Final Report - CS8692

Zihao WEN

December 2019

1 Introduction

The era of autonomous driving is coming. In recent years, various kinds of self-driving vehicles are springing up all over the world, and many of them, like the Google's Waymo, have been running in the real world for a long time. The number of self-driving vehicles running on the road is increasing so rapidly that in a foreseeable future, there will be almost no vehicles which are driven by human drivers. However, before the autonomous driving were applied in large scale, there will be a long time when a considerable number of human drivers still drive on the road, which means self-driving vehicles co-exist with and need to deal with the interaction with human-driving vehicles. Therefore, on one hand, it is crucial for autonomous vehicles to understand human driving behaviors so that they could correctly interpret the intentions of surrounding vehicles and make better driving decisions. On the other hand, both human passengers sitting on the self-driving vehicles and human drivers in the nearby vehicles will feel comfortable if they demonstrate a consistent driving style like human drivers do. Otherwise, human might not understand what they are doing or even be confused, which could raise social distrust on autonomous driving or even cause traffic accidents.

Such problems bring challenges as well as opportunities. The perception, prediction and motion planning of autonomous vehicles have been studied thoroughly in the past decades, but only a few of the studies have ever considered the impact of driving style. In this paper, I am going to go deeper into driving style classification and explore the features that defines a certain driving style. Based on the result of driving style classification, I proposed a criterion to judge whether a behavior is desirable for autonomous driving or not, in other word, to judge if they demonstrate a consistent driving style.

2 Related Works

Human driving style classification and recognition have drawn some attention in the past decade because of its important role in driving safety as well as vehicle energy management. Some of the algorithms are proposed to find the environment-friendly driving style to control emission and achieve Eco-driving [1], [2]. Some aim at improving ADASs design for more comfortable and safer driving [3], or leveraging driving style recognition to prevent car stealing [4], or even helps identify the number of drivers during one trip for insurance estimation [5].

Depending on different aims, the representation of the driving style and the methods of recognition are different as well. The driving style can be represented as discrete classes, e.g. aggressive, moderate and calm [3]–[5] or continuous function [1], [2]. The classification method can be categorized into three classes: rule-based, model-based and learning algorithm-based. The rule-based methods classify driving behaviors by defining threshold over particular variables that allocate them into a driving style [6], and if the rules are too complex, they can be replaced by a fuzzy logic map [7]. Model-based describe driving styles through a set of equations whose parameters can be tuned to fit the data [8]. Because of the advance of machine learning, some common machine learning algorithms and models including GMM[9], hierachical Bayesian regression model[1], k-NN[10] and Markov model[11] are applied to this task as well.

3 Methodology

In this paper, my main goal is to classify human driving styles. Specifically, I try to classify driving styles from the existed trajectory data set without driving style labels through analyzing the features of trajectories and differences among them. To accomplish this task, clustering algorithms are considered because they belong to unsupervised learning method and it is a good way to find a classification using clustering.

Among the clustering algorithms I choose the K-means algorithm which is adopted by [12]. The K-means algorithm is an iterative cluster analysis algorithm. Its steps are to randomly select K objects as the initial cluster center, and then calculate the distance between each object and each seed cluster center, assigning the object to its nearest cluster center. Once all objects are assigned, the cluster center of each cluster is recalculated based on the existing objects in the cluster. This process is repeated until the cluster centers no longer change. The cluster centers and the objects assigned to them represent a cluster.

Algorithm 1 K-means Algorithm

Input: The training data set *data* and the cluster number k
Output: k cluster centers and an array *label* (each of its elements corresponds to a data point in *data*)

Randomly choose k initial cluster centers from *data*
while True **do**
    for $point \in data$ **do**
        Calculate the distance d to every cluster center c_i
        assign *point* to the cluster i whose center c_i has minimum d to it
    end for
    for each cluster i **do**
        Calculate a new cluster center c_i' of data points assigned to this cluster
        Update c_i to c_i'
    end for
    if no cluster centers are changed **then**
        break
    end if
end while

I assume that the ego self-driving vehicle can only observe a short historical trajectory of a surrounding vehicle and use this trajectory for driving style recognition. So first I segment the whole trajectory from the data set into short ones. Specifically, I let

$X^{(t)} = [x^{(t-t_h)}, \dots, x^{(t-1)}, x^{(t)}]$ be the short historical trajectory up to time t, where $x^{(t)}$ is the state (including coordinates, velocity, acceleration, curvature, angular acceleration, jerk, etc) of the vehicle at time t.

The fact that some parameters like curvature and acceleration could be very different in various driving behaviors can make it hard to correctly distinguish different driving styles. Therefore, a two-level classification model is trained in order to efficiently recognize the driving style. These two levels are clustering a small number of driving behaviors and further identifying the detailed driving styles in each type of driving behavior, respectively.

At the first level, the task is to cluster these trajectory segments into several driving behaviors. Considering that the driving behavior of a trajectory is actually time-series based, so I choose the following feature vector:

$$x = (x^{(start)}, y^{(start)}, x^{(25\%t)}, y^{(25\%t)}, x^{(50\%t)}, y^{(50\%t)}, x^{(75\%t)}, y^{(75\%t)}, x^{(end)}, y^{(end)})^T \tag{1}$$

where $(x^{(start)}, y^{(start)})$ is the x, y coordinate of the vehicle at the starting frame of a trajectory, $(x^{(k\%t)}, y^{(k\%t)})$ is the x, y coordinate of the vehicle at the k percentile frame of a trajectory and $(x^{(end)}, y^{(end)})$ is the x, y coordinate of the vehicle at the end frame of a trajectory. The shape of most of the trajectories can be modeled by these five points. Also, each trajectory are translated to move their starting point to the origin $(0, 0)$ in order to exclude the influence of the coordinate of its starting point. By the way, the first two terms of the feature vector can now be removed because $(x^{(start)}, y^{(start)})$ is always $(0, 0)$ under this circumstance.

The training set is divided into k group according to the clustering result of the first level. At the second level, I aim to cluster the driving style of each cluster of driving behaviors. Through using multiple different feature vectors for the clustering, I find that the following parameters are crucial in distinguishing different driving styles: acceleration, curvature, angular acceleration, jerk (derivative of acceleration). And I use some statisic features of these parameters to better recognize the distribution of driving styles. Therefore, the following feature vector is chosen:

$$\begin{aligned} x =(&a_v^{(min)}, a_v^{(25\%)}, a_v^{(50\%)}, a_v^{(75\%)}, a_v^{(max)}, \kappa^{(min)}, \kappa^{(25\%)}, \kappa^{(50\%)}, \kappa^{(75\%)}, \kappa^{(max)}, \\ &a_\kappa^{(min)}, a_\kappa^{(25\%)}, a_\kappa^{(50\%)}, a_\kappa^{(75\%)}, a_\kappa^{(max)}, j_k^{(min)}, j_k^{(25\%)}, j_k^{(50\%)}, j_k^{(75\%)}, j_k^{(max)})^T \end{aligned} \tag{2}$$

where $p^{(k\%)}$ is the k percentile of the parameter p. And the difference between two feature vectors is calculated using Eucildean distance.

Based on the clustering result, here are the criteria to determine whether a trajectory produced by a self-driving vehicle is desirable or not:

- Cut the trajectory into 2-second pieces and get their driving style labels. If over δ percent of 0 the trajectory pieces are of the same driving style, then it is desirable. Otherwise, it is undesirable.

- If the Euclidean distance from a trajectory segment to the cluster center it belongs to is larger than k times of standard deviation of that cluster, that means it does not belong to any human driving style, thus the trajectory is undesirable.

where $\delta > 0$ and $k > 0$ are parameters which can be defined by the users.

4 Experiments

4.1 Data Pre-processing

I use the NGSIM (Next Generation SIMulation) I-80 dataset as the training data set. NGSIM is the most widely used dataset on traffic flow and driver models. NGSIM US-101 and I-80 datasets consists of trajectories of real freeway traffic captured at 10 Hz in 90 minutes from 2005 to 2006, where 9206 vehicles are recorded in a 6 lanes freeway.

I extract the data from the source csv file and assemble the trajectory of every vehicle from the frame information. Since NGSIM only provides a limited data about the feature of a trajectory, such as the curvature and angle at each frame, I have to compute those features manually. Specifically, the angle at i-th frame is approximated to the angle of the segment between the position at i-th frame and the position at $(i-1)$-th frame. And the curvature, angular velocity and angular acceleration at the i-th are computed using the angle information at the $(i-1)$- ,i-, $(i+1)$-th frame. The trajectories are then cut into 2-second-long segments. Segments which do not have a duration of 2 seconds or miss some frame data are discarded. In this way, 116812 segments of trajectory are collected and they are used as the training data set.

4.2 Experiment Design and Results

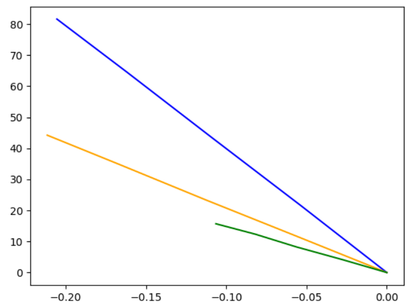
The experiment is divided into two parts, each of which corresponds to one level of the clustering model. The first part is clustering driving behaviors. I extract the relative data from the training data set to form a set of feature vectors shown in function (1), which is the dataset for the first part of the experiment. The result is shown in figure 1 (a). Here I choose $K = 3$.

Here only the shape of the mean trajectory of each cluster is drawn. The x axis is the lateral coordinate and the y axis is the longitudinal coordinate. Obviously, this result is incorrect because all of them move towards the same side. I discover that the difference among these cluster mean trajectories is the longitudinal coordinate, while the lateral coordinate, is more decisive in the clustering of driving behavior, so I modify the function in K-means whose purpose is to compute the distance from a trajectory x to a cluster center x' to apply a weight between x and y coordinate:

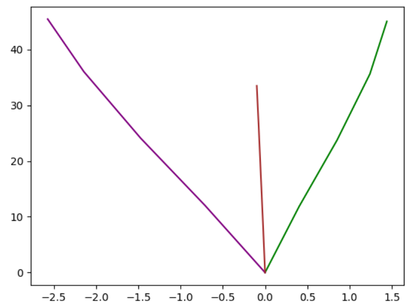
$$d(\boldsymbol{x}, \boldsymbol{x'}) = \left\| \begin{bmatrix} x^{(25\%t)} \\ x^{(50\%t)} \\ x^{(75\%t)} \\ x^{(end)} \end{bmatrix} - \begin{bmatrix} x'^{(25\%t)} \\ x'^{(50\%t)} \\ x'^{(75\%t)} \\ x'^{(end)} \end{bmatrix} \right\|^2 + \lambda \left\| \begin{bmatrix} y^{(25\%t)} \\ y^{(50\%t)} \\ y^{(75\%t)} \\ y^{(end)} \end{bmatrix} - \begin{bmatrix} y'^{(25\%t)} \\ y'^{(50\%t)} \\ y'^{(75\%t)} \\ y'^{(end)} \end{bmatrix} \right\|^2 \tag{3}$$

The figure 1 (b), (c) shows the result after I apply function (3). The mean trajectories now have clear meanings. They represents switching to left lane, going straight forward and switching to right lane respectively. In figure 1 (c), I draw some of the trajectories along with the three means. It proves that such classification do work and the result is correct.

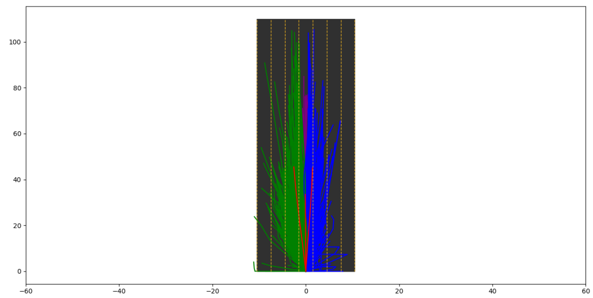
I divide the training dataset into three sub-datasets based on the cluster result of the first part of the experiment. For each sub-dataset, I perform the K-means clustering to classify driving styles. The clustering results are shown in figure 2, 3 and 4. I choose the cluster number $K = 5$ here. It can be seen from these result that the means of acceleration, angular acceleration, curvature and jerk are almost the same of each clusters



(a) Incorrect means



(b) Correct means



(c) Visualization of Clustering

Figure 1: Cluster Result of Driving Behaviors

in every sub-datasets (all around 0), while the minimal and maximal values have relatively large differences. We can also observe that in most cases, the driving style mean which has a large positive maximal also has a large negative minimal and vice versa (Noted that here large is referred to the absolute value). Thus if we order these clusters based on the descending order of their maximal value and we can label them as different degrees from aggressive to conservative.

5 Conclusion and Future Works

From the last chapter I successfully classify different driving styles from different driving behaviors. Suppose now a trajectory was given by a self-driving car, then we could simply apply the criteria proposed in Chapter 3 to determine whether it is desirable or not. In the future, other clustering algorithms could be tried to improve the performance of clustering and replace the criteria with inverse reinforcement learning reward functions.

References

[1] A. Mudgal, S. Hallmark, A. Carriquiry, and K. Gkritza, "Driving behavior at a roundabout : A hierarchical Bayesian regression analysis," *Transportation Research Part D*, vol. 26, pp. 20–26, 2014, ISSN: 1361-9209. DOI: `10.1016/j.trd.2013.10.003`. [Online]. Available: `http://dx.doi.org/10.1016/j.trd.2013.10.003`.

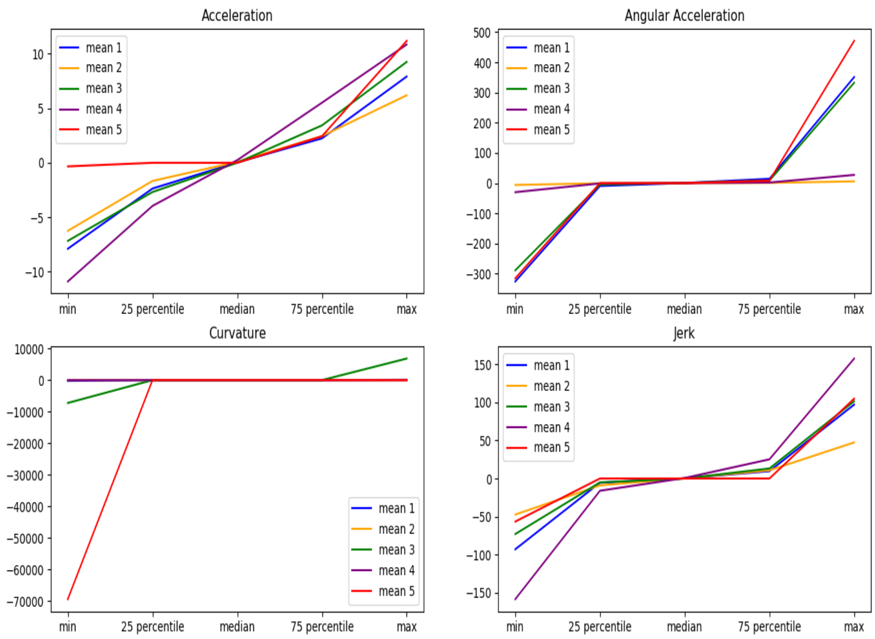


Figure 2: Driving Style of Switching Right

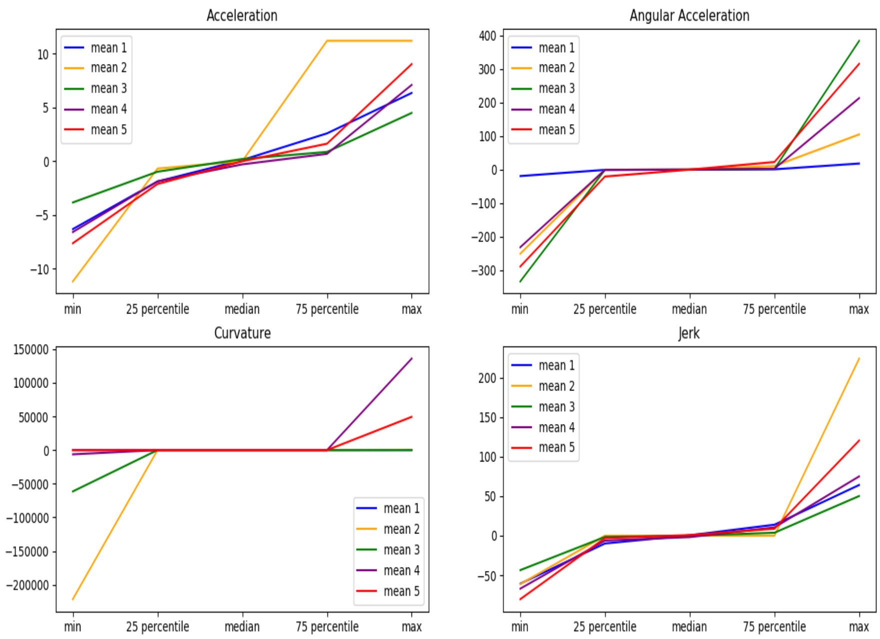


Figure 3: Driving Style of Going Straight

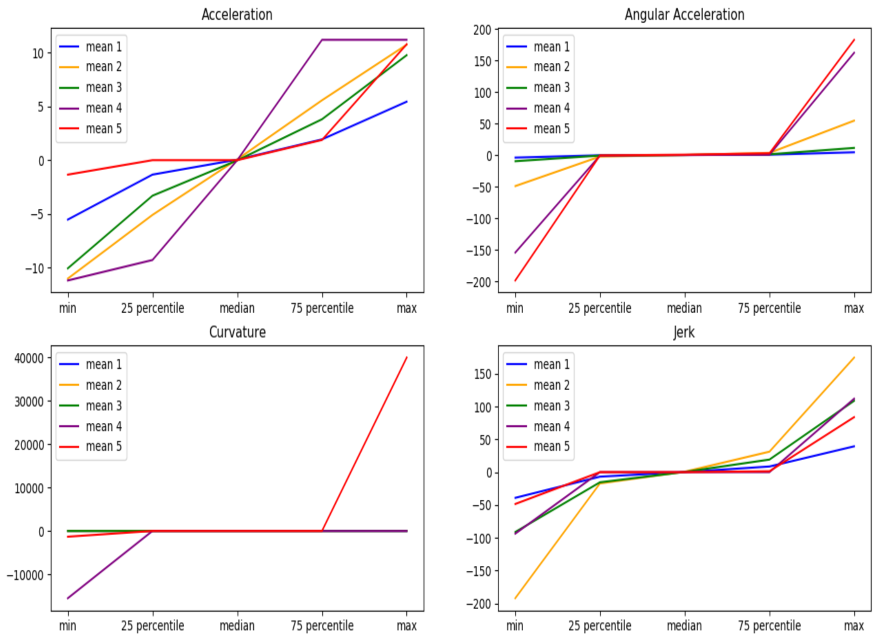


Figure 4: Driving Style of Switching Left

[2] E. Suzdaleva and I. Nagy, "An online estimation of driving style using data-dependent pointer model," *Transportation Research Part C*, vol. 86, no. November 2017, pp. 23–36, 2018, ISSN: 0968-090X. DOI: 10.1016/j.trc.2017.11.001. [Online]. Available: https://doi.org/10.1016/j.trc.2017.11.001.

[3] G. Li, F. Zhu, X. Qu, B. Cheng, S. Li, and P. Green, "Driving Style Classification Based on Driving Operational Pictures," *IEEE Access*, vol. 7, pp. 90 180–90 189, 2019. DOI: 10.1109/ACCESS.2019.2926494.

[4] F. Martinelli, F. Mercaldo, A. Orlando, V. Nardone, A. Santone, and A. K. Sangaiah, "Human behavior characterization for driving style recognition in vehicle system," *Computers and Electrical Engineering*, vol. 0, pp. 1–16, 2018, ISSN: 0045-7906. DOI: 10.1016/j.compeleceng.2017.12.050. [Online]. Available: https://doi.org/10.1016/j.compeleceng.2017.12.050.

[5] W. Dong, T. Yuan, K. Yang, C. Li, and S. Zhang, "Autoencoder Regularized Network For Driving Style Representation Learning," pp. 1603–1609, 2016.

[6] Y. L. Murphey, R. Milton, and L. Kiliaris, "Driver's style classification using jerk analysis," *2009 IEEE Workshop on Computational Intelligence in Vehicles and Vehicular Systems, CIVVS 2009 - Proceedings*, pp. 23–28, 2009. DOI: 10.1109/CIVVS.2009.4938719.

[7] D. Dörr, D. Grabengiesser, and F. Gauterin, "Online driving style recognition using fuzzy logic," *2014 17th IEEE International Conference on Intelligent Transportation Systems, ITSC 2014*, pp. 1021–1026, 2014. DOI: 10.1109/ITSC.2014.6957822.

[8] W. Wang, J. Xi, and H. Chen, "Modeling and Recognizing Driver Behavior Based on Driving Data: A Survey," vol. 2014, 2014. DOI: 10.1155/2014/245641.

[9] "Driver modeling based on driving behavior and its evaluation in driver identification," *Proceedings of the IEEE*, vol. 95, no. 2, pp. 427–437, 2007, ISSN: 00189219. DOI: 10.1109/JPROC.2006.888405.

[10] V. Vaitkus, P. Lengvenis, and G. Žylius, "Driving style classification using long-term accelerometer information," *2014 19th International Conference on Methods and Models in Automation and Robotics, MMAR 2014*, pp. 641–644, 2014. DOI: 10.1109/MMAR.2014.6957429.

[11] C. Guardiola, B. Pla, D. Blanco-Rodríguez, and A. Reig, "Modelling driving behaviour and its impact on the energy management problem in hybrid electric vehicles," *International Journal of Computer Mathematics*, vol. 91, no. 1, pp. 147–156, 2014, ISSN: 00207160. DOI: 10.1080/00207160.2013.829567. [Online]. Available: https://doi.org/10.1080/00207160.2013.829567.

[12] M. Van Ly, S. Martin, and M. M. Trivedi, "Driver classification and driving style recognition using inertial sensors," in *2013 IEEE Intelligent Vehicles Symposium (IV)*, Jun. 2013, pp. 1040–1045. DOI: 10.1109/IVS.2013.6629603.