smooth to an extent but not as smooth as that of
 300 Cluster 3. We suspect that Cluster 1 contains both driving styles. Figure 6
 301 shows individual drivers in each of the four clusters. This gives us some

302 evidence that Cluster 1 contains different driving behaviours, i.e. it is not
 303 as homogeneous as the other clusters. For example, the first one on the
 304 second column of Figure 6a is very smooth while the third one on the first
 305 column of Figure 6a is obviously non-smooth. This indicates that there is
 306 room for improvement of the clustering results of K -means on PCA features,
 e.g. making the clusters purer.

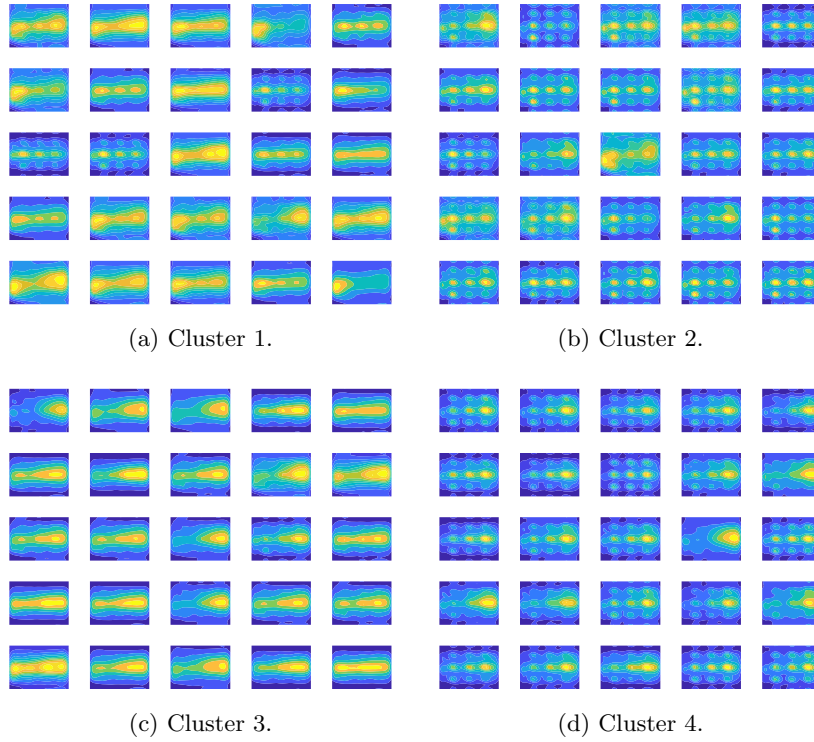


Figure 6: Example heatmap images of the four clusters identified by K -means on PCA features, cluster means are provided in Figure 5.

307

308 To have a further investigation of the physical meanings of PCs, we
 309 visualise the heatmaps via the scatter plot with the first two PCs in Figure 7,
 310 where the four clusters are labelled with different symbols. It seems that the
 311 first PC, i.e. PC1 in Figure 7, indicates the smoothness of the driving style.

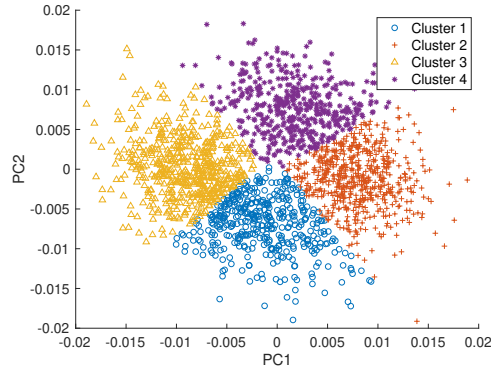


Figure 7: The scatter plot of heatmaps with the first two PCs. The clusters are labelled by K -means on PCA features.

312 Clusters 3 with relatively smooth driving style has small values in PC1 while
 313 Clusters 2 with relatively non-smooth driving style has large values in PC1.

314 4.4. K -means clustering on AlexNet features

315 Here we show the clustering results of K -means on AlexNet features.
 316 Different from the previous two experiments where the input is the data
 317 matrix \mathbf{X} , we export the heatmaps as RGB images (in .png format) and
 318 use these RGB images as input to the pre-trained AlexNet in Matlab².
 319 The high-level features from the fully-connected layer ‘fc7’ in Matlab are
 320 extracted from the pre-trained AlexNet. Because this layer provides a large
 321 number of 4096 extracted features, we reduce this dimension first by PCA,
 322 i.e. we apply PCA on the 4096 features extracted by AlexNet, and then
 323 use these ‘AlexNet+PCA’ features as input to the K -means algorithm. In
 324 the rest of this paper, we call these ‘AlexNet+PCA’ features as ‘AlexNet’
 325 features in short. The first two PCs are used which explain 79% of the total

²<https://uk.mathworks.com/help/deeplearning/ref/alexnet.html>

326 variation of the AlexNet features.

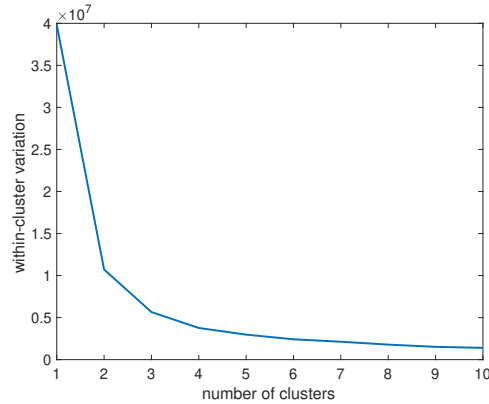


Figure 8: The scree plot of K -means on AlexNet features.

327 Similarly to the previous analysis, we first show the scree plot of the
328 K -means algorithm based on AlexNet features in Figure 8. Compared to
329 Figure 3 with the original data and Figure 4 with PCA features, Figure 8
330 with AlexNet features shows a much clearer elbow. Here we conclude that
331 $K = 4$ is a good choice for the number of clusters, because the reduction of
332 within-cluster variation becomes much smaller when the number of clusters
333 is larger than 4.

334 The four cluster means are shown in Figure 9. It seems that Clusters 1,
335 2 and 3 in Figure 9 with AlexNet features correspond to Clusters 4, 2 and
336 3 in Figure 5 with PCA features. The major difference is between Cluster
337 4 in Figure 9 with AlexNet features and Cluster 1 in Figure 5 with PCA
338 features. The plots show that the smooth driving styles are clustered to
339 Cluster 3 by AlexNet features. Cluster 4 in Figure 9d shows non-smooth
340 driving styles with a certain degree of smoothness in the middle right part,
341 compared with Clusters 1 and 2. We can also observe that the smoothness
342 of driving styles decreases in the order of Clusters 2, 1, 4 and 3.

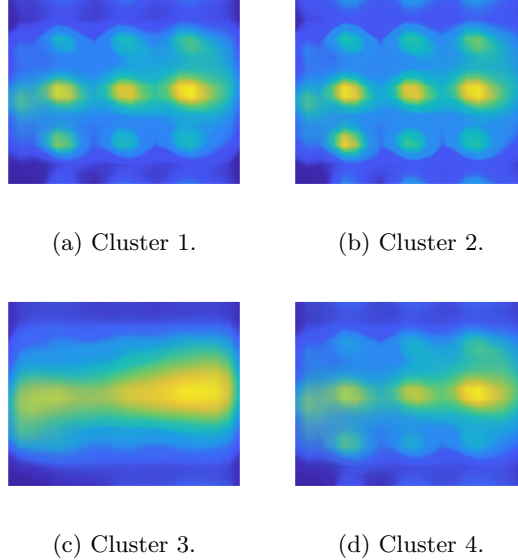


Figure 9: The cluster means of the four clusters identified by K -means on AlexNet features.

343 The improvement in cluster pureness by using AlexNet features is clearer
 344 in Figure 10 of example heatmap images. Cluster 4 examples in Figure 10d
 345 show heatmaps with a certain degree of non-smoothness. We cannot observe
 346 a clear mixture of smooth and non-smooth driving styles as in Cluster 1 with
 347 PCA features in Figure 6a.

348 The visualisation of the heatmap images are also shown as the scatter
 349 plot with the first two PCs of AlexNet features in Figure 11. We can see
 350 that PC1 also indicates the smoothness of the driving styles. The values of
 351 PC1 increase as the driving styles become smoother.

352 To have a closer look at the features extracted by AlexNet, we show the
 353 activation images of two layers for the heatmap image of the first driver.
 354 Each layer in AlexNet is consisting of many 2-dimensional arrays which are
 355 called channels. By visualising the channels, we can examine which parts

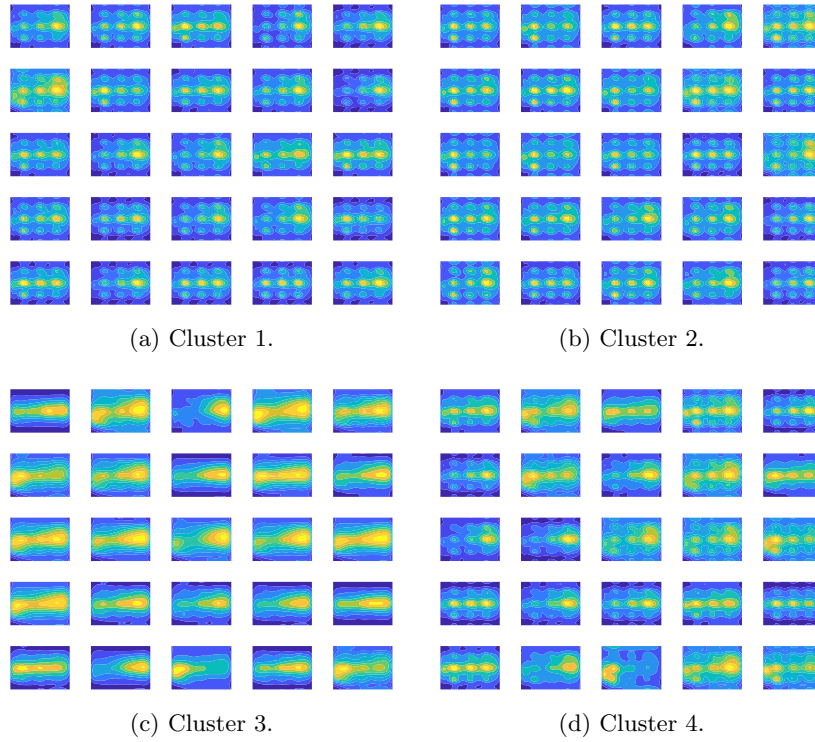


Figure 10: Example heatmap images of the four clusters identified by K -means on AlexNet features, cluster means are provided in Figure 9.

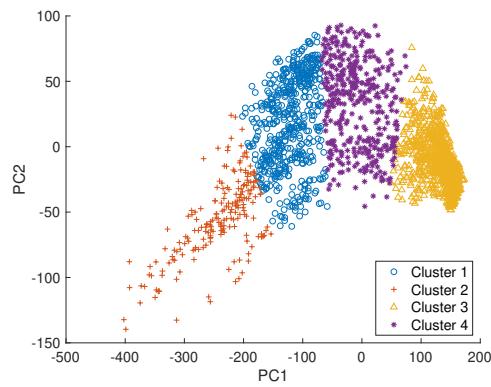


Figure 11: The PC plot of K -means on AlexNet features.

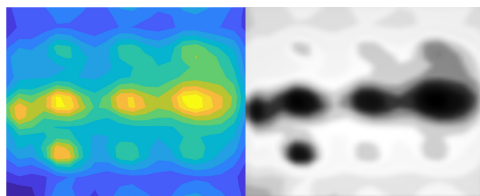


Figure 12: The strongest activation channel in the first convolutional layer, conv1, of driver 1.

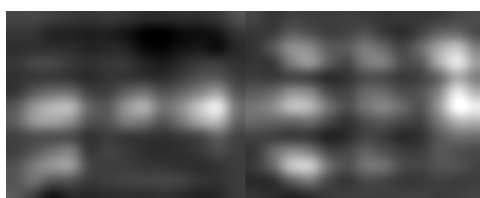


Figure 13: The 14th and 99th channels in the fifth convolutional layer, conv5, of driver 1.

356 of the image are strongly activated or which features are extracted by the
 357 channel. Usually, the channels in early layers extract simple features, e.g.
 358 colour or edge, while those in latter layers extract deep features, e.g. eyes in
 359 face recognition. For the heatmap image of driver 1, the strongest activation
 360 channel in the first convolutional layer, conv1, is shown in Figure 12. The
 361 white part indicates the area that is positively activated while the black
 362 part indicates the area that is negatively activated. It is clear that this layer
 363 extracts the features represented by the light blue area in the heatmap.
 364 Figure 13 shows the 14th and 99th channels in the fifth convolutional layer,
 365 conv5, for the first driver. These two channels extract features representing
 366 the non-smoothness of the heatmap image.

367 4.5. Quantitative measurement of clustering results

368 In previous sections, we have shown the improvement of using AlexNet
 369 features by visualising the elbow plots, the cluster mean images and the

370 example images of each cluster. Here, we aim to quantitatively measure
 371 this improvement. Given the fact that we do not have the ground truth
 372 labels of the heatmaps, it is not possible to compute the purity of the clus-
 373 tering results. Instead of using purity, we choose the average silhouette
 374 value (Rousseeuw, 1987) as our metric, which does not require the knowledge
 375 of ground truth labels. The average silhouette value measures how similar
 376 the heatmaps are to their own clusters and how dissimilar the heatmaps are
 377 to other clusters. The higher the average silhouette value, the better the
 378 clustering results.

379 After applying K -means, we assign each heatmap to one of the clusters
 380 C_1, C_2, \dots, C_K , where K is the predefined number of clusters and in our
 381 experiments it has chosen to be $K = 4$. For the i th heatmap that is assigned
 382 to the s th cluster, we calculate its average distance to all other heatmaps
 383 assigned to the same cluster:

$$a_i = \frac{1}{|C_s| - 1} \sum_{j \in C_s, j \neq i} d(i, j), \quad (4)$$

384 where $|C_s|$ denotes the number of heatmaps in cluster C_s . Thus, a_i measures
 385 how similar the i th heatmap is to its own cluster. Here we use the Euclidean
 386 distance between heatmaps i and j to measure the dissimilarity between
 387 them. We assume that two heatmaps with a small Euclidean distance have
 388 a high similarity while those with a large Euclidean distance have a high
 389 dissimilarity. To measure how dissimilar the i th heatmap is to other clusters,
 390 we calculate

$$b_i = \min_{k \neq s} \frac{1}{|C_k|} \sum_{l \in C_k} d(i, l), \quad (5)$$

391 where $k = 1, 2, \dots, K$.

392 The silhouette value of the i th heatmap is now defined as

$$s_i = \frac{b_i - a_i}{\max\{a_i, b_i\}}. \quad (6)$$

393 We can see that s_i takes values between $[-1, 1]$. The larger the value of
394 s_i , the higher the dissimilarity between the i th heatmap and other clusters
395 while the higher the similarity between the i th heatmap and its own cluster.
396 Thus, a large value of s_i indicates better clustering of the i th heatmap.

397 To measure how well the clustering results are for all heatmaps, we can
398 simply take the average silhouette value of all heatmaps:

$$s_{all} = \frac{1}{N} \sum_{i=1}^N s_i, \quad (7)$$

399 where N is the total number of heatmaps and in our experiment it is $N =$
400 2000.

Table 1: The average silhouette values of all heatmaps when clustering by K -means with $K = 4$.

	Pure K -means	PCA features	AlexNet features
s_{all}	0.4432	0.5769	0.7261

401 We show s_{all} for the clustering results of K -means with $K = 4$ by using
402 the pure K -means, PCA features and AlexNet features in Table 1. This
403 silhouette value shows a clear increase from the original K -means to the
404 AlexNet extracted K -means method, indicating that we receive much more
405 purity when appropriately pre-processing the heatmaps before applying the
406 K -means algorithm.

407 **5. Conclusion**

408 Clustering driving styles by analysing speed-acceleration v - a heatmaps
409 is one interesting topic in studying telematics car driving data. In this
410 study, we propose to process the heatmaps as images and involve supervised
411 information via transfer learning in our clustering task. More specifically,
412 we propose to extract features with supervised information from the pre-
413 trained AlexNet for image classification tasks and conduct clustering based
414 on these features. Experiments on both simulated data and real data show
415 the improvement of clustering results compared with using original data and
416 PCA features. This is verified by comparing the corresponding silhouette
417 values that clearly prefer the pre-trained AlexNet features.

418 **References**

- 419 Ayuso, M., Guillen, M., Pérez-Marín, A.M., 2016a. Telematics and gen-
420 der discrimination: Some usage-based evidence on whether men’s risk of
421 accidents differs from women’s. *Risks* 4, 10.
- 422 Ayuso, M., Guillen, M., Pérez-Marín, A.M., 2016b. Using GPS data to
423 analyse the distance traveled to the first accident at fault in pay-as-you-
424 drive insurance. *Transportation Research Part C: Emerging Technologies*
425 86, 160–167.
- 426 Bian, Y., Yang, C., Zhao, J.L., Liang, L., 2018. Good drivers pay less: A
427 study of usage-based vehicle insurance models. *Transportation Research*
428 *Part A: Policy and Practice* 107, 20–34.
- 429 Boucher, J.P., Côté, S., Guillen, M., 2017. Exposure as duration and dis-

430 tance in telematics motor insurance using generalized additive models.
431 Risks 5, 54.

432 Carfora, M.F., Martinelli, F., Mercaldo, F., Nardone, V., Orlando, A., San-
433 tone, A., Vaglini, G., 2019. A “pay-how-you-drive” car insurance approach
434 through cluster analysis. *Soft Computing* 23, 2863–2875.

435 Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Li, F.F., 2009. ImageNet:
436 A Large-Scale Hierarchical Image Database, in: CVPR09.

437 Denuit, M., Guillen, M., Trufin, J., 2019. Multivariate credibility modelling
438 for usage-based motor insurance pricing with behavioural data. *Annals of*
439 *Actuarial Science* 13, 378–399.

440 European Commission, 2012. EU rules on gender-neutral pricing in
441 insurance industry enter into force URL: [https://ec.europa.eu/
442 commission/presscorner/detail/en/IP_12_1430](https://ec.europa.eu/commission/presscorner/detail/en/IP_12_1430).

443 Gao, G., Meng, S., Wüthrich, M.V., 2019. Claims frequency modeling using
444 telematics car driving data. *Scandinavian Actuarial Journal* 2019, 143–
445 162.

446 Gao, G., Wüthrich, M.V., 2018. Feature extraction from telematics car
447 driving heatmaps. *European Actuarial Journal* 8, 383–406.

448 Ho, S.H., Wong, Y.D., Chang, V.W.C., 2014. Developing singapore driv-
449 ing cycle for passenger cars to estimate fuel consumption and vehicular
450 emissions. *Atmospheric Environment* 97, 353–362.

451 Hung, W.T., Tong, H., Lee, C., Ha, K., Pao, L., 2007. Development of a
452 practical driving cycle construction methodology: A case study in hong

- 453 kong. *Transportation Research Part D: Transport and Environment* 12,
454 115–128.
- 455 James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. *An introduction to*
456 *statistical learning*. volume 112. Springer.
- 457 Jolliffe, I.T., 1986. Principal components in regression analysis, in: *Principal*
458 *component analysis*. Springer, pp. 129–155.
- 459 Kamble, S.H., Mathew, T.V., Sharma, G.K., 2009. Development of real-
460 world driving cycle: Case study of pune, india. *Transportation Research*
461 *Part D: Transport and Environment* 14, 132–140.
- 462 Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet classification
463 with deep convolutional neural networks, in: *Advances in Neural Infor-*
464 *mation Processing Systems*, pp. 1097–1105.
- 465 Lemaire, J., Park, S.C., Wang, K.C., 2016. The use of annual mileage as a
466 rating variable. *ASTIN Bulletin: The Journal of the IAA* 46, 39–69.
- 467 Pan, S.J., Yang, Q., 2009. A survey on transfer learning. *IEEE Transactions*
468 *on Knowledge and Data Engineering* 22, 1345–1359.
- 469 Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation
470 and validation of cluster analysis. *Journal of Computational and Applied*
471 *Mathematics* 20, 53–65.
- 472 Shin, H.C., Roth, H.R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mol-
473 lura, D., Summers, R.M., 2016. Deep convolutional neural networks for
474 computer-aided detection: CNN architectures, dataset characteristics and
475 transfer learning. *IEEE Transactions on Medical Imaging* 35, 1285–1298.

- 476 Sun, S., Bi, J., Guillen, M., Pérez-Marín, A.M., 2020. Assessing driving risk
477 using internet of vehicles data: an analysis based on generalized linear
478 models. *Sensors* 20, 2712.
- 479 Torrey, L., Shavlik, J., 2010. Transfer learning, in: *Handbook of research*
480 *on machine learning applications and trends: algorithms, methods, and*
481 *techniques*. IGI Global, pp. 242–264.
- 482 Verbelen, R., Antonio, K., Claeskens, G., 2018. Unravelling the predictive
483 power of telematics data in car insurance pricing. *Journal of the Royal*
484 *Statistical Society: Series C (Applied Statistics)* 67, 1275–1304.
- 485 Weidner, W., Transchel, F.W., Weidner, R., 2017. Telematic driving profile
486 classification in car insurance pricing. *Annals of Actuarial Science* 11,
487 213–236.
- 488 Wüthrich, M.V., 2017. Covariate selection from telematics car driving data.
489 *European Actuarial Journal* 7, 89–108.