Perception Beyond Sensors Under Uncertainty

Masha Itkina
Department of Aeronautics and Astronautics
Stanford University
mitkina@stanford.edu

Abstract

My research aims to enable spatiotemporal inference in mobile robot perception systems. Specifically, the proposed thesis presents learning-based approaches to the tasks of behavior prediction and occlusion inference that explicitly account for the associated aleatoric and epistemic uncertainty.

Introduction

To safely navigate complex, dynamic environments, autonomous vehicles (AVs) must anticipate the behaviors of other agents and make inferences into occluded regions in the scene, thus informing proactive downstream path planning. Predicting human behavior in time is difficult due to (1) the variability in human decision making processes (e.g., driver aggressiveness, distracted driving, etc.) and (2) the rapid evolution of human behavior in time (e.g., interactions with other drivers). Recent work has successfully used neural networks to perform temporal predictions both in terms of continuous (Chai et al. 2019; Salzmann et al. 2020) and discrete (Itkina, Driggs-Campbell, and Kochenderfer 2019; Lange, Itkina, and Kochenderfer 2020) representations.

Due to the variability and stochasticity of human behaviors, models must account for prediction uncertainty. Real-world systems are subject to two types of uncertainty: *aleatoric* and *epistemic*. The former results from data uncertainty (e.g., two equally valid predictions given an input) and is irreducible. The latter arises from (1) how well the model represents the data and (2) from data distribution shift (Malinin and Gales 2018). Aleatoric uncertainty can often be modelled explicitly during training. For instance, for the trajectory prediction task, parameters for a Gaussian distribution over the output can be learned (Chai et al. 2019) or variational approaches can be used to approximate the prediction uncertainty (Salzmann et al. 2020; Itkina et al. 2020).

Spatially, physical sensor limitations cause occlusions, which result in overly cautious AV behavior. Human drivers, in contrast, use information gathering and inference techniques based on their observed surroundings to progress in driving maneuvers despite spatial uncertainty. For instance, if an observed driver brakes sharply in a neighboring lane, this may indicate the presence of an occluded obstacle (e.g., a pedestrian) ahead. This occluded obstacle may appear in the path of the ego driver, causing the ego to act cautiously. If the observed driver proceeds nominally, a dynamic obstacle is unlikely to be occluded, and the ego may proceed with lower levels of caution. Learning-based approaches (Afolabi et al. 2018; Dequaire et al. 2018; Sun et al. 2019) have shown success in inferring occupancy within occluded regions from observed driver behaviors. However, they do not explicitly model the multimodality of the occluded region's occupancy, failing to capture the aleatoric uncertainty.

Furthermore, although neural networks are capable of modeling aleatoric uncertainty, they have been shown to make unreliable predictions for out-of-distribution (OOD) inputs (Malinin and Gales 2018; Lee et al. 2018). For AVs, it is safety critical that the prediction module is able to provide a level of confidence in its output to a downstream planner. This level of confidence should encode the epistemic uncertainty associated with the input relative to the training set. Existing methods for capturing epistemic uncertainty within neural networks are often applied to simple classification tasks (Sensoy et al. 2020; Malinin and Gales 2018) or small regression problems (Blundell et al. 2015). Encoding the epistemic uncertainty for such a complex, multimodal task as behavior prediction remains an open problem.

The proposed thesis extends AV perception beyond onboard sensors by making inferences in time and in occluded regions, while modeling aleatoric and epistemic uncertainty.

Anticipated Contributions

Environment prediction (Itkina, Driggs-Campbell, and Kochenderfer 2019; Lange, Itkina, and Kochenderfer 2020) We frame the problem of environment prediction in an urban setting as a video frame prediction task. We validate the capacity of a convolutional long short-term memory (ConvLSTM) network to predict the environment in time. We show that a ConvLSTM is able to learn the internal dynamic representation of the environment allowing for prediction from static occupancy grid data without additional dynamic information. We compare the benefits of an evidential occupancy grid to that of a probabilistic alternative.

Sparse multimodal latent spaces for aleatoric uncertainty estimation (Itkina et al. 2020) Prohibitively large discrete latent spaces in conditional variational autoencoders (CVAEs) are required to accurately learn complex data distributions (e.g., for behavior prediction (Salzmann et al. 2020)) causing computational difficulties for downstream tasks, such as motion planning or robot information sharing. We present a post hoc method for identifying the subset of discrete latent classes that is most representative of the input using evidential theory (Dempster 2008), thus sparsifying the latent space while maintaining distributional multimodality. Our algorithm achieves a significant reduction in the discrete latent sample space of CVAEs in image generation and behavior prediction tasks without loss of performance. Our approach outperforms baseline techniques which collapse the multimodality by removing important modes with overly-aggressive filtering.

Variational occlusion inference [In Progress] To better inform occlusion inference, we extend work by Afolabi et al. (2018) to learn a more expressive model that maps observed driver behavior to the environment ahead of the driver, thus capturing interactions between observed drivers and occluded obstacles. Our proposed approach models the multimodality in the potential mappings using a CVAE, thereby accounting for the aleatoric uncertainty associated with the spatial prediction. We introduce a multi-agent fusion mechanism using evidential theory (Dempster 2008) for occlusion inference. The mechanism fuses the spatial predictions, represented as occupancy grids, inferred from multiple observed driver behaviors into the AV's environment map. We demonstrate real-time capability of the occlusion inference algorithm on data collected from a real-world experiment.

Self-aware neural networks for robust behavior prediction [Proposed Work] We propose modeling epistemic uncertainty for behavior prediction by learning parameters for higher-order distributions, which are distributions over sets of distributions (e.g., the Dirichlet is a distribution over categorical distributions). These methods are efficient and have demonstrated state-of-the-art results (Malinin and Gales 2018; Sensoy et al. 2020). We will extend an existing behavior prediction architecture for continuous (Salzmann et al. 2020; Chai et al. 2019) or discrete (Itkina, Driggs-Campbell, and Kochenderfer 2019; Lange, Itkina, and Kochenderfer 2020) environment representations to account for both aleatoric and epistemic uncertainty. The epistemic uncertainty may be learned over a discrete latent space or for the full output prediction, modeling the uncertainty over a classification or a regression task, respectively.

Since the behavior prediction task is high dimensional and it is impossible to predict all potential scenarios encountered on the road, we will not be able to manually select an OOD dataset for training. We will instead learn OOD data following an approach similar to that of Sensoy et al. (2020). OOD samples will be generated that are close to the training data in the continuous latent space of a VAE, but far from the training data in the original data space. This research will output a behavior prediction architecture that is self-aware of its prediction confidence, and thus, able to provide rich information to a downstream planner.

The following is a tentative one year timeline for the proposed work. **(Months 1-2)** I will implement and train an existing behavior prediction network that estimates aleatoric uncertainty. **(Months 3-5)** The network will be augmented to model epistemic uncertainty over discrete modes and trained on publicly available data. Its performance will be tested using several OOD data variations (e.g., trajectories generated with adversarial noise, trajectories from highway versus residential scenes, etc.). **(Months 6-7)** The proposed model will be evaluated against existing baselines. **(Months 8-12)** I will iterate on the method to improve epistemic uncertainty estimation.

References

Afolabi, O.; Driggs-Campbell, K.; Dong, R.; Kochenderfer, M. J.; and Sastry, S. S. 2018. People as Sensors: Imputing Maps from Human Actions. In *International Conference on Intelligent Robots and Systems (IROS)*, 2342–2348. IEEE/RSJ.

Blundell, C.; Cornebise, J.; Kavukcuoglu, K.; and Wierstra, D. 2015. Weight Uncertainty in Neural Networks. In *International Conference on Machine Learning (ICML)*, volume 37, 1613–1622.

Chai, Y.; Sapp, B.; Bansal, M.; and Anguelov, D. 2019. MultiPath: Multiple Probabilistic Anchor Trajectory Hypotheses for Behavior Prediction. In *Conference on Robot Learning (CoRL)*, 86–99.

Dempster, A. P. 2008. A Generalization of Bayesian Inference. *Classic Works of the Dempster-Shafer Theory of Belief Functions* 73–104.

Dequaire, J.; Ondrúška, P.; Rao, D.; Wang, D.; and Posner, I. 2018. Deep Tracking in the Wild: End-to-End Tracking Using Recurrent Neural Networks. *International Journal of Robotics Research* 37(4-5): 492–512.

Itkina, M.; Driggs-Campbell, K.; and Kochenderfer, M. J. 2019. Dynamic Environment Prediction in Urban Scenes using Recurrent Representation Learning. In *International Conference on Intelligent Transportation Systems (ITSC)*, 2052–2059. IEEE.

Itkina, M.; Ivanovic, B.; Senanayake, R.; Kochenderfer, M. J.; and Pavone, M. 2020. Evidential Sparsification of Multimodal Latent Spaces in Conditional Variational Autoencoders. In *Advances in Neural Information Processing Systems (NeurIPS)*.

Lange, B.; Itkina, M.; and Kochenderfer, M. J. 2020. Attention Augmented ConvLSTM for Environment Prediction. *ArXiv* .

Lee, K.; Lee, K.; Lee, H.; and Shin, J. 2018. A Simple Unified Framework for Detecting Out-of-Distribution Samples and Adversarial Attacks. In *Advances in Neural Information Processing Systems (NeurIPS)*, 7167–7177.

Malinin, A.; and Gales, M. 2018. Predictive Uncertainty Estimation via Prior Networks. In *Advances in Neural Information Processing Systems (NeurIPS)*, 7047–7058.

Salzmann, T.; Ivanovic, B.; Chakravarty, P.; and Pavone, M. 2020. Trajectron++: Dynamically-Feasible Trajectory Forecasting With Heterogeneous Data. In *European Conference on Computer Vision (ECCV)*.

Sensoy, M.; Kaplan, L.; Cerutti, F.; and Saleki, M. 2020. Uncertainty-Aware Deep Classifiers Using Generative Models. In *Conference on Artificial Intelligence (AAAI)*.

Sun, L.; Zhan, W.; Chan, C.-Y.; and Tomizuka, M. 2019. Behavior Planning of Autonomous Cars with Social Perception. In *Intelligent Vehicles Symposium (IV)*, 207–213. IEEE.