Proceedings of the 2022 6th CAA International Conference on Vehicular Control and Intelligence (CVCI), Nanjing, China, October 28-30, 2022

An interpretation framework for autonomous vehicles decision-making via SHAP and RF

1st Zhihao Cui
Clean Energy Automotive Engineering Center
School of Automotive Studies
Tongji University
Shanghai, China
2131523@tongji.edu.cn

2nd Meng Li
Department of Control Science and Engineering
Tongji University
Shanghai, China
2010460@tongji.edu.cn

2nd Yanjun Huang
Clean Energy Automotive Engineering Center
Shanghai Research Institute for Intelligent Autonomous Systems
School of Automotive Studies
Tongji University, Shanghai, China
yanjun_huang@tongji.edu.cn

2nd Yulei Wang
Shanghai Research Institute for Autonomous Systems
Department of Control Science and Engineering
Tongji University
Shanghai, China
wangyulei@tongji.edu.cn

2nd Hong Chen*
Clean Energy Automotive Engineering Center
Shanghai Research Institute for Autonomous Systems
Department of Control Science and Engineering
Tongji University, Shanghai, China
chenhong2019@tongji.edu.cn

2022 6th CAA International Conference on Vehicular Control and Intelligence (CVCI) | 978-1-6654-5374-5/22/$31.00 ©2022 IEEE | DOI: 10.1109/CVCI56766.2022.9964561

***Abstract*—Decision-making for autonomous vehicles is critical to achieving safe and efficient autonomous driving. In recent years, deep reinforcement learning (DRL) techniques have emerged as the most promising way to enable intelligent decision-making. However, DRL with 'black box' nature is not widely understood by humans, thus hindering their social acceptance. In this paper, we combine SHapley Additive exPlanation (SHAP) and random forest (RF) techniques to bring transparency to decision-making obtained by DRL. Specifically, we first implement decision-making of autonomous vehicles following in discrete action space based on DRL algorithm with the goal of safety and efficiency. Then we use the SHAP technique to simplify the feature space, which shows that relative distance, longitudinal speed of the ego vehicle, and longitudinal speed of the proceeding vehicle have a critical impact on vehicle following task. Finally, we collect the state-action pairs generated by the DRL model and perform feature filtering, and fit the decision model with an interpretable RF model. The simulation results show that the RF model achieves the behavioral explanation of autonomous vehicle following.**

***Index Terms*—autonomous vehicle, DRL, SHAP, RF, interpretation**

I. Introduction

In recent years, autonomous driving (AD) has attracted a lot of attention benefiting from the boom in artificial intelligence technology [1], [2]. In particular, the AD decision-making module receives the surrounding traffic information from the perception to make safe and efficient behavioral decisions that are then fed to the planning and control to enable AD [3]. Finite state machines [4], [5] are currently the most widely used behavioral decision models because of their simplicity and ease of implementation but ignore the dynamics and uncertainty of the environment. In addition, the division and management of states are tedious when there are many driving scene features, which are mostly applicable to simple scenarios and difficult to perform the behavioral decision tasks in complex road environments with rich structured features [6].

Those limitations might be avoided with the learning-based decision-making model [7], [8], in which, the driving actions are inferred from the trained network obtained through a 'query' mechanism. One branch of learning-based decision models is imitation learning (IL), which establishes a decision model by learning the expert driving data. However, it is difficult to collect sufficient labeled data from skilled drivers. Another branch, deep reinforcement learning (DRL), combines the representational capability of deep learning with the sequential decision capability of reinforcement learning (RL). One advantage of DRL methods is that they can learn optimized policies without requiring model information. Moreover, DRL can learn optimal strategies by interacting with the environment without requiring expert data. However, the deep networks with 'black-box' nature in DRL make it difficult to interpret the decisions. Especially when oriented to applications of autonomous driving systems with high safety requirements, it is significant to give explanations of decisions to enhance user trust [9], [10].

This work proposed an interpretation framework, combining the SHAP and RF techniques to enhance transparency to decision-making obtained by DRL. In our framework, the SHAP-based approach analyzes and simplifies the feature

This work was supported by the National Key Research and Development Program of China 2020AAA0108101.

978-1-6654-5374-5/22/$31.00 ©2022 IEEE

space to find the important features associated with the decision, which in turn processes the data generated by the DRL model. In addition, we fit the processed data with an interpretable random forest model to explain the decisions of the original DRL model. To the authors' best knowledge, this interpretable framework is proposed for the first time in the context of autonomous driving decision applications. The main contributions of this paper can be summarized as follows.

1) This work proposed an interpretation framework, combining the SHAP and RF techniques to enhance transparency of decision-making obtained by DRL.

2) This work validated the effectiveness of the proposed framework in a vehicle-following scenario, and the framework can be easily applied to other autonomous driving scenarios.

The rest of this paper is organized as follows. Section II provides an overview of the work related to the research in this paper. In Section III, we apply DRL method based on value function approximation to achieve efficient and safe vehicle-following decisions in discrete action spaces (including acceleration, deceleration, and hold). In Section IV, we propose a method combining SHAP and RF to analyze the feature space, and finally replace the DRL strategy with RF to achieve transparent decision-making. In Section V, we illustrate the superiority of the proposed method through simulation validation and data analysis.

II. Related Works

This work researches DRL for autonomous driving decision-making and its interpretability, so this section will review related work in terms of both DRL as well as interpretable methods.

A. DRL for autonomous driving

In 2018, Mirchevska et al. implemented autonomous lane change decision-making with desired speed as a reward using the DQN method and feature learning in a three-lane highway environment. And the safety distance is added as a constraint in the deployment phase of the DQN model to ensure autonomous driving safety. Simulation results show that the approach avoids collisions and has a higher average speed than the rule-based approach [11]. Unlike the previous approach, Subramanya et al. embedded safety checks directly into the algorithm and performs safety checks directly at each training. In addition, a dedicated space is set up to store undefined collision scenarios for the purpose of improving safety and learning efficiency. The average returns and learning curves obtained from the simulation in a highway simulation environment are significantly better than the other algorithms. Chen et al. [12] proposed an interpretable end-to-end deep reinforcement learning method for urban driving environments. By introducing a sequential latent model, a semantic bird mask is generated, which significantly reduces the sample complexity of reinforcement learning. Simulation results show that the method outperforms baselines such as DQN, DDPG, TD3, and SAC in vehicle-congested urban scenarios. Xu et al. [13] achieved better learning efficiency and realized overtaking decision-making under realistic highway traffic conditions using feature learning and value function approximation within the DRL framework. Sallab et al. [14] used the Deep Deterministic Actor-Critic (DDAC) algorithm to track lanes and maximize the average velocity. Alizadeh et al. [15] applied Deep Q-learning (DQL) to handle the lane-changing decision-making problem in an uncertain highway environment. Liu et al. [16] adopt the Proximal Policy Optimization (PPO) method to solve the problems of low sampling efficiency. The simulation results show that the collision rate is effectively reduced, the training speed is greatly accelerated.

B. Interpretation methods of DRL

DRL-based autonomous decisions are a network with 'black-box' nature that makes it difficult for the user to access the high-level knowledge behind the driving behavior.

SHAP [17] and LIME [18] are two popular techniques for interpreting DRL. The LIME technique works by constructing an approximate linear model in the vicinity of the instances, which in turn represents the degree of contribution of the input to the output in terms of the weights of the linear model. However, the different selection of kernel functions in LIME methods leads to differences in interpretable representations. There is no research related to the application of LIME-based interpretable methods for automated driving. SHAP applies the shapley value to represent the marginal contributions of all features, and the goal is to quantify the predictions of explainable instances by calculating the contribution of each feature. Liessner et al. [19] demonstrate the application of the SHAP method to autonomous driving based on DRL. In this study, the intelligent vehicle learns longitudinal motion behaviors such as acceleration and deceleration by the DDPG algorithm on a single lane considering the speed limit, and combines them with the SHAP framework for interpretation. In addition, data refitting to interpret the DRL model is also a common approach. Lukas et al. [20] trained a decision-making model for autonomous lane changing using DRL and subsequently used a decision tree model to fit the data generated by the DRL model. The logical relationships between the nodes of the decision tree model can be interpreted and understood more easily.

Inspired by the above research, this paper proposes an architecture to explain the DRL autonomous driving decision model, in which we introduce SHAP to analyze the input features and simplify the feature space, and then use the RF algorithm to fit the decision process of the DRL model to achieve the transparency of the decision process.

III. DRL-based automatic vehicle-following

This section uses DRL to solve the problem of automatic vehicle following.

A. Markov process modeling of automatic vehicle following

Fig. 1 shows the architecture of the DRL-based vehicle following strategy, where the vehicle following problem can

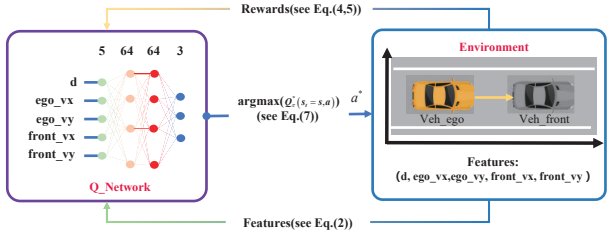


Fig. 1. Deep reinforcement learning-based automatic vehicle following

be abstracted as Markov Decision Process (MDP) [21], [22]:

$$\langle \ S, \quad A, \quad r, \quad P \ \rangle \tag{1}$$

where, S is the state space, containing any state s:

$$S = \ [d, \quad ego\_vx, \quad ego\_vy, \quad front\_vx, \quad front\_vy] \tag{2}$$

where d denotes the relative distance between the ego vehicle and the proceeding vehicle, ego_vx denotes the longitudinal velocity of the ego vehicle, ego_vy denotes the lateral speed of the ego vehicle, $front\_vx$ denotes the longitudinal velocity of the front vehicle, $front\_vy$ denotes the lateral velocity of the front vehicle. The action space A is discrete and defined as:

$$A = \ [Slower, \quad Idle, \quad Faster] \tag{3}$$

where action '$Slower$' denotes deceleration, action '$Idle$' denotes hold, and '$Faster$' denotes acceleration. $r = (s, a, s^{'})$ denotes the reward obtained by agent in the current state s executing action a to the next state $s^{'}$. Three types of rewards are included in this paper, reflecting the need for safe and efficient autonomous driving. Specifically, driving as fast as possible r_v, maintaining a safe distance r_d, and avoiding collisions r_c:

$$r = w_c * r_c + w_d * r_d + w_v * r_v \tag{4}$$

where r_c, r_d, r_v are:

$$r_c = \begin{cases} -1, & if \quad (collision == 1) \\ 0, & otherwise \end{cases} \tag{5a}$$

$$r_d = \begin{cases} -1/(d+1), & if \quad (5 < d < 10) \\ 0, & otherwise \end{cases} \tag{5b}$$

$$r_v = \begin{cases} \frac{V - V_{min}}{V_{max} - V_{min}}, & if \quad (V_{min} < V < V_{max}) \\ 0, & otherwise \end{cases} \tag{5c}$$

where $collision$ denotes whether a collision occurs, $V_{max} = 30m/s$, $V_{min} = 15m/s$, V denotes the maximum velocity, minimum velocity and the current velocity of the ego vehicle, respectively. In order to trade-off between the safety and efficiency of the vehicle following decision-making, the weights w_c, w_d, w_v are obtained through hierarchical analysis [23]: $[0.42, 0.42, 0.16]$.

B. Action-value function approximation

Based on the Markov model of the vehicle following developed in Section III-A, the action-value function is defined as:

$$Q_\pi(s,a) = E\left[\sum_{l=0}^{\infty} \gamma^l r\left(s_{t+l}, a_{t+l}\right) \mid s_t = s, a_t = a\right] \tag{6}$$

where the $\gamma \in (0,1)$ is the reward decay factor. Note that P in the Markov model is difficult to obtain, resulting in the inability to obtain the $Q_\pi(s,a)$ of the model, so the deep Q-network (DQN, Network structure is $5 * 64 * 64 * 3$) is used to approximate the $Q_\pi(s,a)$, and a detailed description of the algorithm is given in [24].

The output of the DQN is the converged $Q_\pi(s,a)$ i.e. $Q^*_\pi(s,a)$, and then the current optimal action a^* is obtained by traversing the $Q^*_\pi(s,a)$ corresponding to the action space A according to the current state s of the agent.

$$a^* = \underset{a}{argmax}\, Q^*_\pi(s_t = s, a) \tag{7}$$

IV. Interpretation Method Using SHAP and RF

Due to the 'black-box' nature of DQN, this section combines SHAP and RF methods to achieve simplification of feature/state space S and transparency of decision-making.

A. Feature importance and feature dependency analysis using SHAP

The complex feature space is not conducive to achieving transparency in decision-making. This section analyzes the feature space based on the SHAP (SHapley Additive exPlanations) framework. The reason for using SHAP is that it is based on shapley theory, which has solid mathematical theoretical foundation. Shapley theory satisfies four axioms: efficiency, symmetry, dummy, and additivity [25]. SHAP calculates shapley values to characterize the impact of features on predictions. The feature values of the instances act as the players/features in the coalition and show how to distribute the predictions equally among the features. The interpretation of SHAP is expressed as an additive feature attribution method, i.e., a linear model. SHAP specifies the interpretation as:

$$g(z^{'}) = \phi_0 + \sum_{j=1}^{M} \phi_j z_j^{'} \tag{8}$$

where g is the interpreted model, $z^{'} \in \{0,1\}^M$ is the coalition vector, M is the maximum coalition size, and ϕ_j is the shapley value of feature s_j. Firstly, obtain multiple sets of random coalitions by sampling, these coalitions consist of combinations of binary numbers {0,1}. Then mapping the sampled coalitions to the original feature space S by $h_s(z_k^{'})$, where $h_s(z_k^{'}) : \{0,1\}^M \longmapsto \mathbb{R}^p$ (The upper corner p indicates the length of the state space vector S (see Eq.2)). The predicted value of each coalition is obtained applying model.

$$f(h_s(z_k^{'})): \quad f(\cdot) \triangleq Q^*_\pi(\cdot) \tag{9}$$

After that, the weight of each feature coalition is calculated according to the following equation:

$$\pi_s(z_k^{'}) = \frac{M-1}{(\binom{M}{|z_k^{'}|})|z_k^{'}|(M-|z_k^{'}|))} \tag{10}$$

where $|z^{'}|$ denotes the number of exist features in coalition. Finally, based on the Eq. 8, the Eq. 9 and the Eq. 10, a loss function L is constructed to optimize g:

$$L(f,g,\pi_s(z_k^{'})) = \sum_{z_k^{'} \in Z} [f(h_s(z_k^{'})) - g(z_k^{'})]^2 \pi_x(z_k^{'}) \tag{11}$$

where Z denotes all sampled coalitions. The estimated coefficients ϕ_j are can be obtained by optimally solving for Eq.11. So we filter the S_{set} in sample set $D = \{S_{set}, A_{set}\}$ generated by DRL model and then construct a new sample set $D^* = \{S^*_{set}, A^*_{set}\}$ based on the ϕ_j i.e. feature importance and feature dependency.

B. Post hoc interpretation based on random forest

This section trains the RF model to achieve transparency in decision-making based on sample set D^*. In 2001, Breiman et al. [26] firstly proposed RF, a very efficient prediction method consisting of multiple decision trees (DT). Each DT in a forest is constructed using data based on the bootstrap sample method [27]. The tree base learner is typically grown in the data training process using classification and regression tree (CART) methodology [28] illustrated in Fig. 2.

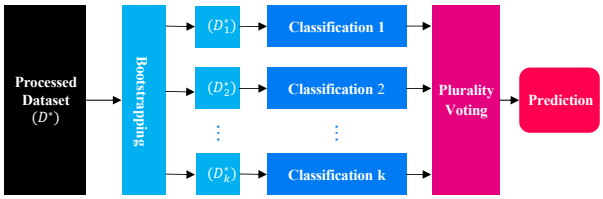


Fig. 2. Flow chart of RF prediction

V. Experiments

This section constructed a vehicle following scenario based on a highway environment [29] as shown in Fig. 1 and use the classical DQN algorithm introduced in Section III to implement automatic vehicle following, with hyperparameters defined in the Table I. From the Fig. 3, we can see that the average reward of each episode is stable at about 25,000 steps, indicating that the $Q_\pi(s,a)$ has nearly converged.

We took a total of 1000 steps of state-action data $D = \{S_{set}, A_{set}\}$. Then we calculate the shapley value of all features based on data and $Q^*_\pi(s,a)$. As shown in the Fig. 4, (a), (c), and (e) represent the feature importance vs. average absolute shapley value plots for $Q^*_\pi(s,a)$ of actions '$Slower$', '$Idle$', '$Faster$', respectively. The larger the average absolute of shapley value, the greater the influence of the feature on the decision-making in the tested sample. (b), (d), and (f)

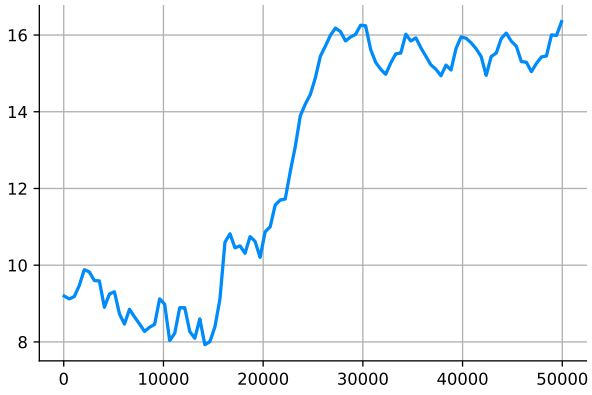


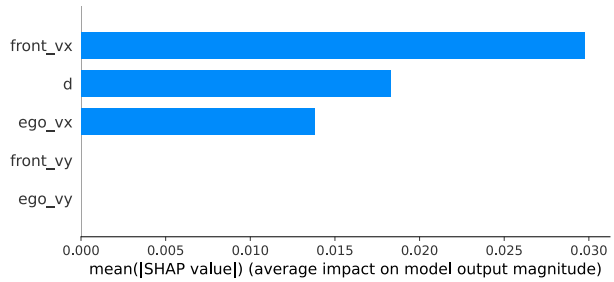
Fig. 3. Average reward per episode (During training)

TABLE I
HYPERPARAMETERS

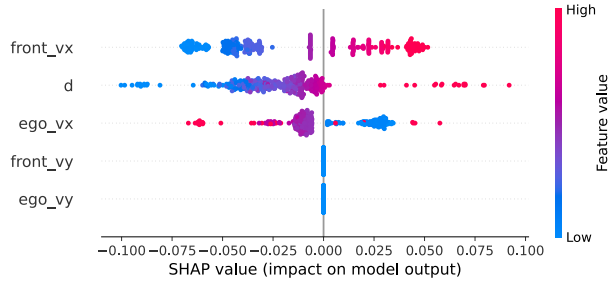
Hyperparameters	Description	Value [Unit]
γ	Reward decay factor	0.5
Q_{net}	Action-value approximation network	5× 64× 64× 3
Lr	Learning rate	0.0002
Tui	Target update interval	50
Af	Activation function	ReLU
Tt	Total timesteps	50,000

represent the distribution of shapley values of features for $Q^*_\pi(s,a)$ of '$Slower$', '$Idle$', '$Faster$', respectively. The shapley values of d, ego_vx, $front\_vx$ are presented on both sides of 0, which contributes more to the $Q^*_\pi(s,a)$. Meanwhile, the shapley values of the other two types of features: ego_vy, $front\_vy$ are almost constant around 0, indicating that they have almost no influence on the vehicle-following decision-making. The above results show that the d, ego_vx, $front\_vx$ have the significant effect on the decision-making, which objectively reflects the attention mechanism of human driving.

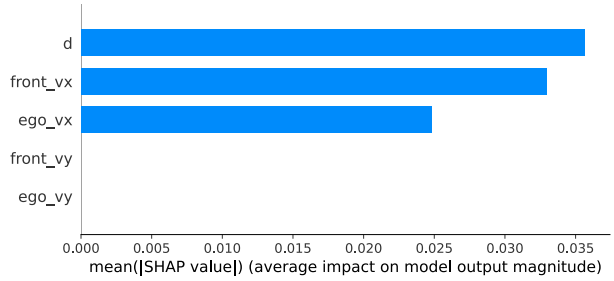
Fig. 5 is the SHAP dependency plot of features,take Q value of 'Faster' as an example, which shows the dependence between features. From left to right are the feature dependency plots between the three features [ego_vx, d, $'front\_vx'$] that contribute most to the Q value of the acceleration and the other features, as can be seen in (e) and (f) of Fig. 4. Fig. 4 shows the degree of influence of each feature individually on the output, while the dependency plot of SHAP allows us to see more details, such as the interactions with other features. The first row shows the dependency plot between each feature and the feature with the strongest interaction, with the interaction decreasing from top to bottom, while the last row shows the dependency plot between the three most important features



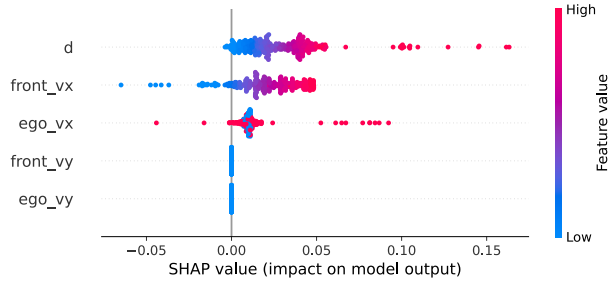
(a) Feature Importance for Q value of 'Slower'



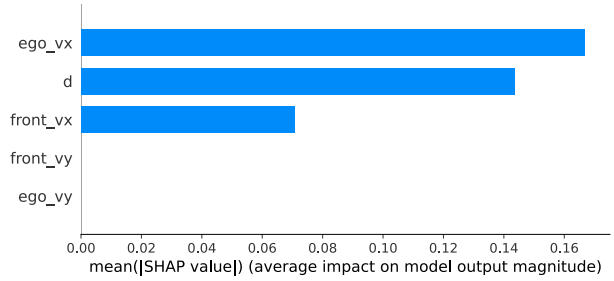
(b) Distribution of shapley values of features for Q value of 'Slower'



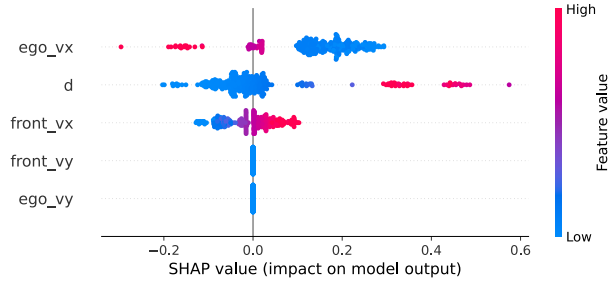
(c) Feature Importance for Q value of 'Idle'



(d) Distribution of shapley values of features for Q value of 'Idle'



(e) Feature Importance for Q value of 'Faster'



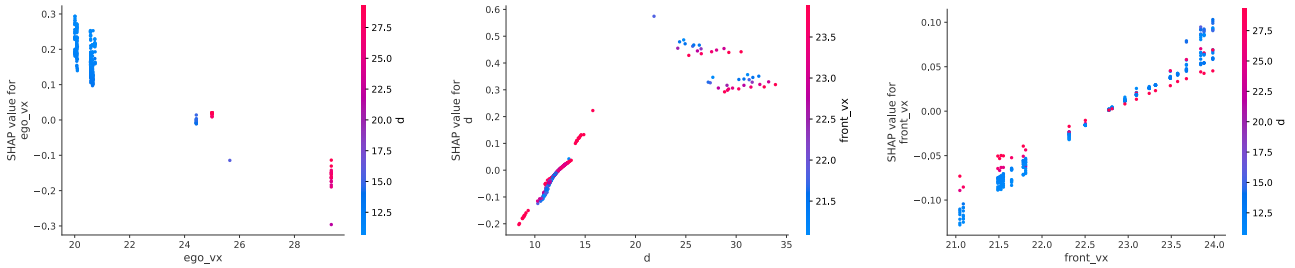
(f) Distribution of shapley value values of features for Q value of 'Faster'

Fig. 4. Feature importance ranking and distribution of shapley values of features diagram: expresses the degree of influence of different features on the corresponding $Q^*_\pi(s,a)$ output in the action space.

$[ego\_vx,\ d,\ 'front\_vx']$ and $'ego\_vy'$. Obviously, there is no interaction between them in the last row, and similarly, there is no interaction between $[ego\_vx,\ d,\ 'front\_vx']$ and $'front\_vy'$ (not shown in the figure). So we can know that $'ego\_vy'$ and $'front\_vy'$ have little effect on autonomous driving following decision-making. We can also see many details from Fig. 5, for example, in (b), the contribution of feature d to the output increases as the distance increases; for local, from the upper right corner, we can see that when the distance $'d'$ is larger, the utility of feature d has a positive contribution whether the front car is fast or slow; from the lower left corner, we can see that for a certain speed of the front car, a smaller distance has a negative contribution value and a larger distance has a positive contribution value.

Based on the above feature importance and feature dependence analysis, we filter the S_{set} in sample set $D = \{S_{set}, A_{set}\}$, i.e., deleted the features ego_vy, $front\_vy$. Then, we construct a new sample set $D^* = \{S^*_{set}, A^*_{set}\}$, and S^*_{set} contains only three types of features: $[d,\ ego\_vx,\ front\_vx]$.

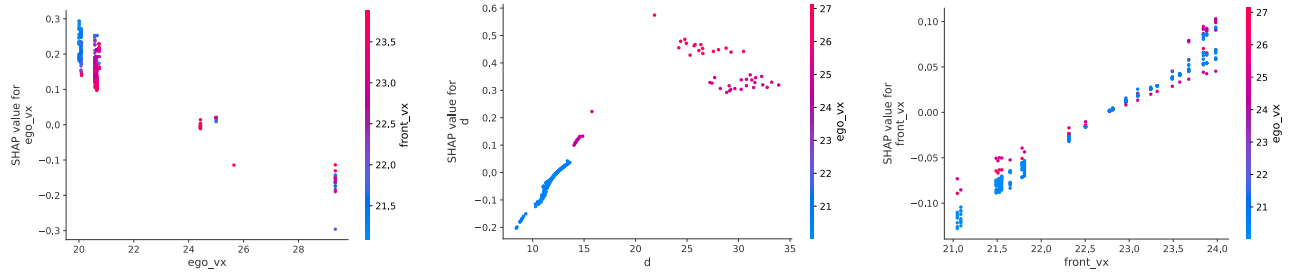
We generated a random forest model introduced in Section IV-B with the dataset D^*. Fig. 6 shows the first decision tree of the RF model with blue bars in each tree indicating 'Slower', purple bars indicating 'Idle', and red bars indicating 'Faster'. Starting from the root node, the decision tree divides each branch according to the values of the features until it splits to each leaf node, forming multiple complete decision paths. In addition, we analyze the decision process for the instance with state: $d = 11.56, ego\_vx = 20.00, front\_vx = 21.17$. In decision tree 1, the instances correspond to the paths indicated by the orange arrows, which constitute a logical rule. From top to bottom, the longitudinal speed $front\_vx = 21.17$ of the front vehicle in the instance is less than the judgment node 22.67 therefore enters the second layer through the left path, where the relative distance $d = 11.56$ in the instance is greater than the node threshold 10.73 therefore the right path is taken to reach the judgment node in the third layer, where the deceleration action is output because the longitudinal speed $ego\_vx = 20.00$ of the ego vehicle in the instance is less than the node threshold 22.54.



(a) SHAP-dependent plot of $'ego\_vx'$ and its SHAP value with respect to $'d'$

(b) SHAP-dependent plot of $'d'$ and its SHAP value with respect to $'front\_vx'$

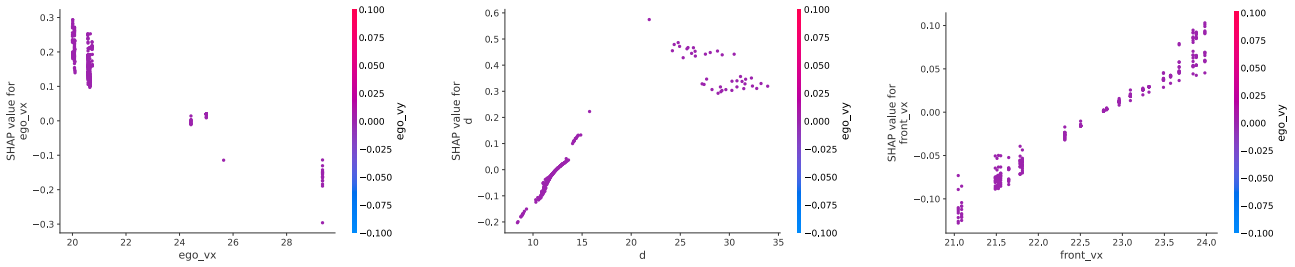
(c) SHAP-dependent plot of $'front\_vx'$ and its SHAP value with respect to $'d'$



(d) SHAP-dependent plot of $'ego\_vx'$ and its SHAP value with respect to $'front\_vx'$

(e) SHAP-dependent plot of $'d'$ and its SHAP value with respect to $'ego\_vx'$

(f) SHAP-dependent plot of $'front\_vx'$ and its SHAP value with respect to $'ego\_vx'$



(g) SHAP-dependent plot of $'ego\_vx'$ and its SHAP value with respect to $'ego\_vy'$

(h) SHAP-dependent plot of $'d'$ and its SHAP value with respect to $'front\_vy'$

(i) SHAP-dependent plot of $'front\_vx'$ and its SHAP value with respect to $'ego\_vy'$

Fig. 5. SHAP dependence plot for Q value of 'Faster'

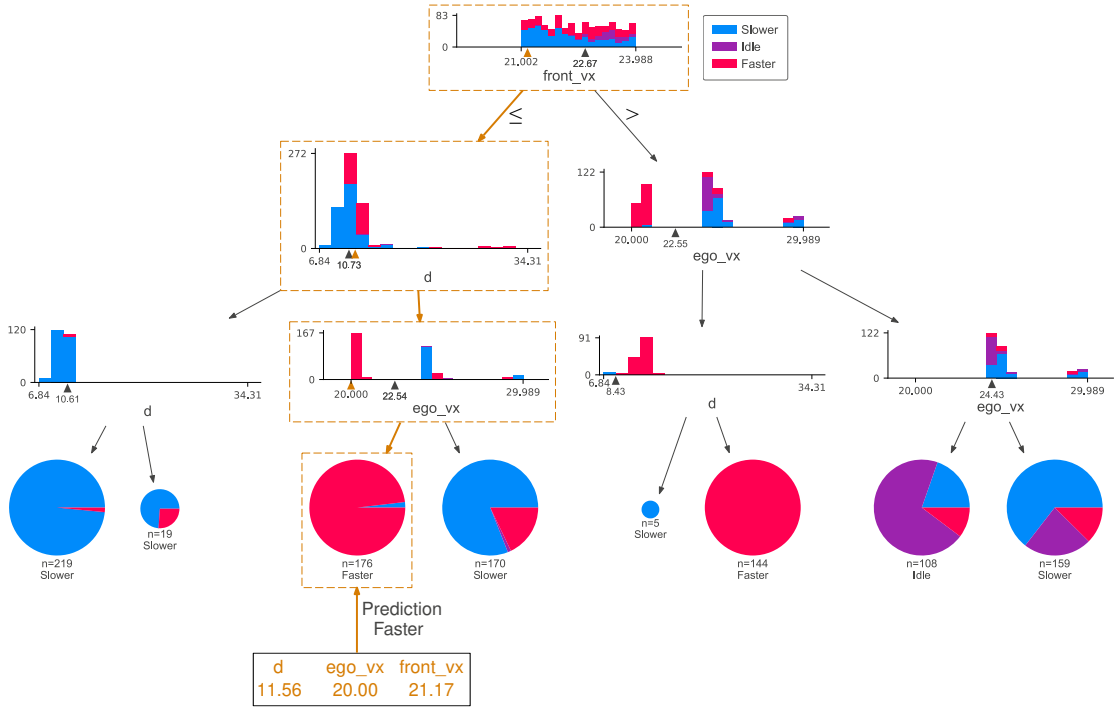


Fig. 6. Decision tree 1

VI. Conclusion

In this paper, we proposed an interpretation framework, combining the SHAP and RF techniques to enhance transparency to decision-making obtained by DRL. In our framework, the SHAP-based approach analyzes and simplifies the feature space to find the important features associated with the decision, which in turn processes the data generated by the DRL model. At the same time, the feature analysis allows us to obtain objective explanations, some of which are consistent with partial human perceptions. In addition, we fit the processed data with an interpretable RF model to explain the decisions of the original DRL model. Simulation results show that our proposed framework enhances the interpretability of DRL-based autonomous following decisions. However, there is still much work to be further investigated. Future work is to select a shapley base value for autonomous driving rather than a mean value so that the interpretation of the results is more in alignment with the way humans perceive; it is necessary to classify the data more finely and then perform interpretable analysis; in addition, the analysis of natural driving behavior based on machine learning is also an interesting research direction.

Acknowledgment

This work was supported by the National Key Research and Development Program of China 2020AAA0108101.

References

[1] Y. Ma, Z. Wang, H. Yang, and L. Yang, "Artificial intelligence applications in the development of autonomous vehicles: a survey," *IEEE/CAA Journal of Automatica Sinica*, vol. 7, no. 2, pp. 315–329, 2020.

[2] J. Li, H. Cheng, H. Guo, and S. Qiu, "Survey on artificial intelligence for vehicles," *Automotive Innovation*, vol. 1, no. 1, pp. 2–14, 2018.

[3] W. Schwarting, J. Alonso-Mora, and D. Rus, "Planning and decision-making for autonomous vehicles," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 1, pp. 187–210, 2018.

[4] X. Wang, X. Qi, P. Wang, and J. Yang, "Decision making framework for autonomous vehicles driving behavior in complex scenarios via hierarchical state machine," *Autonomous Intelligent Systems*, vol. 1, no. 1, pp. 1–12, 2021.

[5] P. Wang, S. Gao, L. Li, S. Cheng, and H. Zhao, "Research on driving behavior decision making system of autonomous driving vehicle based on benefit evaluation model," *Archives of transport*, vol. 53, 2020.

[6] J. Perez, V. Milanes, E. Onieva, J. Godoy, and J. Alonso, "Longitudinal fuzzy control for autonomous overtaking," in *2011 IEEE International Conference on Mechatronics*. IEEE, 2011, pp. 188–193.

[7] A. D. Lattner, J. D. Gehrke, I. J. Timm, and O. Herzog, "A knowledge-based approach to behavior decision in intelligent vehicles," in *IEEE Proceedings. Intelligent Vehicles Symposium, 2005.* IEEE, 2005, pp. 466–471.

[8] G. Dimitrakopoulos, G. Bravos, M. Nikolaidou, and D. Anagnostopoulos, "Proactive, knowledge-based intelligent transportation system based on vehicular sensor networks," *IET Intelligent Transport Systems*, vol. 7, no. 4, pp. 454–463, 2013.

[9] D. Omeiza, H. Web, M. Jirotka, and L. Kunze, "Towards accountability: providing intelligible explanations in autonomous driving," in *2021 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2021, pp. 231–237.

[10] D. Omeiza, H. Webb, M. Jirotka, and L. Kunze, "Explanations in autonomous driving: A survey," *arXiv preprint arXiv:2103.05154*.

[11] B. Mirchevska, C. Pek, M. Werling, M. Althoff, and J. Boedecker, "High-level decision making for safe and reasonable autonomous lane changing using reinforcement learning," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2018, pp. 2156–2162.

[12] J. Chen, S. E. Li, and M. Tomizuka, "Interpretable end-to-end urban autonomous driving with latent deep reinforcement learning," *IEEE Transactions on Intelligent Transportation Systems*, 2021.

[13] X. Xu, L. Zuo, X. Li, L. Qian, J. Ren, and Z. Sun, "A reinforcement learning approach to autonomous decision making of intelligent vehicles on highways," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, no. 10, pp. 3884–3897, 2018.

[14] A. E. Sallab, M. Abdou, E. Perot, and S. Yogamani, "Deep reinforcement learning framework for autonomous driving," *Electronic Imaging*, vol. 2017, no. 19, pp. 70–76, 2017.

[15] 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 *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*. IEEE, 2019, pp. 1399–1404.

[16] T. Liu, H. Wang, B. Lu, J. Li, and D. Cao, "Decision-making for autonomous vehicles on highway: Deep reinforcement learning with continuous action horizon," *arXiv preprint arXiv:2008.11852*.

[17] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proceedings of the 31st international conference on neural information processing systems*, 2017, pp. 4768–4777.

[18] M. T. Ribeiro, S. Singh, and C. Guestrin, "" why should i trust you?" explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 2016, pp. 1135–1144.

[19] R. Liessner, J. Dohmen, and M. A. Wiering, "Explainable reinforcement learning for longitudinal control." in *ICAART (2)*, 2021, pp. 874–881.

[20] L. M. Schmidt, G. Kontes, A. Plinge, and C. Mutschler, "Can you trust your autonomous car? interpretable and verifiably safe reinforcement learning," in *2021 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2021, pp. 171–178.

[21] M. L. Puterman, *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.

[22] C. H. Papadimitriou and J. N. Tsitsiklis, "The complexity of markov decision processes," *Mathematics of operations research*, vol. 12, no. 3, pp. 441–450, 1987.

[23] S. W. Raudenbush and A. S. Bryk, *Hierarchical linear models: Applications and data analysis methods*. sage, 2002, vol. 1.

[24] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing atari with deep reinforcement learning," *arXiv preprint arXiv:1312.5602*.

[25] E. Winter, "The shapley value," *Handbook of game theory with economic applications*, vol. 3, pp. 2025–2054, 2002.

[26] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[27] C.-C. Yeh, D.-J. Chi, and Y.-R. Lin, "Going-concern prediction using hybrid random forests and rough set approach," *Information Sciences*, vol. 254, pp. 98–110, 2014.

[28] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and regression trees*. Routledge, 2017.

[29] E. Leurent, "An environment for autonomous driving decision-making," *GitHub repository*, 2018.