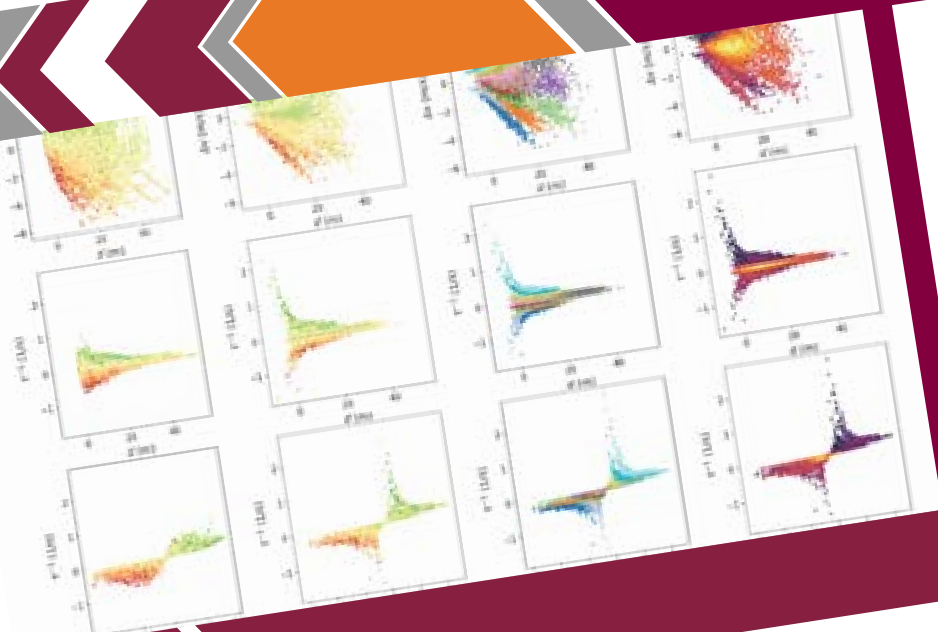
Enhancing Automated Vehicle Safety Through Testing with Realistic Driver Models

December 2023 | Final Report



SAFE-D
SAFETY THROUGH DISRUPTION

VIRGINIA TECH TRANSPORTATION INSTITUTE
VIRGINIA TECH

Texas A&M Transportation Institute

San Diego State University
Leadership Starts Here

Disclaimer

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. 06-009	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Enhancing Automated Vehicle Safety Through Testing with Realistic Driver Models		5. Report Date December 2023 6. Performing Organization Code:
7. Author(s) Alfredo Garcia Anthony D. McDonald		8. Performing Organization Report No.
9. Performing Organization Name and Address: Safe-D National UTC Texas A&M Transportation Institute 3135 TAMU College Station, Texas 77843-3135		10. Work Unit No. 11. Contract or Grant No. 69A3551747115/Project 06-009
12. Sponsoring Agency Name and Address Office of the Secretary of Transportation (OST) U.S. Department of Transportation (US DOT)		13. Type of Report and Period Final Research Report Start-01-2022 End-12-2023 14. Sponsoring Agency Code

15. Supplementary Notes

This project was funded by the Safety through Disruption (Safe-D) National University Transportation Center, a grant from the U.S. Department of Transportation – Office of the Assistant Secretary for Research and Technology, University Transportation Centers Program.

16. Abstract

Driver process models play a central role in the testing, verification, and development of automated and autonomous vehicle technologies. Prior models developed from control theory and physics-based rules are limited in automated vehicle applications due to their restricted behavioral repertoire. Data-driven machine learning models are more capable than rule-based models but are limited by the need for large training datasets and their lack of interpretability. In this project we developed a novel car following modeling approach using active inference, which has comparable behavioral flexibility to data-driven models while maintaining interpretability. We assessed the proposed model, the Active Inference Driving Agent (AIDA), through a benchmark analysis against several benchmarks. The models were trained and tested on a real-world driving dataset using a consistent process. The testing results showed that the AIDA predicted driving controls significantly better than the rule-based Intelligent Driver Model and had similar accuracy to the data-driven neural network models in three out of four evaluations. Subsequent interpretability analyses illustrated that the AIDA's learned distributions were consistent with driver behavior theory and that visualizations of the distributions could be used to directly comprehend the model's decision-making process and correct model errors attributable to limited training data.

17. Key Words Perception and action model, partially observable Markov Decision Process, active inference, bi-level optimization, car-following task		18. Distribution Statement No restrictions. This document is available to the public through the Safe-D National UTC website, as well as the following repositories: VTechWorks, The National Transportation Library, The Transportation Library, Volpe National Transportation Systems Center, Federal Highway Administration Research Library, and the National Technical Reports Library.	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 11	22. Price $0

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

Abstract

Driver process models play a central role in the testing, verification, and development of automated and autonomous vehicle technologies. Prior models developed from control theory and physics-based rules are limited in automated vehicle applications due to their restricted behavioral repertoire. Data-driven machine learning models are more capable than rule-based models but are limited by the need for large training datasets and their lack of interpretability. In this project we developed a novel car following modeling approach using active inference, which has comparable behavioral flexibility to data-driven models while maintaining interpretability. We assessed the proposed model, the Active Inference Driving Agent (AIDA), through a benchmark analysis against several benchmarks. The models were trained and tested on a real-world driving dataset using a consistent process. The testing results showed that the AIDA predicted driving controls significantly better than the rule-based Intelligent Driver Model and had similar accuracy to the data-driven neural network models in three out of four evaluations. Subsequent interpretability analyses illustrated that the AIDA's learned distributions were consistent with driver behavior theory and that visualizations of the distributions could be used to directly comprehend the model's decision-making process and correct model errors attributable to limited training data.

Acknowledgements

This project was funded by the Safety through Disruption (Safe-D) National University Transportation Center, a grant from the U.S. Department of Transportation – Office of the Assistant Secretary for Research and Technology, University Transportation Centers Program.

Table of Contents

TABLE OF CONTENTS III

LIST OF FIGURES IV

INTRODUCTION 1

METHOD 2

RESULTS 4

DISCUSSION 7

CONCLUSIONS AND RECOMMENDATIONS 8

ADDITIONAL PRODUCTS 8

Education and Workforce Development Products 9

Technology Transfer Products 9

Data Products 9

REFERENCES 10

List of Figures

Figure 1. Active Inference Driving Agent (AIDA): o = instantaneous observation, a = control action, b = instantaneous belief, *G* = expected free energy 3

Figure 2. Visualizations of the dataset and AIDA model components. (a), plot of random observations sampled from the dataset; (b), (c), and (d) plot of samples from model............5

Figure 3. Offline evaluation MAE-IQM. Each point corresponds to a random seed used to initialize model training and its color corresponds to the testing condition of either same-lane or new-lane. 6

Figure 4. Online evaluation ADE-IQM. Each point corresponds to a random seed used to initialize model training and its color corresponds to the testing condition of either same-lane or new-lane. 7

Introduction

The rapid development of automated and connected vehicle technologies has created a corresponding demand for models of driver behavior that can be used to calibrate design parameters [1, 2], evaluate technologies [3, 4], and refine real-time decision making [5]. To be effective in these tasks, driver models must be flexible, generalizable, and interpretable. Model flexibility is the ability of the model to mimic nuanced social behavior of human drivers [6]. Generalizability is the ability of the model to extend to new environments with minimal modeler intervention. Interpretability refers to both a clear connection between model mechanics and predicted behavior and a grounding in human psychology [7]. These elements facilitate model inspection and diagnostics which are essential for interpretable models [8].

Car following is an important driving sub-task as it represents a large portion of current driving time and crashes involving automated vehicles [9, 10]. Therefore, it is important to develop flexible, generalizable, and interpretable car following models for automated vehicles and future transportation systems.

Existing car following models can be partitioned into rule-based models and data-driven models. Rule-based models generate acceleration behavior based on a function specified by the modeler. Typically, this function is grounded in known observations or driver behavior theory [11]. For example, the Intelligent Driver Model (IDM) predicts driver acceleration based on deviations from a desired speed and distance headway. While rule-based models have a clear connection between model mechanics and predicted behavior, they are limited in their flexibility and generalizability. Because the rules in rule-based models are designed to replicate driving behavior in specific contexts and depict driver characteristics with small parameter sets, they are limited in the behavioral repertoire and in generalizing to scenarios outside of those governed by rules beyond their initial rule set. For example, research has shown that rule-based models designed for car following do not generalize to emergency scenarios and crashes [12]. Despite these limitations, rule-based models are still widely used for automated vehicle analyses and thus offer a valid benchmark for new models.

In contrast to rule-based models, data-driven models learn a function that maps observations or features to acceleration behaviors using an algorithm. Recent works have used neural networks [13], reinforcement learning [14], and adversarial imitation learning [15] to model car following behavior. These approaches have shown considerable flexibility in replicating human behavior, however, data-driven models still struggle to reproduce well-known traffic phenomena such as stop-and-go oscillation and their generalizability is constrained by the chosen machine learning technique. Furthermore, the complexity of existing data-driven models prohibits interpretability both in the connection between input and output and in their grounding in human psychology. Despite these shortcomings, data-driven models are more generalizable to complex scenarios which are difficult for manual model specification. One important class of data-driven models is Behavior Cloning (BC) known for their simplicity and general effectiveness. Neural network-based BC models have been widely adopted for developing and evaluating automated vehicle algorithms and are a common benchmark for evaluating novel data-driven models [16].

The relative strengths of rule-based and data-driven approaches suggest that there is a role for model structure (to aid in interpretability), especially structure grounded in psychological theory and learning from data (to aid in flexibility) in car following model development. The incorporation of these two concepts requires a shift to contemporary theories of human cognition. One relevant theory is active inference [17,18], a framework developed from Bayesian principles of cognition. The central ideas of active inference are that 1) humans have internal probabilistic generative models of the environment and that 2) humans leverage their model of the environment to make inferences about action courses that reduce surprise in terms of both distance from their desired states of the environment and uncertainty. Importantly, these principles have been translated into a quantitative framework for modeling human behavior and cognition [19]. The quantitative framework includes an explicit representation of agent belief dynamics to facilitate agent decision making and action selection in response to observed perceptual signals. Due to this structure, the model is fundamentally interpretable (i.e., actions can be traced back to beliefs and observations at a given time). On the other hand, the increased complexity and probabilistic nature of the model compared to rule-based frameworks also increase its flexibility and potentially its generalizability. Recently, the active inference framework has been extended to driving to depict driving behavior during emergency scenarios with some success [20, 21]. However, the application to broader scenarios has been limited.

Our goal in this project is to develop an Active Inference Driving Agent (AIDA), evaluate its flexibility and generalizability relative to rule-based and data-driven benchmarks, and illustrate the interpretability of the model and the resulting insights it provides into car following behavior. The details of this work have been described in an article that is currently available as a preprint [25].

Method

Active inference [17, 18] is a novel framework for cognition and behavior according to which the agent jointly perceives and acts upon the world to minimize the mismatch between perceived vs preferred states of the world. The process of developing an Active Inference Driving Agent requires model specification, model fitting, and model validation. Model specification requires translating Active inference theory to a quantitative framework then identifying observations, actions, and model belief updates.

In this work, we chose to implement the active interference model using a Partially Observable Markov Decision Process (POMDP). A POMDP describes a dynamic process in which the state of the environment evolves as a function of the driver action's and the current state of the driving environment. The state of the environment is assumed to be not directly observable but obeservations related to the state are available to the driver. This situation is consistent with many driving scenarios; for example, while following a lead vehicle a driver may observe brake lights and a change in the lead vehicle's speed but may not observe the reason for the lead vehicle braking (e.g., slowing to turn, slowing to prevent a collision). For our AIDA implementation, we specified that the driver observations to include the relative velocity between the driver's vehicle and the lead vehicle, the distance to the lead vehicle, and the visual looming (i.e., τ^{-1}) of the lead vehicle. These parameters were selected based on prior driver braking models and driving theory. We further specified that the driver's actions consisted of continuous acceleration input.

After initial specification, we specified the model action selection process. We based this process on existing active inference implementations in other contexts. In our implementation the driver uses the history of driving observations to infer the underlying state of the environment at every time step by using Bayes rule (see Figure 1 below), then selects an action that minimizes a quantity called free energy (G in Figure 1). Free energy has two components: (i) a measure of *pragmatic* value that quantifies the extent to which the driver's beliefs about the state of the world resulting from implementing a given control action differs from their preferred distribution of the state of the world and (ii) a measure of *epistemic* value defined as a measure of the uncertainty about future observations induced by control actions given the driver's current beliefs about the environment.

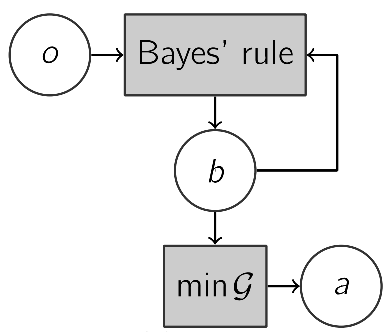


Figure 1. Active Inference Driving Agent (AIDA): o = instantaneous observation, a = control action, b = instantaneous belief, *G* = expected free energy

The specified active inference model contains free parameters associated with the observation distributions and belief transition matricies. We learned these parameters from driver behavior captured by the INTERACTION dataset. The INTERACTION dataset [22] is a publicly available driving dataset recorded using drones on fixed road segments in the USA, Germany, and China. The dataset provides a set of time-indexed trajectories of the positions, velocities, and headings of each vehicle in the road segment in the map's coordinate system at a sampling frequency of 10 Hz, and the vehicle's length and width for each road segment. The dataset contains a variety of traffic behaviors, including car following, free-flow traffic, and merges. We used the trajectories from this dataset to estimate the model free parameters using maximum likelihood estimation.

In addition to the AIDA model, we trained three comparison algorithms: IDM, BC (BC-MLP), and a recursive BC (BC-RNN) using the same maximum likelihood estimation process. These models were selected for comparison to understand the differences between AIDA and rule-based (i.e., IDM) and data-based (i.e, BC-MLP; BC-RNN) modeling approaches. We compared the models according to their accuracy in predicting trajectories and their interpretability. Note that additional details on the model specification, fitting, and evaluation process can be found in [25].

Results

Model Interpretability

Initial insights into the AIDA model input and output connections can be gained by visualizing the AIDA components, specifically its policy (Figure 2b), observation distribution (shown in Figure 2c), and preference distribution (Figure 2d). These figures show 200 random samples from each state of the AIDA's state-conditioned observation distribution, plotted on each pair of observation modalities. Color is used to highlight relevant quantities of interest.

We further used samples drawn from the INTERACTION dataset, plotted in Figure 2a and colored by the recorded accelerations, to facilitate interpreting the AIDA samples. Figure 2b illustrates the observation samples by the model's chosen control actions. The top chart shows the samples using distance headway (d; x-axis) by relative velocity to the lead vehicle (Δv ; y-axis), the middle chart shows distance headway by τ^{-1} which is defined as the rate of change of the visual angle of the lead vehicle from the ego driver's seat position divided by the angle itself. Finally, the bottom chart shows relative velocity by τ^{-1}. The shape of the sampled points matches the contour of the empirical dataset (Figure 2a), particularly in the middle and bottom visualizations, which suggests that the model's learned observations align with the recorded observations in the dataset. Darker green and red colors correspond to larger acceleration and deceleration magnitudes, respectively, and light-yellow color corresponds to near zero control inputs. The color gradient at different regions in Figure 2b is consistent with that of the empirical dataset shown in Figure 2a. This shows that the model learned a similar observation to action mapping as the empirical dataset. The mapping can be interpreted as the tendency to choose negative accelerations when the relative speed and τ^{-1} are negative and the distance headway is small, and positive accelerations in the opposite case. Furthermore, the sensitivity of the red and green color gradients with respect to distance headway shows that the model tends to accelerate whenever there is positive relative velocity, regardless of the distance headway. However, it tends to input smaller deceleration at large distance headway for the same level of relative speed. Figure 2c shows the observation samples colored by their associated discrete states. The juxtaposition of color clusters in the top panel shows that the AIDA learned to categorize observations by relative speed and distance headway and its categorization for relative speed is more fine-grained at small distance headways and spans a larger range of values. The middle and bottom panels show that its categorization of relative speed is highly correlated with τ^{-1}.

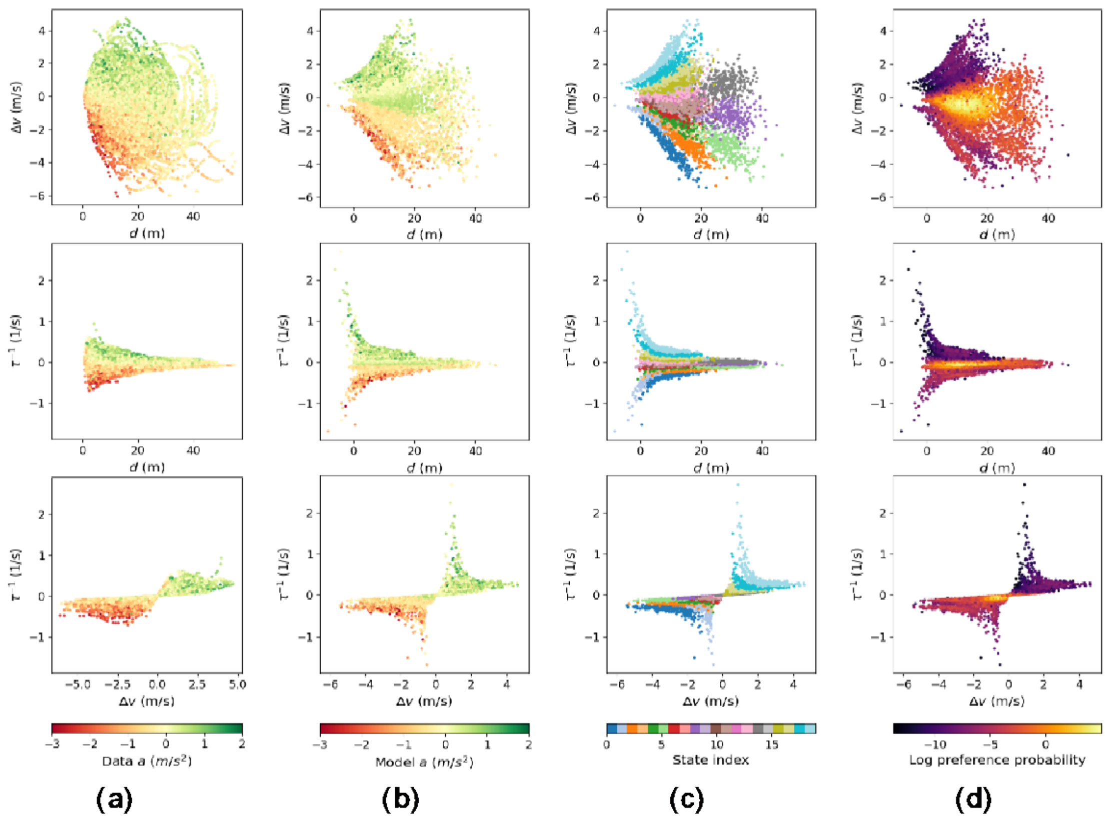


Figure 2. Visualizations of the dataset and AIDA model components. (a), plot of random observations sampled from the dataset; (b), (c), and (d) plot of samples from model.

Statistical Comparison with Rule-based (IDM) and Data-Driven Models (BC).

Following the recommendations in [23, 24] for evaluating learned control policies, we represented the central tendency of a model's offline prediction and online control performance using the interquartile mean (IQM) of the offline mean absolute error (MAE) and online average deviation error (ADE). For collision rate, we compute the regular mean instead of IQM to account for the collision rate lower bound of 0. The IQMs are computed by 1) ranking all tested trajectories by their respective performance metrics and 2) computing the mean of the performance metrics ranked in the middle 50%.

Figure 3 shows the offline evaluation results for each model with the model type on the x-axis and the IQMs of acceleration prediction MAEs averaged across the testing dataset on the y-axis. The color of the points in the figure represents the testing condition and each point corresponds to a random seed's result. The points are randomly distributed around each x-axis label for clarity. Dispersion on the y-axis indicates sensitivity in the model to initial training conditions. The plot illustrates that the AIDA had the lowest MAE-IQM in the same-lane tests, followed by BC-RNN, BC-MLP, and IDM. T

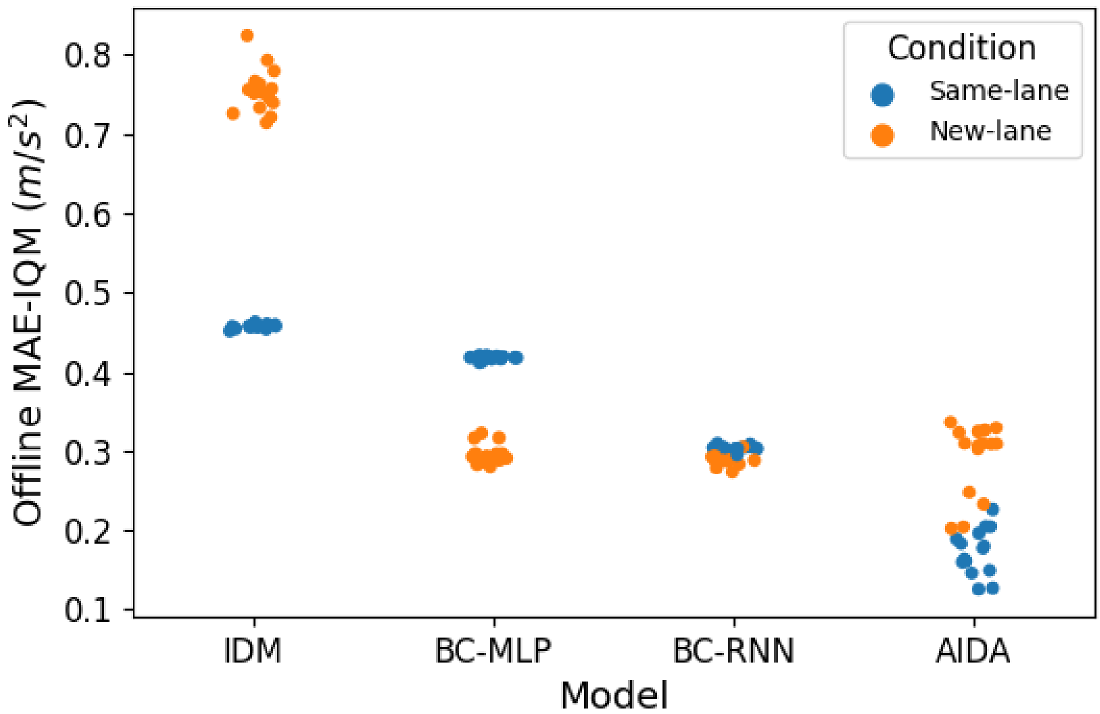


Figure 3. Offline evaluation MAE-IQM. Each point corresponds to a random seed used to initialize model training and its color corresponds to the testing condition of either same-lane or new-lane.

Figure 4 shows the IQM of each model's ADEs from data set trajectories in the online evaluations using the same format as the offline evaluation results. In the same-lane testing condition, all models had an ADE-IQM values between 1.8 m and 2.8 m, which is less than the length of a standard sedan (approx. 4.8 m). Among all models, BC-MLP achieved the lowest ADE values for both the same-lane and new-lane conditions, followed by the AIDA, IDM, and BC-RNN.

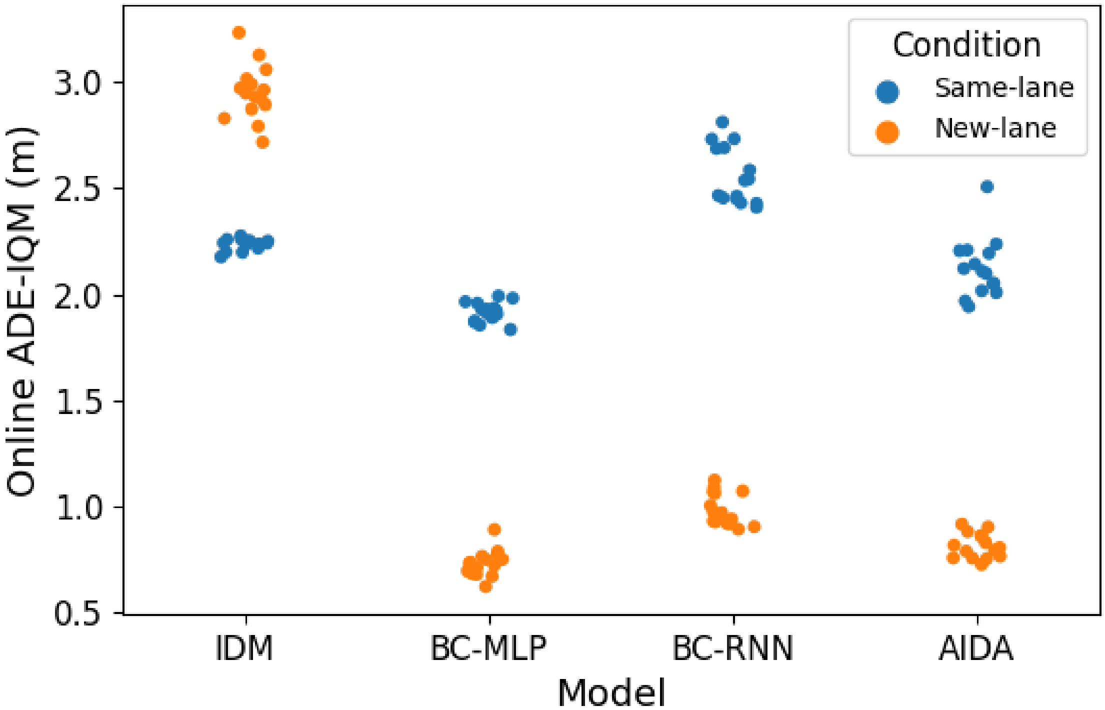


Figure 4. Online evaluation ADE-IQM. Each point corresponds to a random seed used to initialize model training and its color corresponds to the testing condition of either same-lane or new-lane.

To compare the central performance difference between the AIDA and baseline models, we performed two-sided Welch's t-tests with 5 percent rejection level on the MAE-IQM and ADE-IQM values computed from different random seeds with the assumption that the performance distributions between two models may have different variances. These are reported in [25].

Discussion

In this project, we introduced and evaluated a novel active inference model of driver car following behavior (AIDA). The proposed AIDA significantly outperformed the IDM and neural network BC models in offline predictions in the same-lane condition and outperformed the IDM while performing similarly to BC models in the new-lane condition. Additionally, the AIDA achieved significantly lower average deviation error than the rules-based model IDM and data-driven model BC-RNN in the online control settings. However, the results showed that the AIDA was sensitive to initial training conditions, which resulted in higher rates of lead vehicle collisions in the same-lane condition compared to the IDM and BC-MLP. While BC had comparable or better performance than the AIDA in action prediction and control, the AIDA is substantially more interpretable than BC models. In contrast to approximate explanatory methods for BC neural networks, we showed that the AIDA's decision making process can be directly accessed by sampling and visualizing the AIDA distributions. Further, we illustrated how the AIDA's joint belief and action trajectories could be used to understand model errors and correct them. This level of understanding and diagnostic analysis is central to real world model inspection and verification which are essential components of interpretability.

These results partially confirm our hypothesis that balancing the relative strengths of rule-based and data-driven models, specifically using the active inference framework, results in better predictions of driver behavior and more nuanced understanding of driver cognitive dynamics during car following. In contrast to fixed rule-based models like the IDM, the AIDA can incorporate additional "rules" in its state and policy priors while maintaining the flexibility provided by its probabilistic representation. In contrast to purely data-driven models, learning in the AIDA is constrained by its probability distributions and structure. This balance preserves interpretability but still allows the model to be flexible to new data. Our findings here suggest that this flexibility comes at a cost of sensitivity to local optima in the training process as evidenced by the collision rates across random seeds in online evaluations. Our findings here extend prior applications of active inference theory in driving and driver models and illustrate the value of rule-based modeling.

Our work is limited by the following aspects. First, we have assumed three driver observation modalities: distance headway, relative speed, and τ^{-1} with respect to the lead vehicle. However, human drivers are known to monitor other surrounding vehicles while driving and have broader visual sampling. Second, our parameterization of discrete states has limited the expressivity of the model and prevented inductive biases such as the smoothness of physical dynamics from being encoded. The limited dataset coverage, e.g., the lack of crashes, prevented the learned dynamics from generalizing to some out-of-distribution scenarios. The combination of model and data insufficiency led to the difficulty of recognizing near-crash states and resulted in substantially more lead vehicle crashes than BC and the IDM.

Conclusions and Recommendations

We proposed a novel active inference model of driver behavior (AIDA). Using car following data, we showed that the AIDA significantly outperformed the rule-based IDM on all metrics and performed comparably with the data-driven neural network benchmarks. Using an interpretability analysis, we showed that the structure of the AIDA provides superior transparency of its input-output mechanics than the neural network models. Future work should focus on training with data from more diverse driving environments and examining model extensions that can capture heterogeneity across drivers.

While we anticipate incorporating additional observations and higher state space dimension and application to alternative driving scenarios to be easy under the current model formulation, doing so would impose additional requirements on dataset quality and diversity. We thus recommend future work to consider general methods for incorporating domain knowledge in more expressive generative models to combat dataset limitations and modeling heterogeneity in naturalistic driver behavior. The results here suggest that these extensions may alleviate many of the current model limitations.

Additional Products

Publications:

1. **Wei R., McDonald A. D., Garcia A. and Alambeigi H. (2022) Modeling Driver Responses to Automation Failures With Active Inference, *IEEE Transactions on Intelligent Transportation Systems* 23 (10) 18064-18075, doi: 10.1109/TITS.2022.3155381.**

2. **Wei R., Garcia, A., McDonald A.D., Markkula G., Engstrom J., Supeene I. and O'Kelly M., (2022) World Model Learning From Demonstrations With Active Inference: Application to Driving Behavior, 3rd International Workshop on Active Inference (IWAI).**

3. **Wei R., McDonald A.D., Garcia, A., Markkula G., Engstrom J. and O'Kelly M., (2023) An active inference model of car following: Advantages and applications, doi: 10.48550/arXiv.2303.15201**

4. **Wei R., Zeng S., Li Ch., Garcia A., McDonald A. and Hong M. (2023) Robust Inverse Reinforcement Learning Through Bayesian Theory of Mind, ICML 2023 Workshop on Theory of Mind in Communicating Agents**

Education and Workforce Development Products

1. PhD Thesis "LEARNING REPRESENTATIONS OF COGNITIVE DYNAMICS AND DECISION MAKING IN HUMAN DRIVERS" by Ran Wei, (June 2023).

2. Lecture material covering Bayesian Models of perception and action for the course ISEN 427 "Decision Making Under Uncertainty"

Technology Transfer Products

Nothing to report.

Data Products

Nothing to report.

References

1. Scheel, O., L. Bergamini, M. Wolczyk, B. Osiński, and P. Ondruska, Urban driver: Learning to drive from real-world demonstrations using policy gradients. In Conference on Robot Learning, PMLR, 2022, pp. 718-728.
2. Scanlon, J. M., K. D. Kusano, T. Daniel, C. Alderson, A. Ogle, and T. Victor, Waymo simulated driving behavior in reconstructed fatal crashes within an autonomous vehicle operating domain. Accident Analysis & Prevention, Vol. 163, 2021, p. 106-154.
3. Bärgman, J., C.-N. Boda, and M. Dozza, Counterfactual simulations applied to SHRP2 crashes: The effect of driver behavior models on safety benefit estimations of intelligent safety systems. Accident Analysis & Prevention, Vol. 102, 2017, pp. 165-180.
4. Roesener, C., J. Hiller, H. Weber, and L. Eckstein, How safe is automated driving? Human driver models for safety performance assessment. In 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), IEEE, 2017, pp. 1-7.
5. Rhinehart, N., R. McAllister, K. Kitani, and S. Levine, Precog: Prediction conditioned on goals in visual multi-agent settings. In Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 2821-2830.
6. Brown, B., The social life of autonomous cars. Computer, Vol. 50, No. 2, 2017, pp. 92-96.
7. Markkula, G., Y.-S. Lin, A. R. Srinivasan, J. Billington, M. Leonetti, A. H. Kalantari, Y. Yang, Y. M. Lee, R. Madigan, and N. Merat, Explaining human interactions on the road requires large-scale integration of psychological theory, 2022.
8. Räukur, T., A. Ho, S. Casper, and D. Hadfield-Menell, Toward transparent AI: A survey on interpreting the inner structures of deep neural networks. arXiv preprint arXiv:2207.13243, 2022.
9. Alambeigi, H., A. D. McDonald, and S. R. Tankasala, Crash Themes in Automated Vehicles: A Topic Modeling Analysis of the California Department of Motor Vehicles Automated Vehicle Crash Database, 2020.
10. Novakazi, F., M. Johansson, H. Strömberg, and M. Karlsson, Levels of what? Investigating drivers' understanding of different levels of automation in vehicles. Journal of Cognitive Engineering and Decision Making, Vol. 15, No. 2-3, 2021, pp. 116-132.
11. Saifuzzaman, M. and Z. Zheng, Incorporating human-factors in car-following models: a review of recent developments and research needs. Transportation research part C: emerging technologies, Vol. 48, 2014, pp. 379-403
12. Hamdar, S. H. and H. S. Mahmassani, From existing accident-free car-following models to colliding vehicles: exploration and assessment. Transportation research record, Vol. 2088, No. 1, 2008, pp. 45-56.
13. Zhou, M., X. Qu, and X. Li, A recurrent neural network based microscopic car following model to predict traffic oscillation. Transportation research part C: emerging technologies, Vol. 84, 2017, pp. 245-264
14. Zhu, M., X. Wang, and Y. Wang, Human-like autonomous car-following model with deep reinforcement learning. Transportation research part C: emerging technologies, Vol. 97, 2018, pp. 348-368.
15. Bhattacharyya, R., B. Wulfe, D. Phillips, A. Kuefler, J. Morton, R. Senanayake, and M. Kochenderfer, Modeling human driving behavior through generative adversarial imitation learning. arXiv preprint arXiv:2006.06412, 2020.

16. Igl, M., D. Kim, A. Kuefler, P. Mougin, P. Shah, K. Shiarlis, D. Anguelov, M. Palatucci, B. White, and S. Whiteson, Symphony: Learning Realistic and Diverse Agents for Autonomous Driving Simulation. arXiv preprint arXiv:2205.03195, 2022
17. Friston, K., The free-energy principle: a unified brain theory? Nature reviews neuroscience, Vol. 11, No. 2, 2010, pp. 127-138.
18. Friston, K., T. FitzGerald, F. Rigoli, P. Schwartenbeck, and G. Pezzulo, Active inference: a process theory. Neural computation, Vol. 29, No. 1, 2017, pp. 1-49.
19. Smith, R., P. Badcock, and K. J. Friston, Recent advances in the application of predictive coding and active inference models within clinical neuroscience. Psychiatry and Clinical Neurosciences, Vol. 75, No. 1, 2021, pp. 3-13
20. Wei, R., A. D. McDonald, A. Garcia, and H. Alambeigi, Modeling Driver Responses to Automation Failures With Active Inference. IEEE Transactions on Intelligent Transportation Systems, 2022.
21. Engström, J., S.-Y. Liu, A. Dinparastdjadid, and C. Simoiu, Modeling road user response timing in naturalistic settings: a surprise-based framework. arXiv preprint arXiv:2208.08651, 2022.
22. Zhan, W., L. Sun, D. Wang, H. Shi, A. Clausse, M. Naumann, J. Kummerle, H. Konigshof, C. Stiller, A. de La Fortelle, et al., Interaction dataset: An international, adversarial and cooperative motion dataset in interactive driving scenarios with semantic maps. arXiv preprint arXiv:1910.03088, 2019.
23. Colas, C., O. Sigaud, and P.-Y. Oudeyer, A hitchhiker's guide to statistical comparisons of reinforcement learning algorithms. arXiv preprint arXiv:1904.06979, 2019.
24. Agarwal, R., M. Schwarzer, P. S. Castro, A. C. Courville, and M. Bellemare, Deep reinforcement learning at the edge of the statistical precipice. Advances in neural information processing systems, Vol. 34, 2021, pp. 29304-29320
25. **Wei R., McDonald A.D., Garcia, A., Markkula G., Engstrom J. and O'Kelly M., (2023) An active inference model of car following: Advantages and applications, doi: 10.48550/arXiv.2303.15201**