Extended Abstract: Benchmarking Behavior Prediction Models in Gap Acceptance Scenarios

Schumann, Julian F.; Kober, Jens; Zgonnikov, Arkady

DOI
10.1109/IV55156.2024.10588499

Publication date
2024

Document Version
Final published version

Published in
Proceedings of the 35th IEEE Intelligent Vehicles Symposium, IV 2024

Citation (APA)
Schumann, J. F., Kober, J., & Zgonnikov, A. (2024). Extended Abstract: Benchmarking Behavior Prediction Models in Gap Acceptance Scenarios. In *Proceedings of the 35th IEEE Intelligent Vehicles Symposium, IV 2024* (pp. 3148). (IEEE Intelligent Vehicles Symposium, Proceedings). IEEE. https://doi.org/10.1109/IV55156.2024.10588499

2024 IEEE Intelligent Vehicles Symposium (IV)
June 2-5, 2024. Jeju Island, Korea

2024 IEEE Intelligent Vehicle Symposium (IV) | 979-8-3503-4881-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/IV55156.2024.10588499

Extended Abstract: Benchmarking Behavior Prediction Models in Gap Acceptance Scenarios

Julian F. Schumann, Jens Kober, Arkady Zgonnikov

***Abstract*— Autonomous vehicles currently suffer from a time-inefficient driving style caused by uncertainty about human behavior, which could be improved by accurate and reliable prediction models enabling more efficient trajectory planning. However, the evaluation of such models is commonly over-simplistic, ignoring the asymmetric importance of prediction errors and the heterogeneity of the datasets used for testing. We examine the potential of recasting interactions between vehicles as gap acceptance scenarios and evaluating models in this structured environment. To that end, we develop a framework aiming to facilitate the evaluation of any model, by any metric, and in any scenario. We then apply this framework to state-of-the-art prediction models, which all show themselves to be unreliable in the most safety-critical situations.**

Introduction: Under the goal of fulfilling the highest safety standards, the uncertainty inherent to human behavior forces current versions of autonomous vehicles (AVs) to adopt a very conservative driving style. This can be mitigated by using accurate and reliable prediction models limiting this uncertainty, with the current state-of-the-art mainly relying on large deep-learning models [1]. However, as those models are usually tested on random subsets of large datasets, their reliability in safety critical scenarios is uncertain.

Consequently, we suggest to test such models specifically in safety-critical interactions, most of which can be classified as so called gap acceptance scenarios [2]. There, a human without priority has to decide to either cut across in front of an oncoming AV or let it pass. Under this definition, we can estimate the criticality of a particular situation, allowing us to assemble a test set including the most difficult prediction problems. To that end, we propose a framework for benchmarking prediction models in gap acceptance (Figure 1), and apply it to a number of state of the art prediction models.

Methods and Results: Our framework, consisting out of four main parts, is designed with the goal of an easy addition of new instances. This are the dataset, which contains some form of gap acceptance scenarios, its splitting into train and test set (either random, or prioritizing safety-critical scenarios for testing), the prediction model trained on the training data, and the metrics used for its evaluation. Here, metrics and models can be be based on trajectory prediction or the binary classification of the gap acceptance problem. The framework allows the transformation between both kind of model outputs, by either extracting the binary decision or using the predicted binary probability as input to a conditional trajectory predictor.

We use the framework with a number of state-of-the-art trajectory prediction methods and binary classification methods on different gap acceptance scenarios (Figure 1). We find that earlier prediction generally lead to worse binary predictions, especially for trajectory predictors, which in many cases could be outperformed by simple binary classifiers. However, both kind of models consistently showed a significant decrease in performance when tested on the most safety-critical scenarios. Finally, we found that combining binary classifiers with a conditional trajectory predictor can lead to small improvements.

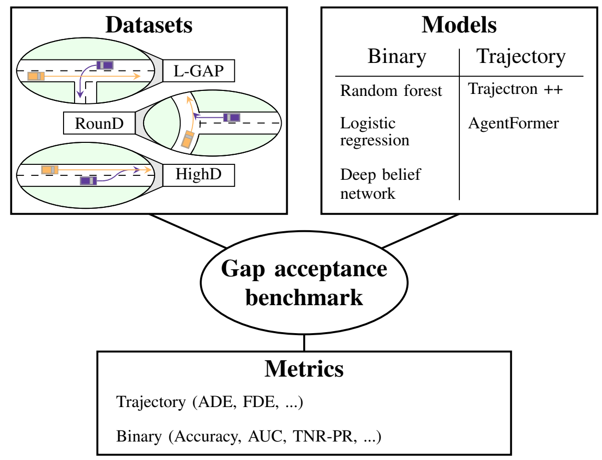

Fig. 1. The proposed framework for evaluating prediction models for human behavior with several metrics on different gap acceptance datasets [3]

Conclusion: Our proposed framework showed that state-of-the-art behavior prediction models are unreliable in safety-critical gap acceptance scenarios. However, we found that such scenarios are generally rare, so new data sets should focus primarily of including as many of them as possible. The framework could benefit from expanding to also include non-gap-acceptance scenarios, enabling a more holistic testing. Crucially, we argue that only testing prediction models on a randomly selected test cases that do not involve any critical interaction is insufficient for an effective evaluation of these models' usefulness as a part of an autonomous vehicle's planning algorithm.

Full paper: See [3] for further details.

All authors are with the Cognitive Robotics Department, TU Delft, 2628 CB Delft, The Netherlands {J.F.Schumann, J.Kober, A.Zgonnikov}@tudelft.nl

[1] S. Mozaffari, O. Y. Al-Jarrah, M. Dianati, P. Jennings, and A. Mouzakitis, "Deep Learning-Based Vehicle Behavior Prediction for Autonomous Driving Applications: A Review," *IEEE Trans. on Intell. Transp. Syst.*, vol. 23, pp. 33–47, Jan. 2022.
[2] G. Markkula, R. Madigan, D. Nathanael, E. Portouli, Y. M. Lee, A. Dietrich, J. Billington, A. Schieben, and N. Merat, "Defining interactions: a conceptual framework for understanding interactive behaviour in human and automated road traffic," *Theor. Issues Ergon. Sci.*, 2020.
[3] J. F. Schumann, J. Kober, and A. Zgonnikov, "Benchmarking Behavior Prediction Models in Gap Acceptance Scenarios," *IEEE Trans. on Intell. Veh.*, vol. 8, pp. 2580–2591, Mar. 2023.