An Inverse Reinforcement Learning Approach for Customizing Automated Lane Change Systems

Jundi Liu, *Student Member, IEEE*, Linda Ng Boyle, *Member, IEEE*, and Ashis G. Banerjee, *Senior Member, IEEE*

Abstract—**Vehicle automation seeks to enhance road safety and improve the driving experience. However, a standard system does not account for variations in users and driving conditions. Customizing vehicle automation based on users' preferences aims to improve the user experience and adoption of the technologies. This study introduces a systematic paradigm that starts with naturalistic driving data to identify the driving behaviors and styles for a customized automated lane change system. The driving behaviors are first extracted using Multivariate Functional Principal Component Analysis (MFPCA) with minimum prior expert knowledge. The driving styles are identified by clustering the extracted driving behaviors. An Inverse Reinforcement Learning (IRL) algorithm is then used to train the automated lane change system from grouped demonstrations of the identified driving styles to capture the preferences of a group of drivers with a similar driving style. The performance of the proposed customized automated lane change system is compared to (1) a non-customized system trained on all the sample trips, (2) customized systems built on expert-coded reward functions, and (3) customized systems trained using a Generative Adversarial Imitation Learning (GAIL) algorithm. The results show that our method outperforms all the other systems with respect to the prediction accuracy of the lane change actions. Additionally, our method gains insights on the representative behaviors of different driving styles to enable customization of automated lane change systems.**

Index Terms—**Customized automated lane change systems, driving style identification, inverse reinforcement learning, multivariate functional principal component analysis.**

I. Introduction

THE field of vehicle automation is evolving rapidly. Although fully autonomous vehicles are not expected to be publicly available for several more years, most vehicles are equipped with some degree of vehicle automation to assist in the driving tasks. Vehicles with automation technologies promise significant societal benefits, including decreases in car crashes, injuries and deaths, enhanced mobility, increased road efficiency, and better utilization of parking and lands [1]. Even though research has shown substantial potential benefits of vehicle automation, the acceptance of these technologies is still not apparent, partly due to the significant increase in the variety of driving conditions and potential users.

Non-customized vehicle automation with a moderate driving style may be too conservative for more aggressive drivers and too aggressive for more passive drivers [2]. Some systems are customized to account for users' preferences using an in-vehicle interface [3], [4]. However, an effective interface design would require extensive usability studies with multiple real-world scenarios. Another approach is not based on user testing but rather on observing driving styles based on existing data to capture driver's behavior [5].

The purpose of this study is to demonstrate the ability to customize lane change systems based on naturalistic data made available to the study team from the University of Michigan Transportation Research Instiute (UMTRI). The system is targeted toward a Society of Automotive Engineers (SAE) Level 2 system. The data provide insights on various driving styles for multiple drivers; this is an advantage over other models [2], [6] that focus only on one driver.

The framework consists of three components (see Fig. 1): a driving style identification method using Multivariate Functional Principal Component Analysis (MFPCA) and clustering analysis to generate grouped demonstrations for each driving style; automated lane change systems trained with an Inverse Reinforcement Learning (IRL) method using the previously identified driving style demonstrations; and testing and validation of the proposed framework. We support the existence of the resulting driving styles with literature from behavioral psychology, and the interpretability of our driving style identification method ensures that the grouped demonstrations capture the representative behaviors of each driving style. In addition, the explicit forms of the reward functions extracted from the IRL method provide insights into the resulting behaviors of the customized automated lane change systems. The testing and validation highlight the improved action prediction performance of the proposed framework. In summary, the contributions of this work are:

Manuscript received 7 December 2021; revised 2 April 2022; accepted 9 May 2022. Date of publication 31 May 2022; date of current version 19 September 2022. This work was supported by the National Science Foundation (NSF) under Grant CPS 1739085. The review of this article was coordinated by Prof. Mahdi Imani. *(Corresponding author: Linda Ng Boyle.)*

Jundi Liu is with the Department of Industrial & Systems Engineering, University of Washington, Seattle, WA 98195 USA (e-mail: jundiliu@uw.edu).

Linda Ng Boyle is with the Department of Industrial & Systems Engineering Department of Civil & Environmental Engineering, University of Washington, Seattle, WA 98195 USA (e-mail: linda@uw.edu).

Ashis G. Banerjee is with the Department of Industrial & Systems Engineering Department of Mechanical Engineering, University of Washington, Seattle, WA 98195 USA (e-mail: ashisb@uw.edu).

This work involved human subjects in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Research Board at the University of Michigan, as described in the US DOT Report No. DOT HS 812 171.

This article has supplementary downloadable material available at https://doi.org/10.1109/TVT.2022.3179332, provided by the authors.

Digital Object Identifier 10.1109/TVT.2022.3179332

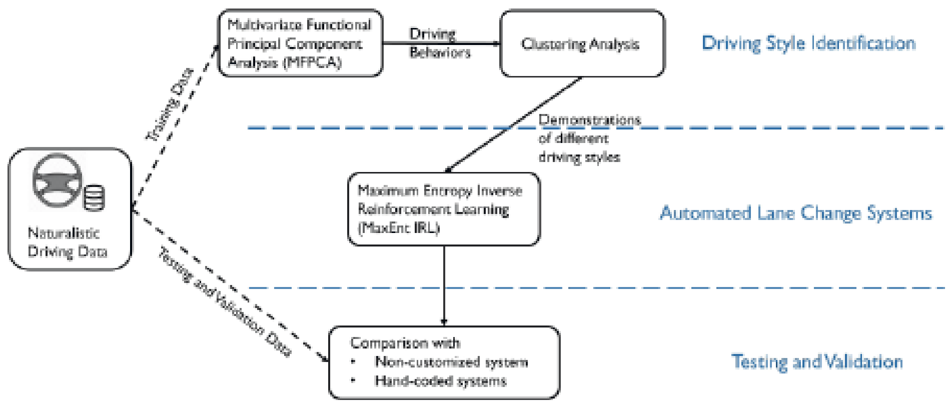


Fig. 1. The overall framework of the proposed method.

1) extraction of driving behaviors from naturalistic data without expert knowledge;
2) identification of driving styles;
3) customization of automated lane change systems using driving styles; and
4) improved prediction accuracy of lane change actions for unobserved driving segments.

II. Related Work

Customizing a system based on the operator's capabilities and limitations is a data-driven approach. The premise of successful customization is identifying accurate and appropriate driving styles with representative driving behaviors using naturalistic driving data. In this section, we provide an extensive literature review of driving style identification. We then review the methods for customized automated lane change systems. The gaps observed in the literature are discussed in the context of the study motivation.

A. Driving Style Identification

Driving style is generally defined as the habitual ways drivers choose to drive [7]. It is crucial to accurately identify driving styles to help better design customized vehicle automation. However, driving style refers to all the activities performed by the driver, which contains many aspects. To date, a uniform method for quantifying or identifying driving styles has not been identified. However, qualitative definitions are available for some driving styles. This includes the actions that fall under the category of risky driving defined by the National Highway Traffic Safety Administration (NHTSA): drunk driving, distracted driving, speeding, frequent brakes, tailgating, etc. (https://www.nhtsa.gov/risky-driving). Some qualitative criteria can be transformed into quantitative measurements using continuous sensory data such as speed and distances to the surrounding vehicles. The speed profile provides information on the fluctuations and peaks related to aggressive or conservative driving styles [8]. The distance to the surrounding vehicles provides the behavior related to tailgating and abrupt lane changes. As long as these patterns are extracted in a reasonable manner, we can identify an effective driving style. A survey by Martinez *et al.* [9] shows the typical process of recognizing meaningful driving styles. A figure summarizing the current driving style identification methods is presented in Section I of the supplementary material.

Driving behaviors are typically captured using continuous sensory data from the vehicles. For example, speed profiles are collected over time and can, therefore, be examined using time series analysis methods [10]–[13]. Speed profiles can also be examined using Functional Data Analysis (FDA). As compared to time series analysis methods, FDA does not require the data points to be evenly spaced and can model both temporal and spatial trends present in the data. In the past, FDA was attempted by [14] to examine the drivers' speed profiles in naturalistic driving data collected on rural two-lane curves. The results provide insights on driver behaviors for the individual road segments of different curvatures.

Once driving behaviors are successfully identified as constructed features, we can use supervised or unsupervised learning to distinguish the different driving styles. Supervised learning requires labeling the observations in the training dataset in advance [15]–[19]. Since naturalistic driving data is fairly noisy, the labeling process would be inaccurate and biased by expert knowledge and personal judgments about the driving styles. The subjective labels can lead to underfitting or overfitting certain driving styles. Unsupervised learning, on the other hand, can assign each observation to a cluster based on the measured distances among the observations [20]–[23]. Although it does not require extra effort to label the data, it might generate latent driving styles that were not previously known. For this reason, some manual identification or naming convention will be needed after the groups (or styles) are determined. Therefore, an ideal driving style identification method will need to include (1) automated feature extraction without complicated feature engineering and (2) selection using prior assumptions and interpretable clusters with minimum expert knowledge.

B. Customization in Automated Lane Change Systems

Personalization and customization have similar meanings and are sometimes used interchangeably. In this paper, customization refers to the adaptation of automation technologies to driving styles, whereas personalization captures the detailed individual preferences in driving tasks. Hence, we adopt the former term in our study but review the literature related to both in terms of vehicle automation.

The primary approach for designing a customized vehicle automation system is learning from demonstrations [5]. Here, the demonstrations refer to the operation data collected from the driving tasks. The widely adopted method is to observe human driving tasks and learn the behaviors using Inverse Reinforcement Learning (IRL) [24] or Imitation Learning (IL) [25], [26]. Multiple studies adopted variants of these methods for personalized vehicle automation [27]–[30].

For automated lane change systems, Butakov *et al.* [2] developed a method for modeling individual driver behaviors during lane changes. They used a sinusoidal lane change kinematic model and a Gaussian Mixture Model (GMM) to adjust the parameters to fit individual driving styles. However, the framework was validated on only three participants with 717 lane changes. The relatively high number of lane change behaviors for each participant may be inefficient when applied to new users, and the limited number of participants can compromise the evaluation of the proposed method. In another study, Vallon *et al.* [6] proposed a data-driven classifier to predict the moment a lane change initiation occurs while also considering the preferences of the driver. This decision logic was then integrated into a model predictive control framework for lane change control. However, the limited number of participants impacts the validity of the model.

In another study, Zhu *et al.* [31] proposed a Personalized Lane-Change Warning System based on three components: surrounding vehicle information, the predicted safety distance based on the driving style, and a lane change feasibility judgment system. However, these modules were rule-based decision-making systems that did not learn or adapt to changes in the driver behaviors. Another recent work constructed a customized lane-change assistance system using deep learning and spatial-temporal modeling [8]. The results on real-world driving data showed that the lane-change model is capable of learning latent features and achieved better performance than the current system. However, stacking three sophisticated black-box algorithms weakened the interpretability of the model, and the reasoning process for the action became unclear to the user. In a very recent study, Huang *et al.* [3] identified the user's preferences using the fuzzy linguistic preference relation method, including measurements on driving safety, ride comfort, vehicle stability. However, the preferences need to be collected interactively with the passengers using a touch screen rather than passively identified from naturalistic driving data.

III. Methods

Customized automated lane change systems are built using a two-step method. The first step identifies the driving styles with representative behaviors from naturalistic driving data by clustering the driving trips into different groups. The second step customizes the automated lane change systems to the identified driving styles by applying the IRL method to the clustered trips (demonstrations). The IRL algorithm learns the latent preferences and objectives in the form of reward functions from the demonstrations. The reward functions encode the underlying intents in the lane change tasks, which ensures better performance than directly imitating the lane change behaviors. To achieve task-level decision making for lane changes using IRL, we need sequential lane change demonstrations for each driving style. Therefore, it is essential to identify the appropriate driving styles at the trip-level.

A. Driving Style Identification Method

The current gaps in driving style identification (as discussed in Section II-A) are addressed using MFPCA and cluster analysis. Naturalistic driving data is modeled as functional data to characterize the driver's behavior at the trip level. Such modeling can be used to identify the underlying patterns behind the driver behaviors and requires less prior knowledge and effort to calculate the characteristic features. These features are then used as inputs to identify the most relevant clusters for all the trips that lead to improved interpretability of the clustered driving styles. Finally, each cluster is labeled with an appropriate driving style based on similar behaviors.

1) Multivariate Functional Principal Component Analysis: The naturalistic driving data are collected continuously over time. Functional Data Analysis (FDA) is a collection of effective methods for analyzing functional data. It considers each individual datum as a function and does not impose any particular assumptions on the independence of the different values within a functional datum [32]. It aims to detect the important patterns and explain the reasons behind the variations in the data [33]. These characteristics make FDA an appropriate method to extract features from naturalistic driving data.

Specifically, we use the Basis Expansion model to represent the functional data:

$$x(t) = \sum_{i=1}^{K} c_j \phi_j(t) = \Phi(t)^T \mathbf{c} \tag{1}$$

where $x(t)$ is a single measurement in a continuum t, $\Phi(t)$ is a basis system, and $\mathbf{c}$ is the coefficient vector. Commonly-used basis systems include Fourier, B-spline, and Polynomial. However, they require us to define parameters, such as knots and orders, in advance. To address this issue, Functional Principal Component Analysis (FPCA) represents functional data in the most parsimonious way by using the eigenfunction basis to explain most of the variation. As the name suggests, FPCA is a variation of the traditional PCA in a functional basis system instead of a coordinate system.

An extension of univariate FPCA to the multivariate case is used to examine multiple dimensions of the sensory data (i.e., speed and relative locations of the surrounding vehicles). Happ and Greven [34] provide an algorithm to estimate the multivariate Principal Components (PCs) based on their univariate counterparts using the following steps.

1) Calculate the PC functions and scores using the Karhunen–Loève decomposition on the covariance surface.
2) Estimate the joint covariance matrix by combining all the PC scores into one large matrix.

3) Find the eigenvectors and eigenvalues of the estimated joint covariance matrix.
4) Calculate the estimated multivariate PC functions and scores by adjusting the univariate PC functions and scores from step 1 using the eigenvectors and eigenvalues estimated in step 3.

2) Clustering Analysis: A k-means clustering algorithm is applied to the scores of the functional PCs. The k-means algorithm computes the centroid of each cluster iteratively and updates the assignment of each observation. While converging to stable assignments, k-means finds the clusters that minimize the within-cluster variances. There have been previous attempts to implement k-means clustering on functional PCs scores, and results confirm the effectiveness of the method [35]. Driving style identification is achieved once each observation is assigned to an appropriate cluster.

B. Customized Automated Lane Change System

The driving style identification step clusters the driving trips into different driving styles. In the second step, we leverage the grouped driving demonstrations generated by the previous step and adapt a data-driven method to customize the automated lane change systems. Each customized system is trained separately from the driving demonstrations.

The vehicle automation system can be regarded as an intelligent agent that can be trained using Reinforcement Learning (RL). RL provides the learning ability to make sequential decisions in order to improve the agent's experience through interactions with its environment instead of explicit programming. This is achieved by maximizing the expected cumulative reward of the agent's actions, where the environment is typically modeled as a Markov Decision Process (MDP). An MDP is defined by four components: a set of states S, a set of actions A, a reward R, and a state transition probability P. The state is a collection of variables that characterize the system (agent and its environment) at any point in time, and the action is the decision made by the agent to transition from one state to the next to maximize the cumulative reward over a given time horizon. Here, the state is characterized by the variables that affect lane change decisions such as the speed of the ego car and its distance to the surrounding vehicles.

RL has been shown to be immensely successful in a wide variety of applications. However, unlike most other applications, no specific reward signals are directly generated from the driving tasks in our study. We therefore used an Inverse Reinforcement learning (IRL) algorithm to extract the specific form of the reward function.

The IRL algorithm can enhance the interpretability of the policy and behaviors, but can also lead to degeneracy and multi-optimality. In other words, it is not easy to choose the optimal reward function without some premise. To address this issue, we consider the Maximum Entropy Inverse Reinforcement learning (MaxEnt IRL) [36]. There are a few definitions introduced by the MaxEnt IRL algorithm:

- $\zeta = \{(s_1, a_1), (s_2, a_2), (s_3, a_3), \dots\}$ is a demonstrated trajectory, where s_i and a_i are the state and action at time instant i, respectively.
- $\theta \in \mathbb{R}^k$ is the reward function parameters. The reward function is modeled as $r(s) = f(s, \theta)$.
- $P(\zeta)$ is the probability of observing the trajectory ζ.
- $P(s)$ is the probability of visiting the state s, also called state visitation frequency.
- $R(\zeta)$ is the reward of a trajectory ζ.

The probability of a trajectory demonstrated by the expert (in our case, the human driver) has to be exponentially greater for larger rewards as compared to smaller rewards. Mathematically,

$$P(\zeta) \propto e^{R(\zeta)}. \tag{2}$$

The MaxEnt IRL algorithm solves for optimal θ^* by maximizing the likelihood of the demonstrated trajectories,

$$\begin{aligned}\theta^* &= \arg\max_\theta L(\theta) \\ &= \arg\max_\theta \frac{1}{M}\sum_{\zeta_D} R(\zeta_D) - \log\sum_{\zeta} e^{R(\zeta)}.\end{aligned} \tag{3}$$

Here, M is the number of demonstrated trajectories, ζ_D is the observed trajectory from demonstration, and ζ is any trajectory that can be observed from the task. Since the above objective function is convex, an optimal solution is obtained using any gradient descent based method.

The MaxEnt IRL algorithm handles expert suboptimality and stochasticity by operating on the distribution over possible trajectories. The reward function is estimated using any form of function approximation. In our work, the driving task is complex and nonlinear functions are needed to approximate the reward. Hence, we adopt Maximum Entropy Deep Inverse Reinforcement Learning (DeepIRL) [37], which uses a neural network to approximate the reward function. Since the gradients are computed efficiently using back-propagation for neural networks, the optimal reward function parameter θ^* is obtained by gradient descent-based algorithms.

IV. Naturalistic Driving Data

The dataset used for this study comes from the UMTRI Safety Pilot Model Deployment (SPMD) naturalistic driving data [38]. The data collection period was from August 2012 to June 2017. This dataset is a part of the Connected Vehicle Safety Pilot Program, a research initiative that features real-world implementation of connected vehicle safety technologies, applications, and systems using everyday drivers. SPMD data includes approximately 3000 vehicles (some with more than one driver) equipped with vehicle-to-vehicle (V2V) communication in real-world conditions at a test site with diverse traffic [38].

Fig. 2 shows the layout of the test site and the location of the roadside equipment used in data collection. There were five data acquisition systems that worked cooperatively to record four sets of data: contextual (e.g., weather, lighting), driving (from a vehicle's data acquisition system (DAS)), basic safety messages (BSM), and data from roadside equipment (RSE). The SPMD data contains more than four terabytes of raw data. We used the publicly available data, which covered a 7-month period from October 2012 to April 2013 (https://catalog.data.gov/sv/dataset/

Fig. 2. Safety pilot model deployment site plan in Ann Arbor, Michigan. Modified from [38]. Note: Red circles denote the approximate starting (1) and ending (2) locations of the trips for the highway segment of interest (red solid line).

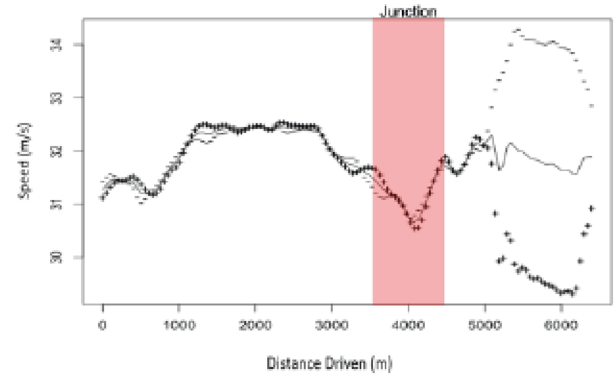


Fig. 3. The effect of functional PC 2 on the average speed profile. FPC 2 explains 21.1% of the variability in the clockwise direction. Note: The solid line denotes the average driving profile of all the trips (coefficient=32.2). A coefficient greater than 32.2 will move the speed profile closer to the "+" line; otherwise, the profile is closer to the "-" line. More specifically, the "+" and "-" lines are generated by adding or subtracting two standard deviation of the functional PC from the average speed profile, respectively.

safety-pilot-model-deployment-data). This dataset does not include any personal identifiers and is exempt from additional IRB review. The original IRB process is described in [38].

The road environment varied from low-speed urban roads to high-speed freeways. For our analysis, a segment of the freeway driving route was selected. It starts on location 1 and ends on location 2 (shown in red in Fig. 2). The trips in the counterclockwise direction were also selected for the same freeway segment, which starts on location 2 and ends on location 1. The selected route is approximately 7 kilometers with 3.5 minutes of travel time. Note that the selected segment has a junction from U.S. Route 23 to U.S. Route 14 in the elbow transition part. Specifically, we filtered the trips based on the GPS coordinates from the DAS to select the road segments of interest. We then extracted four variables for analysis: (1) speed of the ego car, (2) distance to the vehicle in front, and (3) average distances to the vehicles on the left and (4) the right. Previous studies have shown that these features reflect the driving behaviors and significantly affect the lane change decisions [8], [39]. As a result, a total of 105 trips, including 46 in the clockwise direction and 59 in the counterclockwise direction, were selected for further analysis.

V. Results

The experiments were carried out on a MAC Big Sur version 11.6 Operating System using R programming language version 3.6.1 and Python version 2.7. A 2.6 GHz Intel Core i7 quad-core processor with 16 GB 2133 MHz LPDDR3 RAM was used. R packages mfpca [40] and other supportive packages were employed for implementing MFPCA. Python packages Theano [41] and numpy [42] were used for implementing IRL algorithm.

A. Driving Style Identification

Four variables from the naturalistic driving data were selected to identify the driving styles. The speed profiles provided data on speeding, and changes in speed over time (i.e., acceleration/deceleration). The distances to the surrounding vehicles were used to identify aggressive driving behaviors (e.g., tailgating and abrupt lane changes). We then applied the MFPCA method to extract the driving behaviors.

For the clockwise direction, the top four FPCs are able to explain 92.3% of the total variability. The distances to the surrounding vehicles were missing in many observations. Hence, most of the variability came from the speed profile, and only a small amount was from the distance to the vehicles in front. The functional PC 2 is shown in Fig. 3. The corresponding coefficient of the average speed profile for this functional PC is 32.2. Therefore, a coefficient greater than 32.2 indicates that the trip has a lower than average speed and faster deceleration than average trips; otherwise, the trip has the opposite behavior. Similarly, other driving behaviors can be revealed for the other functional PCs:

1) Functional PC 1: Variability is observed before entering the junction segment. A coefficient larger than the average speed profile for FPC1 indicates that the driver tends to have a greater than average speed and faster than average deceleration before entering the junction.
2) Functional PC 2: Variability is observed at the long open road after the junction. A coefficient larger than the average speed profile for FPC2 (in this case 32.2) corresponds to lower than average speed and larger than average distance to the vehicle in front.
3) Functional PC 3: Variability is spotted when entering the freeway. A coefficient larger than the average speed profile for FPC3 means the driver has a slower than average speed and prefers larger than average distance to the vehicle in front.
4) Functional PC 4: Variability is observed before entering the junction segment. A coefficient larger than the average speed profile for FPC4 denotes greater speeding behavior on the open road before entering the junction segment.

We also applied k-means clustering on the extracted MFPCA scores. The silhouette analysis showed that three clusters

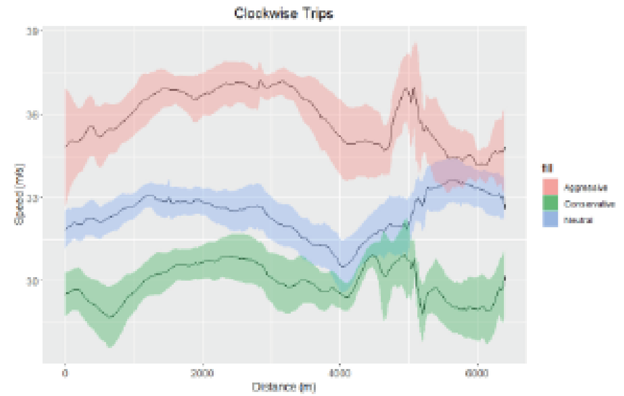


Fig. 4. Three clusters for the clockwise direction: each cluster is depicted by its mean speed and the corresponding 95% confidence interval.

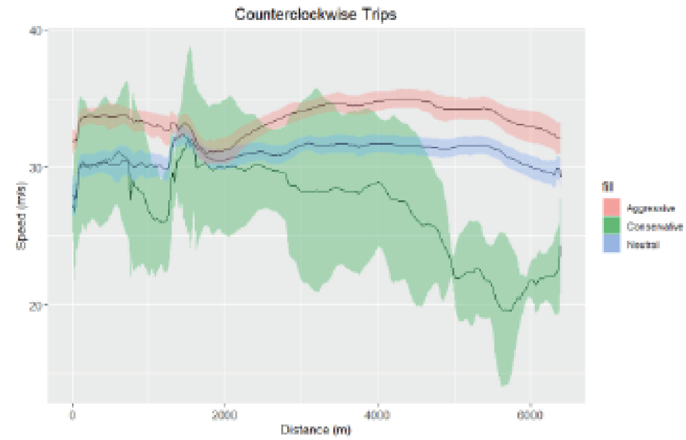


Fig. 5. Three clusters for the counterclockwise direction: each cluster is depicted by its mean speed and the corresponding 95% confidence interval.

provided the largest average silhouette coefficient (0.56) among clusters varying from two to six. This is consistent with studies from traffic psychology [43], [44] that show three broad categories of driving styles: aggressive, neutral, and conservative.

We visualized the three resulting clusters using the speed profile in Fig. 4 since it encodes the most variability in the data. The clusters are labeled as aggressive trips, neutral trips, and conservative trips. The aggressive trips tend to have higher speeds, faster accelerations, and a small distance to the front vehicles on the open road after and in the junction segments. The conservative trips demonstrate the opposite behaviors, and the neutral trips lie between the two extremes.

The same analysis steps were conducted for the counterclockwise direction. The top four functional PCs explained 86.1% of the total variability. The same issue with missing values in the distances to the surrounding vehicles was observed. The top four FPCs represented the following behaviors:

1) Functional PC 1: Variability is observed when leaving the highway. A coefficient larger than the average speed profile for FPC1 indicates that the driver tends to have a relatively high speed when leaving the highway.
2) Functional PC 2: Variability is observed before the junction segment. A coefficient larger than the average speed profile for FPC2 denotes fast speed, rapid deceleration, and small distance to the vehicle in front.
3) Functional PC 3: Variability is spotted on the long open road after the junction segment. A coefficient larger than the average speed profile for FPC3 indicates speeding behavior in the long open road after the junction.
4) Functional PC 4: Variability is spotted in the open road after the junction. A coefficient larger than the average speed profile for FPC4 indicates that the driver is braking first after the junction, but accelerating fast thereafter. Otherwise, the driver accelerates fast at first, but then slows down.

Fig. 5 summarizes the speed profiles of the three clusters identified by the k-means clustering method for the counterclockwise direction trips. Since the conservative group contains a small number of trips, its confidence intervals are wider than the other two groups.

In summary, out of 105 total trips, 27 are classified as aggressive, 56 as neutral, and 22 as conservative. These behaviors, as revealed by the spatial and temporal characteristics of the driving trips using MFPCA, match the qualitative description of the corresponding driving styles discussed in Section II-A. Previous studies have also shown that these factors significantly impact lane change decision making [39], [45]. Therefore, the results confirm that our driving style identification method is able to extract representative driver behaviors, and clustering analysis successfully separates the different driving styles with minimal feature engineering and prior knowledge about the driving styles.

B. *Customized Automated Lane Change Systems*

We customized the automated lane change systems using the IRL method, wherein three separate systems were trained on the clusters generated by the driving style identification method. Specifically, three reward functions were learned from the clustered demonstrations to capture the preferences of the different driving styles. Accordingly, the lane change systems followed the optimal (stochastic) policies corresponding to the reward functions to adapt or customize to the different driving styles.

The lane change system is considered one of the most challenging vehicle automation problems, since it involves both longitudinal and lateral vehicle control, and requires constant monitoring of the surroundings. For this reason, we selected the same four features for the IRL method: the speed of the ego vehicle, the distance to the lead vehicle, the average distance to the right vehicles, and the average distance to the left vehicles.

The IRL algorithm is computationally expensive in recovering reward functions and stochastic policies. The surrounding vehicle information has a high percentage of missing values due to noisy and imperfect naturalistic driving data. Therefore, we discretize the continuous state space into a discrete state space. Each continuous variable is discretized into four levels based on the first, second, and third quantile values along with an additional level to encode either missing information or no car

TABLE I
DISCRETIZATION OF THE CONTINUOUS VARIABLES: SPEED, FORWARD DISTANCE, LEFT DISTANCE, AND RIGHT DISTANCE

Var. Names	Speed	Forward Dist.	Left Dist.	Right Dist.
List of Levels	Slow Medium Fast Very Fast	No Car Very Small Small Medium Large	No Info Very Small Small Medium Large	No Info Very Small Small Medium Large

Note: The 'No Info' level corresponds to a situation with no car or no lane information.

in a given direction. The detailed numerical intervals for each level are provided in Table VIII of the supplementary material. As a result, the state space contains 500 ($4 \times 5 \times 5 \times 5$) distinct states and 19 ($4 + 5 + 5 + 5$) feature variables (see Table I). The action space comprises three actions: staying in the current lane, changing to the right lane, or changing to the left lane.

The transition probability is obtained by taking the maximum likelihood estimate of the transition matrix [46]. However, not all states are guaranteed to be observed in the demonstrated trajectories. To account states with no observations, we adopt the common ϵ-greedy ($\epsilon = 0.1$) algorithm to get a trade-off between exploration and exploitation. The experiments are carried out using a Feed-forward Neural Network (FNN) with two hidden layers of 30 hidden units in each layer. We select the discount factor as 0.9, and the learning rate as 0.01. The reward values are normalized to a continuous scale between -1 and 1. The results are presented for the recovered reward functions and stochastic policies. The optimal policies are recovered using approximate value iteration in each update of the reward function.

Summary statistics of the reward functions for four representative driving scenarios are shown in Table II. We selected these scenarios to demonstrate the interpretability and effectiveness of the IRL method. In the first three scenarios, we fix the ego vehicle speed and the distance to the forward vehicle, and show the summary statistics as a result of varying the other two variables (average distance to the left vehicles and right vehicles). When the ego vehicle is slow and there is no lead vehicle, the aggressive drivers have the least number of non-negative reward values and the conservative drivers have the most number of such values. The reward values for the neutral drivers lie in between those for the aggressive and conservative drivers. Moreover, the conservative drivers have the largest minimum, median, and mean reward values, which means that they are more comfortable staying in this scenario as compared to the other two driving groups.

On the other hand, the scenario with fast ego vehicle speed and very small forward distance is preferred by the aggressive driver group. Likewise, the scenario with the medium ego vehicle speed and large forward distance is preferred by the neutral driver group. We also fixed the average distance to the right vehicles and left vehicles as medium, and examined the reward values of the other two variables. In this scenario, the neutral drivers have the largest minimum, median, mean, and maximum reward values. Thus, we observe that the optimal policies obtained by the learned reward function facilitate model interpretability.

The recovered policies are used to validate the proposed IRL method by comparing the behaviors associated with the different driving styles. Tables III, IV, V present the stochastic policy calculated for aggressive drivers, neutral drivers, and conservative drivers based on the recovered reward functions, respectively. The results show that the aggressive drivers prefer to switch lanes in more than half of the scenarios. The optimal actions would still maximize safety by choosing actions with the highest probabilities. For example, although the probability of changing to left or right lanes in the small distance scenarios is 0.286 for aggressive drivers (see Table III), the optimal action is to stay in the current lane since it has the highest probability of 0.429. Alternatively, the conservative drivers are quite comfortable traveling at a low speed in the current lane even when there is no lead vehicle. Finally, the neutral drivers prefer to stay in the current lane for most scenarios, but are willing to switch to other lanes in a few cases. To better demonstrate the performance of the recovered optimal policies, additional policy tables are provided for a second scenario of fast driving speed and a very small distance to the lead vehicle in the supplementary material.

C. Performance Evaluation

The performance of the proposed customized automated lane change system is compared to (1) a non-customized system trained on all the sample trips, (2) customized systems built on expert-coded reward functions, and (3) customized systems trained using a Generative Adversarial Imitation Learning (GAIL) algorithm.

The non-customized system is expected to have an average driving style across all the sample trips. The optimal policy is trained using the same IRL method on all the trips. Specifically, we design three expert-coded reward functions, one for each driving style group, using an approach similar to An *et al.* [47] and Hoel *et al.* [48]. Their primary approach is to balance the efficiency and safety of automated lane change systems. To maximize safety, the reward function is purposely coded to give a large negative reward if the vehicle lands in a collision or near-collision state. We define the near-collision state as being very close to the surrounding vehicles and use similar reward values (i.e., -5, -10, and -20 for aggressive, neutral, and conservative drivers, respectively). An additional negative reward is designated if the system tries to choose an action that results in a near-collision state (i.e., -1.5, -3, and -6, respectively). In terms of efficiency, a positive reward is given such that it is proportional to either the driving distance or the average speed between two consecutive actions. Since we discretize speed into 4 levels and the data is recorded at 10 Hz, we assign the same positive reward for any particular speed level in each of three driving styles, i.e., 1, 3, 5 and 7 for slow, medium, fast and very fast speed, respectively.

Compared to the learned reward function using IRL, the expert-coded reward depends much less on the state levels. It assigns the same reward value for every near-collision state. However, the learned reward function distinguishes between the different near-collision states. For instance, if the distance from the vehicle on the right is very small, but the distance from the vehicle on the left is large, the reward value is positive, whereas the expert-coded reward function generates negative values.

TABLE II
SUMMARY STATISTICS OF THE REWARD VALUES IN FOUR EXAMPLE SCENARIOS

Scenarios (variable: level)	Driving Style	Non-negative portion*	Min	Median	Mean	Max	SD
Speed: Slow Forward Dist.: No Car	Aggressive	12/25	-0.418	-0.014	0.045	**1**	0.275
	Neutral	15/25	**-0.059**	0.074	0.048	0.398	0.255
	Conservative	**21/25**	-0.114	**0.210**	**0.285**	**1**	**0.318**
Speed: Fast Forward Dist.: Very Small	Aggressive	**8/25**	**-0.101**	**-0.037**	**-0.004**	**0.403**	0.111
	Neutral	7/25	-0.269	-0.046	-0.048	0.158	0.080
	Conservative	5/25	-0.343	-0.174	-0.120	0.293	**0.172**
Speed: Medium Forward Dist.: Large	Aggressive	11/25	**-0.087**	-0.018	0.044	**0.520**	0.165
	Neutral	**18/25**	-0.279	**0.069**	**0.065**	0.357	0.128
	Conservative	10/25	-0.489	-0.079	-0.077	0.320	**0.203**
Left Dist.: Medium Right Dist: Medium	Aggressive	7/16	-0.137	**-0.048**	-0.001	0.281	0.114
	Neutral	**8/16**	**-0.108**	**-0.048**	**0.009**	**0.326**	0.109
	Conservative	3/16	-0.324	-0.146	-0.129	0.160	**0.118**

Note: *Non-negative values indicate that the driver is more likely to stay in the scenario. **Bold** represents the largest value for a particular scenario and driving style.

TABLE III
STOCHASTIC POLICY OF THE AGGRESSIVE DRIVERS FOR SLOW DRIVING WITHOUT ANY LEAD CAR

Dist. to Left Vehicles	Dist. to Right Vehicles	Probability of Action		
		Stay in Current Lane	Change to Right Lane	Change to Left Lane
No info	No info	0.476	0.313	0.211
No info	Very small	0.152	0.424	0.424
No info	Small	0.323	0.292	0.385
No info	Medium	0.26	0.288	0.452
No info	Large	0.266	0.718	0.016
Very small	No info	0.634	0.221	0.144
Very small	Very small	0.429	0.286	0.286
Very small	Small	0.296	0.352	0.352
Very small	Medium	0.355	0.171	0.474
Very small	Large	0.133	0.433	0.433
Small	No info	0.273	0.63	0.097
Small	Very small	0.499	0	0.501
Small	Small	0.47	0.265	0.265
Small	Medium	0.346	0.327	0.327
Small	Large	0.36	0.269	0.371
Medium	No info	0.34	0.268	0.391
Medium	Very small	0.364	0.318	0.318
Medium	Small	0.843	0.078	0.078
Medium	Medium	0.571	0.215	0.215
Medium	Large	0.42	0.29	0.29
Large	No info	0.522	0.2	0.278
Large	Very small	0.285	0.357	0.357
Large	Small	0.38	0.42	0.2
Large	Medium	0.491	0.255	0.255
Large	Large	0.119	0.055	0.826

Note: Blue shade denotes the optimal action.

TABLE IV
STOCHASTIC POLICY OF THE NEUTRAL DRIVERS FOR SLOW DRIVING WITHOUT ANY LEAD CAR

Dist. to Left Vehicles	Dist. to Right Vehicles	Probability of Action		
		Stay in Current Lane	Change to Right Lane	Change to Left Lane
No info	No info	0.546	0.407	0.047
No info	Very small	0.406	0.332	0.262
No info	Small	0.398	0.172	0.43
No info	Medium	0.527	0.267	0.206
No info	Large	0.432	0.363	0.206
Very small	No info	0.266	0.389	0.345
Very small	Very small	0.346	0.252	0.402
Very small	Small	0.509	0.436	0.054
Very small	Medium	0.147	0.62	0.233
Very small	Large	0.322	0.576	0.102
Small	No info	0.683	0.154	0.162
Small	Very small	0.367	0.42	0.213
Small	Small	0.424	0.288	0.288
Small	Medium	0.39	0.343	0.267
Small	Large	0.378	0.444	0.177
Medium	No info	0.57	0.43	0
Medium	Very small	0.662	0.168	0.171
Medium	Small	0.493	0.253	0.253
Medium	Medium	0.696	0.017	0.287
Medium	Large	0.718	0.141	0.141
Large	No info	0.452	0.407	0.141
Large	Very small	0.279	0.23	0.491
Large	Small	0.549	0.221	0.23
Large	Medium	0.718	0.141	0.141
Large	Large	0.783	0.109	0.109

Note: Blue shade denotes the optimal action.

TABLE V
STOCHASTIC POLICY OF THE CONSERVATIVE DRIVERS FOR SLOW DRIVING WITHOUT ANY LEAD CAR

Dist. to Left Vehicles	Dist. to Right Vehicles	Probability of Action		
		Stay in Current Lane	Change to Right Lane	Change to Left Lane
No info	No info	0.63	0.326	0.044
No info	Very small	0.425	0.287	0.287
No info	Small	0.359	0.321	0.321
No info	Medium	0.521	0.246	0.233
No info	Large	0.544	0.228	0.228
Very small	No info	0.951	0.025	0.025
Very small	Very small	0.691	0.155	0.155
Very small	Small	0.506	0.247	0.247
Very small	Medium	0.397	0.302	0.302
Very small	Large	0.501	0.25	0.25
Small	No info	0.973	0.014	0.014
Small	Very small	0.592	0.131	0.277
Small	Small	0.753	0.124	0.124
Small	Medium	0.709	0.146	0.146
Small	Large	0.758	0.121	0.121
Medium	No info	0.715	0.199	0.086
Medium	Very small	0.415	0.2	0.385
Medium	Small	0.403	0.299	0.299
Medium	Medium	0.452	0.274	0.274
Medium	Large	0.695	0.153	0.153
Large	No info	0.917	0.042	0.042
Large	Very small	0.935	0.032	0.032
Large	Small	0.576	0.212	0.212
Large	Medium	0.677	0.162	0.162
Large	Large	0.888	0.056	0.056

Note: Blue shade denotes the optimal action.

Lastly, we implemented a GAIL algorithm to learn the optimal policies from the demonstrated driving styles. The GAIL algorithm used in this study was adopted by Ho and Ermon [25], and was shown to be equivalent to the MaxEnt IRL algorithm for a specific form of the discriminator loss function [49]. However, this GAIL algorithm is a form of imitation learning that learns the optimal policies directly from the demonstrations without any specific form of the reward functions.

For brevity, we only present the stochastic policy of the conservative drivers in Table VI, where we fix the ego car speed to be slow and do not have any lead vehicle. A significant difference is observed between the optimal policies for the expert-coded and IRL methods. Table V suggests that the optimal policy for the IRL method is to always stay in the current lane. However, the optimal policy for the expert-coded reward function shows some unexpected behaviors. For example, it tries to change to the right lane even though there is already a very small distance to the vehicles on the right. Additionally, there are several states that have an equal probability of changing to the right and left lanes. It suggests the agent is not learning useful information from the

TABLE VI
STOCHASTIC POLICY OF THE CONSERVATIVE DRIVERS USING THE EXPERT-CODED REWARD FUNCTION FOR SLOW DRIVING WITHOUT ANY LEAD CAR

Dist. to Left Vehicles	Dist. to Right Vehicles	Probability of Action		
		Stay in Current Lane	Change to Right Lane	Change to Left Lane
No info	No info	0.592	0.408	0
No info	Very small	0	0.5	0.5
No info	Small	1	0	0
No info	Medium	0.05	0.95	0
No info	Large	1	0	0
Very small	No info	0	0.5	0.5
Very small	Very small	0	0.5	0.5
Very small	Small	0	0.5	0.5
Very small	Medium	0	0.5	0.5
Very small	Large	0	0.5	0.5
Small	No info	1	0	0
Small	Very small	0	1	0
Small	Small	1	0	0
Small	Medium	1	0	0
Small	Large	1	0	0
Medium	No info	0.069	0.078	0.853
Medium	Very small	0	1	0
Medium	Small	1	0	0
Medium	Medium	0.624	0.188	0.188
Medium	Large	1	0	0
Large	No info	1	0	0
Large	Very small	0	0.5	0.5
Large	Small	0.986	0.007	0.007
Large	Medium	0.996	0.002	0.002
Large	Large	0.931	0.035	0.035

Note: Blue shade denotes the optimal action.

expert-coded reward function. Lastly, most probabilities in the table are 0, 0.5, or 1, which means the optimal policy is not able to capture the stochasticity in the actions.

The policy tables for the aggressive and neutral drivers are provided in Section II of the supplementary material. For the aggressive policy recovered from the expert-coded reward function, there are more lane change actions than for the slow speed and no lead vehicle scenario, and most probabilities are not simply 0, 0.5 and 1 as compared to the policy table for the conservative drivers. However, a significant proportion of the scenarios (8/18) still assigns equal probabilities for changes to the right and left lanes. In comparison, the aggressive policy recovered from the learned reward function has a substantially smaller proportion of scenarios (4/14) that are unable to distinguish between the changes to the right and left lanes. Overall, the neutral policy shows comparable behaviors for the expert-coded and learned reward functions. However, a few problematic actions are observed in the policy of the expert-coded reward function. For example, in the slow speed and no lead vehicle scenario, if the left distance is large and right distance is very small, the optimal actions of the expert-coded reward and the learned reward are to change to the right lane and left lane, respectively. To summarize, our method maximizes safety and yields more reasonable driving behaviors.

Finally, we test the prediction accuracy of the actions of all the four systems on the unobserved trips of other highway segments for the same set of drivers identified by our driving style identification method. We have 368, 398, and 229 shorter trips for aggressive, neutral, and conservative drivers, respectively. The results are shown in Table VII. Our proposed method outperforms the other three systems in the three categories of the testing trips. More specifically, our customized automated lane change system with an aggressive driving style improved the prediction accuracy by 81.6%, 141.9% and 21.5%, respectively.

TABLE VII
PREDICTION ACCURACY FOR UNOBSERVED TRIPS

Methods	Prediction Accuracy (%)		
	aggressive	neutral	conservative
Non-customized	40.64	65.33	60.20
Customized using expert-coded rewards	30.52	61.67	74.4
Customized using GAIL	60.75	68.42	70.07
Customized using IRL (our method)	73.82	70.52	85.40

Note: The customized systems are tested for the corresponding trips with the same driving style.

In summary, our proposed method is capable of accurately predicting the driver actions for all the different driving styles, with best predictive power for the conservative drivers and largest improvement for the aggressive drivers.

VI. CONCLUSION

In this paper, we propose a two-step method for building customized automated lane change systems directly from naturalistic driving data. In the first step, our proposed method applies MFPCA for automated feature extraction and clustering analysis to identify driving styles. Our approach is able to reveal three distinct styles: aggressive, neutral, and conservative. The aggressive drivers tend to have the highest speed and smallest distance to the lead vehicles in open roads and the fastest acceleration when leaving the junction segment. On the other hand, the conservative drivers tend to decelerate very fast when entering the junction segment and maintain the lowest speed on the whole freeway segment and largest distance to the lead vehicles. Lastly, the neutral drivers have a smooth transition of speed when entering and leaving the junction segment. As for the open roads, the neutral drivers travel at speeds that are not as fast as the aggressive drivers but also not as slow as the conservative drivers. Our method, therefore, addresses the challenges of identifying driving styles quantitatively using driving behaviors. It also simplifies the interpretability of the results from unsupervised learning. In other words, minimal expert knowledge are needed to interpret the driving styles, which reduce excessive and potentially biased labeling efforts.

In the second step, we apply the IRL method to build customized lane change systems for each driving style. From the extracted optimal policies and reward functions, we observe that aggressive drivers prefer faster speeds, stay closer to the lead vehicles, and change lanes more frequently when compared to the neutral and conservative drivers. The neutral drivers prefer moderate speeds and do not stay too close or too far from the surrounding vehicles. They also adapt their speeds and locations by executing lane changes as required by their current situations. Lastly, the conservative drivers tend to minimize the number of lane changes by continuing in their current lanes and staying far away from the surrounding vehicles.

The performance of the customized lane change systems on unobserved trips was examined and compared to a non-customized system and two other customized systems. The results show the effectiveness of our proposed method and confirm that the learned reward function is more useful than its expert-coded counterpart. Our two-step method leveraged a clustering-based method to capture the behaviors of multiple drivers. Moreover, we improved the performance on real-world unobserved driving trips by using noisy naturalistic driving data instead of limited samples from a controlled driving environment. The value of using a two-step method is demonstrated through the computed rewards, which connects the identified driving styles in step one with the driving behaviors from the optimal stochastic policies in step two.

There are limitations of the proposed method. First, the naturalistic driving data can exhibit different driving behaviors for certain scenarios that are all within the safety threshold. For example, if an aggressive driver has to merge into one lane from a four-lane road, they need to take action regardless of their preferred driving style. This limitation can generate irrational behaviors in the training set, which will affect the reward values and optimal policies. However, the highest probabilities of the reasonable behavior (stay in the current lane) in the training set can, however, maximize safety when implementing the policy in future testing of the proposed systems. Second, the naturalistic driving data is purely observational and lacks experimental control. For this reason, some state transitions or state-action pairs may not be observed in the data. This exploration deficiency compromises the model performance. Third, the computational cost of the IRL method can be huge for long trajectories.

Future studies can consider more realistic and complex scenarios that better account for the continuous state and action spaces. A more robust method may also have to be considered to overcome the irrational behaviors observed in the training set. Nevertheless, in general, our approach is promising for customizing automated lane change systems to driving styles. While the data used in this study does not provide examples of all driving situations, we can expand on this approach to learn other meaningful driving behaviors.

ACKNOWLEDGMENT

Any opinions, findings, and conclusions or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of NSF.

REFERENCES

[1] D. V. McGehee *et al.*, "Review of automated vehicle technology: Policy and implementation implications.," Iowa. Dept. Transportation Tech. Rep. MATC-MU: 276, 2016.

[2] V. A. Butakov and P. Ioannou, "Personalized driver/vehicle lane change models for ADAS," *IEEE Trans. Veh. Technol.*, vol. 64, no. 10, pp. 4422–4431, Oct. 2015.

[3] C. Huang *et al.*, "Personalized trajectory planning and control of lane-change maneuvers for autonomous driving," *IEEE Trans. Veh. Technol.*, vol. 70. no. 6, pp. 5511–5523, Jun. 2021.

[4] A. Trende, D. Gräfing, and L. Weber, "Personalized user profiles for autonomous vehicles," in *Proc. 11th Int. Conf. Automot. User Interfaces Interactive Veh. Appl.: Adjunct Proc.*, 2019, pp. 287–291.

[5] M. Kuderer, S. Gulati, and W. Burgard, "Learning driving styles for autonomous vehicles from demonstration," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2015, pp. 2641–2646.

[6] C. Vallon, Z. Ercan, A. Carvalho, and F. Borrelli, "A machine learning approach for personalized autonomous lane change initiation and control," in *Proc. IEEE Intell. Veh. Symp.*, 2017, pp. 1590–1595.

[7] F. Sagberg, Selpi, G. F. Bianchi Piccinini, and J. Engström, "A review of research on driving styles and road safety," *Hum. Factors*, vol. 57, no. 7, pp. 1248–1275, 2015.

[8] Z. Gao, T. Sun, and H. Xiao, "Decision-making method for vehicle longitudinal automatic driving based on reinforcement Q-learning," *Int. J. Adv. Robotic Syst.*, vol. 16, no. 3, 2019, Art. no. 1729881419853185.

[9] C. M. Martinez, M. Heucke, F.-Y. Wang, B. Gao, and D. Cao, "Driving style recognition for intelligent vehicle control and advanced driver assistance: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 3, pp. 666–676, Mar. 2018.

[10] L. Mozaffari, A. Mozaffari, and N. L. Azad, "Vehicle speed prediction via a sliding-window time series analysis and an evolutionary least learning machine: A case study on san francisco urban roads," *Eng. Sci. Technol., An Int. J.*, vol. 18, no. 2, pp. 150–162, 2015.

[11] D. Wang, X. Pei, L. Li, and D. Yao, "Risky driver recognition based on vehicle speed time series," *IEEE Trans. Human-Mach. Syst.*, vol. 48, no. 1, pp. 63–71, Feb. 2018.

[12] Y. V. Tan, M. R. Elliott, and C. A. Flannagan, "Development of a real-time prediction model of driver behavior at intersections using kinematic time series data," *Accident Anal. Prevention*, vol. 106, pp. 428–436, 2017.

[13] W. Wang, J. Xi, and D. Zhao, "Driving style analysis using primitive driving patterns with Bayesian nonparametric approaches," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 8, pp. 2986–2998, Aug. 2019.

[14] B. Wang, "Modeling drivers' naturalistic driving behavior on rural two-lane curves," *Ph.D. dissertation*, Iowa State Univ., Ames, IA, USA, 2015.

[15] C.-C. Lin, S. Jeon, H. Peng, and J. Moo Lee, "Driving pattern recognition for control of hybrid electric trucks," *Veh. Syst. Dyn.*, vol. 42, no. 1/2, pp. 41–58, 2004.

[16] A. Wahab, C. Quek, C. K. Tan, and K. Takeda, "Driving profile modeling and recognition based on soft computing approach," *IEEE Trans. Neural Netw.*, vol. 20, no. 4, pp. 563–582, Apr. 2009.

[17] G. S. Aoude, V. R. Desaraju, L. H. Stephens, and J. P. How, "Driver behavior classification at intersections and validation on large naturalistic data set," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 724–736, Jun. 2012.

[18] M. Sundbom, P. Falcone, and J. Sjöberg, "Online driver behavior classification using probabilistic ARX models," in *Proc. 16th Int. IEEE Conf. Intell. Transp. Syst.*, 2013, pp. 1107–1112.

[19] M. Van Ly, S. Martin, and M. M. Trivedi, "Driver classification and driving style recognition using inertial sensors," in *Proc. IEEE Intell. Veh. Symp.*, 2013, pp. 1040–1045.

[20] Z. Constantinescu, C. Marinoiu, and M. Vladoiu, "Driving style analysis using data mining techniques," *Int. J. Comput. Commun. Control*, vol. 5, no. 5, pp. 654–663, 2010.

[21] B. Higgs and M. Abbas, "Segmentation and clustering of car-following behavior: Recognition of driving patterns," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 1, pp. 81–90, Feb. 2015.

[22] B. Shi *et al.*, "Evaluating driving styles by normalizing driving behavior based on personalized driver modeling," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 45, no. 12, pp. 1502–1508, Dec. 2015.

[23] A. Bender, G. Agamennoni, J. R. Ward, S. Worrall, and E. M. Nebot, "An unsupervised approach for inferring driver behavior from naturalistic driving data," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 6, pp. 3325–3336, Dec. 2015.

[24] C. You, J. Lu, D. Filev, and P. Tsiotras, "Advanced planning for autonomous vehicles using reinforcement learning and deep inverse reinforcement learning," *Robot. Auton. Syst.*, vol. 114, pp. 1–18, 2019.

[25] J. Ho and S. Ermon, "Generative adversarial imitation learning," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 4565–4573.

[26] A. Kuefler, J. Morton, T. Wheeler, and M. Kochenderfer, "Imitating driver behavior with generative adversarial networks," in *Proc. IEEE Intell. Veh. Symp.*, 2017, pp. 204–211.

[27] X. Chen, Y. Zhai, C. Lu, J. Gong, and G. Wang, "A learning model for personalized adaptive cruise control," in *Proc. IEEE Intell. Veh. Symp.*, 2017, pp. 379–384.

[28] S. Elmalaki, H.-R. Tsai, and M Srivastava, "Sentio: Driver-in-the-loop forward collision warning using multisample reinforcement learning," in *Proc. 16th ACM Conf. Embedded Netw. Sensor Syst.*, 2018, pp. 28–40.

[29] W. Wang, D. Zhao, W. Han, and J. Xi, "A learning-based approach for lane departure warning systems with a personalized driver model," *IEEE Trans. Veh. Technol.*, vol. 67, no. 10, pp. 9145–9157, Oct. 2018.

[30] D. Nava, G. Panzani, P. Zampieri, and S. M. Savaresi, "A personalized adaptive cruise control driving style characterization based on a learning approach," in *Proc. IEEE Intell. Transp. Syst. Conf.*, 2019, pp. 2901–2906.

[31] B. Zhu, S. Yan, J. Zhao, and W. Deng, "Personalized lane-change assistance system with driver behavior identification," *IEEE Trans. Veh. Technol.*, vol. 67, no. 11, pp. 10293–10306, Nov. 2018.

[32] J. O. Ramsay and B. W. Silverman, *Applied Functional Data Analysis: Methods and Case Studies*. Berlin, Germany: Springer, 2007.

[33] J. O. Ramsay and B. W. Silverman, *Functional Data Analysis*. New York, NY, USA: Springer, 2005.

[34] C. Happ and S. Greven, "Multivariate functional principal component analysis for data observed on different (dimensional) domains," *J. Amer. Stat. Assoc.*, vol. 113, no. 522, pp. 649–659, 2018.

[35] F. Fortuna and F. Maturo, "K-means clustering of item characteristic curves and item information curves via functional principal component analysis," *Qual. Quantity*, vol. 53, no. 5, pp. 2291–2304, 2019.

[36] B. D. Ziebart, A. L. Maas, J. A. Bagnell, and A. K. Dey, "Maximum entropy inverse reinforcement learning.," in *Proc. AAAI Conf. Artif. Intell.*, 2008, pp. 1433–1438.

[37] M. Wulfmeier, P. Ondruska, and I. Posner, "Maximum entropy deep inverse reinforcement learning," in *Proc. NIPS Workshop Deep Reinforcement Learn.*, 2015, p. 10.

[38] D. Bezzina and J. Sayer, "Safety pilot model deployment: Test Conductor Team Report," Rep. no. DOT HS 812 171, 2015.

[39] A. Alizadeh, M. Moghadam, Y. Bicer, N. K. Ure, U. Yavas, and C. Kurtulus, "Automated lane change decision making using deep reinforcement learning in dynamic and uncertain highway environment," in *Proc. IEEE Intell. Transp. Syst. Conf.*, 2019, pp. 1399–1404.

[40] C. Happ-Kurz, "Object-oriented software for functional data," *J. Stat. Softw.*, vol. 93, no. 5, pp. 1–38, 2020. [Online]. Available: https://www.jstatsoft.org/index.php/jss/article/view/v093i05

[41] T. T. D. Team *et al.*, "Theano: A python framework for fast computation of mathematical expressions," 2016, *arXiv:1605.02688*.

[42] C. R. Harris *et al.*, "Array programming with NumPy," *Nature*, vol. 585, no. 7825, pp. 357–362, Sep. 2020. [Online]. Available: https://doi.org/10.1038/s41586-020-2649-2

[43] A. Aljaafreh, N. Alshabatat, and M. S. N. Al-Din, "Driving style recognition using fuzzy logic," in *Proc. IEEE Int. Conf. Veh. Electron. Saf.*, 2012, pp. 460–463.

[44] W. Wang, J. Xi, A. Chong, and L. Li, "Driving style classification using a semisupervised support vector machine," *IEEE Trans. Human-Mach. Syst.*, vol. 47, no. 5, pp. 650–660, Oct. 2017.

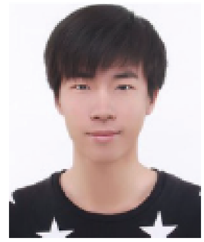
[45] S. Moridpour, G. Rose, and M. Sarvi, "Effect of surrounding traffic characteristics on lane changing behavior," *J. Transp. Eng.*, vol. 136, no. 11, pp. 973–985, 2010.

[46] C. Sherlaw-Johnson, S. Gallivan, and J. Burridge, "Estimating a Markov transition matrix from observational data," *J. Oper. Res. Soc.*, vol. 46, no. 3, pp. 405–410, 1995.

[47] H. An and J. Jung, "Decision-making system for lane change using deep reinforcement learning in connected and automated driving," *Electronics*, vol. 8, no. 5, 2019, Art. no. 543.

[48] C.-J. Hoel, K. Wolff, and L. Laine, "Automated speed and lane change decision making using deep reinforcement learning," in *Proc. 21st Int. Conf. Intell. Transp. Syst.*, 2018, pp. 2148–2155.

[49] C. Finn, P. Christiano, P. Abbeel, and S. Levine, "A connection between generative adversarial networks, inverse reinforcement learning, and energy-based models," 2016, *arXiv:1611.03852*.



Jundi Liu (Student Member, IEEE) received the B.S. degree in computer science and engineering from Shanghai Jiao Tong University, Shanghai, China, in 2016, the M.S degree in industrial and systems engineering in 2018 from the University of Washington, Seattle, WA, USA where he is currently working toward the Ph.D. degree in industrial and systems engineering. His research interests include human-in-the-loop reinforcement learning, human factors, and human-vehicle interactions.

Linda Ng Boyle (Member, IEEE) received the B.S. degree in industrial engineering from the State University of New York at Buffalo, Buffalo, NY, USA, in 1986, the M.S. degree in inter-engineering and the Ph.D. degree in civil and environmental engineering from the University of Washington (UW), Seattle, WA, USA, in 1994 and 1998, respectively. She was on the Faculty of The University of Iowa, Iowa City, IA, USA, and a Research Scientist with the U.S. Department of Transportation-Volpe Center, Cambridge, MA, USA. She is currently a Professor of Industrial and Systems Engineering at the University of Washington and has a joint appointment in Civil and Environmental Engineering. Her research interests include human factors, human behavior modeling, and human-systems interactions.

Ashis G. Banerjee (Senior Member, IEEE) received the B.Tech. degree in manufacturing science and engineering from the Indian Institute of Technology Kharagpur, Kharagpur, India, in 2004, the M.S. and Ph.D. degrees in mechanical engineering from the University of Maryland, College Park, MD, USA, in 2006 and 2009, respectively. He was a Research Scientist and Postdoctoral Associate with the Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA. He was a Research Scientist with GE Global Research, Niskayuna, NY, USA. He is currently an Associate Professor of Industrial and Systems Engineering and Mechanical Engineering at the University of Washington, Seattle, WA, USA. His research interests include autonomous robotics, predictive analytics, and statistical learning.