Classification Of Drivers Based On Their Driving Patterns

Rutul Patel*, Saumil Patel†, Kunj Kanzariya‡, and Ayush Patel§
*†‡§Computer Science and Engineering, Ahmedabad University, Ahmedabad, India
{rutul.p, saumil.p, kunj.k1, ayush.p3}@ahduni.edu.in

***Abstract*—Driving behavior analysis is crucial for enhancing road safety and developing intelligent transportation systems. Traditional evaluation methods often fall short in capturing the complexity of driving patterns. However, recent advancements in data analytics and sensor technologies enable the collection of rich datasets, facilitating more comprehensive analyses. In this study, we propose a methodology that leverages dynamic time warping (DTW) and K-means clustering to categorize driving behavior based on risk patterns derived from UAV-collected data. By representing driving patterns as variable-length time series, we create classification models that capture subtleties and temporal relationships. The K-means algorithm, guided by DTW similarity metrics, partitions driving patterns into clusters, revealing distinct behavior profiles. Statistical techniques aid in determining the optimal number of clusters, ensuring effective characterization of driving patterns. Our results demonstrate the effectiveness of the proposed approach in uncovering valuable insights for road safety and transportation decision-making. Personalized approaches derived from identified behavior clusters can significantly contribute to improving road safety and driving experiences.**

***Index Terms*—Unsupervised Learning, Clustering, Driving risk, Time-series**

I. Introduction

A. Background

Driving behavior is a complex interaction of various factors, including individual habits, external conditions, and situation. Understanding and analyzing driving patterns have become important for enhancing road safety, designing intelligent transportation systems, and developing advanced driver-assistance systems (ADAS). Conventional techniques for evaluating driving behavior frequently depend on subjective observations or constrained metrics, which may not adequately represent the entire range of driving behaviors. But because to developments in data analytics and sensor technologies, it's now possible to collect rich datasets encompassing diverse driving scenarios and behaviors.

B. Motivation

To utilize the abundance of driving data accessible today to create advanced and accurate models for categorizing driving behavior. Our goal is to uncover valuable insights that can enhance road safety and enable better decision-making in transportation. The dynamic and evolving nature of driving behavior is acknowledged by presenting driving patterns as variable length time-series. By using this method, we are able to produce classification models that are more resilient by capturing the subtleties and temporal relationships present in driving data. Furthermore, by using machine learning methods to examine these patterns, we can find undiscovered correlations, pinpoint risk factors, and eventually aid in the creation of intelligent systems that can foresee and stop unfavorable driving outcomes.

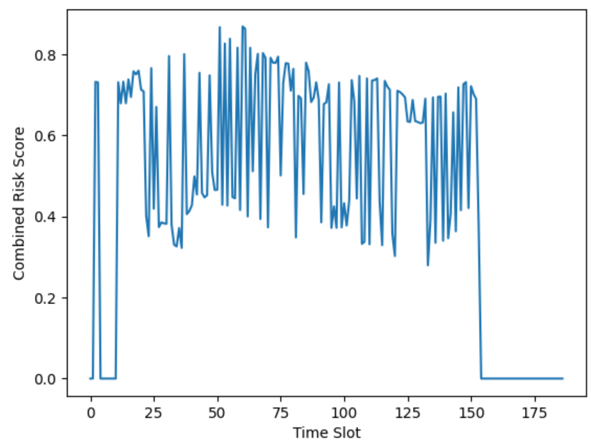

Fig. 1. Risk vs Time slot graph of an example vehicle

II. Dataset Description

The Data is collected using UAV. An camera mounted Drone was used to get top view of the road. YOLOv5 was used to detect the position of moving vehicles on road. Where each data point was captured every 33ms. So, If a car is going fast it will have lower number of time-series data compared to a slow moving vehicle. Each data point contains the coordinates of the car at that particular time instant, velocity and direction. Then a risk value was assigned to each car at each instant. In Fig. 1 we visualize the risk vs time of a particular car.

III. Problem formulation

Our task is to find similar drivers based on their driving pattern. A particular driving pattern has a risk associated with it at each time instant which is labelled by hand. The driving pattern is quantized by this risk values. A time-series of risk values is generated for each driver that appeared in the frame. Note that each time-series will have variable length depending on time it stayed in the frame. We have to find similarities in these risk time-series and make clusters of similar driving patterns.

we will denote this time series for a driver i as $D^i = \{d_{1i}, d_{2i}, d_{3i}, ..., d_{Ti}\}$, where T is the number of time instances, and d_{ti} is the risk value at time t. In time series analysis, dynamic time warping (DTW) is one of the algorithms for measuring similarity between two temporal sequences that do not align exactly in time, speed, or length. As our problem has variable length time-series, for Similarity measurement we'll use Dynamic Time Warping (DTW) to measure the similarity between two time series. DTW computes the optimal alignment between two time series by warping them in time to minimize the distance between corresponding points. The DTW distance between two time series D_i and D_j can be calculated as:

$$\text{DTW}(D_i, D_j) = \min \sqrt{\sum_{(t_k, t'_k) \in \text{path}} (d_{it_k} - d_{jt'_k})^2}$$

The objective function for k-means clustering with Dynamic Time Warping (DTW) as the similarity metric can be written as:

$$J = \sum_{i=1}^{k} \sum_{x \in C_i} \text{DTW}(x, \mu_i)^2$$

k is the number of clusters. C_i is the ith cluster. x is a time series in cluster C_i. μ_i is the centroid of cluster C_i.where the minimum is taken over all possible warping paths.

IV. Methodology

We use K-means clustering approach to find similar driving patterns. We start by pre-processing the data. In which we remove all the time-series with zero risk at all time instances. Also handled any missing values, outliers, or noise, ensuring data uniformity in format and scale through normalization or standardization techniques.

TABLE I
SILHOUETTE SCORES FOR DIFFERENT NUMBERS OF CLUSTERS

Number of Clusters	Silhouette Score
2	0.373
3	0.319
4	0.134
5	0.154
6	0.160

Next, we employ statistical techniques such as the elbow method or silhouette score (Table 1) to determine the optimal number of clusters (k) for the k-means algorithm. Through experimentation and evaluation of clustering performance, we identify the most suitable k that effectively captures distinct driving behavior patterns. Utilizing the k-means algorithm, we initialize k centroids randomly and proceed to partition the driving patterns into k clusters based on similarity. This

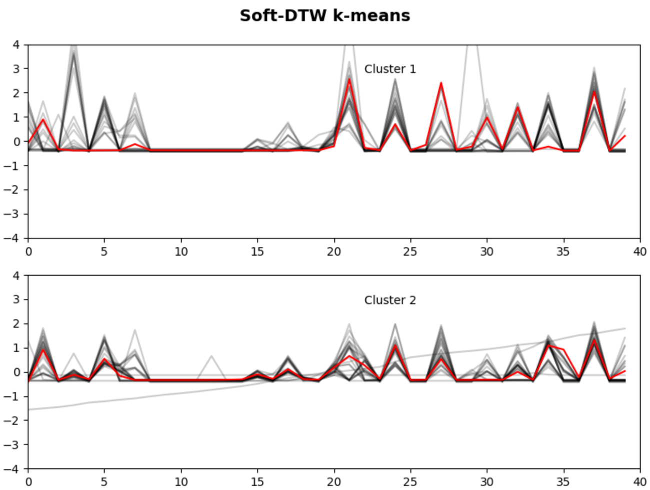

Fig. 2. Clusters produced when k=2

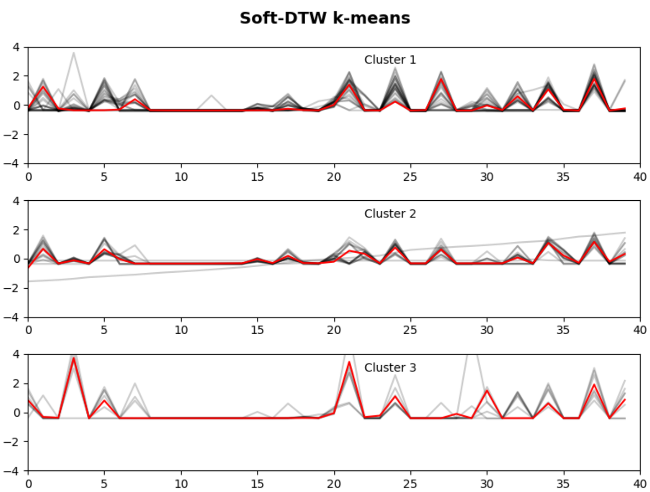

Fig. 3. Clusters produced when k=3

iterative process involves assigning data points to the nearest centroid and updating centroids until convergence.

We leverage the capabilities of the tslearn[1] framework, a powerful toolkit designed specifically for time series analysis and clustering tasks. By integrating tslearn into our workflow, we gain access to a wide range of functionalities tailored for handling variable-length time-series data effectively.

V. Results/Inferences

As illustrated in Figures 2 and 3, the identified clusters provide valuable insights into the diverse driving behaviors observed within the dataset. These figures visually represent the distinct characteristics and trends within each cluster. Understanding this information enables targeted interventions and personalized approaches for improving road safety, enhancing traffic management strategies, and developing effective driver assistance systems.

VI. Conclusion

By identifying and characterizing these distinct driving behavior clusters, our analysis offers valuable insights for various applications in the realm of transportation and road safety. Furthermore, these results underscore the importance of personalized approaches in promoting safer and more efficient driving practices, ultimately contributing to goal of improving road safety and enhancing the driving experience for all road users.

References

[1] R. Tavenard et al., ”Tslearn, A Machine Learning Toolkit for Time Series Data,” *Journal of Machine Learning Research*, vol. 21, no. 118, pp. 1-6, 2020.