2017 IEEE Intelligent Vehicles Symposium (IV)
June 11-14, 2017, Redondo Beach, CA, USA

Generalizable Intention Prediction of Human Drivers at Intersections

Derek J. Phillips, Tim A. Wheeler, and Mykel J. Kochenderfer

Abstract— **Effective navigation of urban environments is a primary challenge remaining in the development of autonomous vehicles. Intersections come in many shapes and forms, making it difficult to find features and models that generalize across intersection types. New and traditional features are used to train several intersection intention models on real-world intersection data, and a new class of recurrent neural networks, Long Short Term Memory networks (LSTMs), are shown to outperform the state of the art. The models predict whether a driver will turn left, turn right, or continue straight up to 150 m with consistent accuracy before reaching the intersection. The results show promise for further use of LSTMs, with the mean cross validated prediction accuracy averaging over 85% for both three and four-way intersections, obtaining 83% for the highest throughput intersection.**

I. Introduction

This work seeks models that can effectively predict the intentions of human drivers as they approach an intersection. Intersection navigation is a crucial task for any urban driving platform, and models for predicting driver intentions at intersections are crucial to the development of safe and effective automated driving policies and for human behavior models in simulation-based safety validation.

Intention prediction at intersections has been the subject of prior research, with work covering intention prediction in isolation or alongside motion planning [1]–[5]. Models for predicting complete trajectories have been developed as well, with intention inferred in the process. Such work has used Hidden Markov Models [2], [6], Gaussian Processes [7], Dynamic Bayesian Networks [8], Support Vector Machines [9], and inverse reinforcement learning [3].

An additional body of work approaches the problem of populated environment navigation and path planning, often framing the problem as a Markov Decision Process (MDP), including Partially Observable MDPs (POMDPs) [10], Interactive POMDPs [11], and Mixed Observability MDPs [12]. Unfortunately, these approaches must often rely on coarse discretizations of the state and actions spaces in order to make the problem tractable. Nonetheless, it is apparent that using intention to supplement vehicle states in navigating an intersection is a promising line of research. Sezer, Bandyopadhyay, Rus, *et al.* examined the decision marking problem of a car turning right at a T-intersection and succeeded in lowering the accident probability and intersection navigation duration by capitalizing on intention-aware planning [12].

The work of Tang, Khokhar, and Gupta motivates this paper. The authors build lane-level maps of intersections and use a variety of models to predict driver intentions as they approach intersections [2]. They achieve 90% turn prediction accuracy 2.8 meters before the car enters the intersection. Their work is limited to four-way intersections, whereas our work generalizes between any type of intersection, and we evaluate on three- and four-way intersections.

We contribute to intersection prediction modeling in several ways. First, we address limitations in prior intersection prediction models that limit them from applying to any intersection composition. Second, we investigate several feature categories and suggest new features that generalize well across intersection types. Third, the proposed approach does not require high-fidelity maps of the intersection to be known ahead of time, and works well on simple layout representations.

II. Problem Definition

This work develops and evaluates models that classify the action a human driver will take at an upcoming intersection. A driver can take three possible actions a at the intersection, turning *left*, *right*, or continuing *straight*. Each model represents a conditional probability distribution $P(a \mid f)$ given a vector of features f. A good model will assign high likelihood to the correct action eventually taken by the driver at the intersection. Such models can be used in risk assessment for motion planning or driver behavior modeling.

Several challenges must be overcome in developing practical intersection intention prediction models. These challenges include developing models that generalize across unseen intersections and unseen drivers; that predict over longer prediction horizons, here with evaluating predictions up to 150 m before an intersection; and that use features that are generally available to an autonomous driving system or simulator, and thus do not require driver eye tracking or similar inputs. All models are probability distributions rather than mere classifiers, allowing them to capture the inherent stochasticity in human driving behavior.

The models presented below are evaluated according to both their classification accuracy and the likelihood assigned to withheld data. A high accuracy indicates good classification performance whereas a high likelihood indicates a good distribution fit.

III. Dataset

This work evaluates models on real driving data from the Next Generation Simulation (NGSIM) program conducted in 2005 by the Federal Highway Administration [13], [14]. Both the Lankershim and Peachtree datasets are used, totaling about 1 hour at 10 Hz. The NGSIM dataset provides the

The Department of Aeronautics and Astronautics & the Department of Computer Science, Stanford University, Stanford CA, 94305 USA. {djp42, wheelert, mykel}@stanford.edu

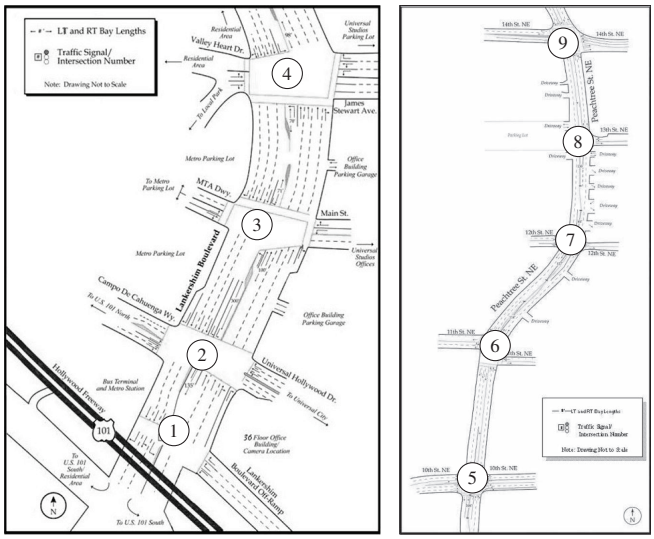

Fig. 1: Road networks for the Lankershim Blvd (left) and Peachtree Street (right) datasets and intersection labels.

positions for all vehicles over large road sections, providing a complete view of the intersection environment.

The training data was split into folds across the 9 intersections in order to assess the ability of models to generalize to unseen intersections. There are four intersections in the Lankershim dataset and five in the Peachtree dataset, as shown in Fig. 1. The roadways exhibit both differing road characteristics (wide versus narrow) and a combination of intersection types, including both three and four-way intersections.

IV. Features

A set of 104 features was extracted from the NGSIM data. The set includes *base features* from ego position and dynamics, *history features* from past states, *traffic features* based on neighboring vehicles, and *rule features* indicating legal actions at the upcoming intersection. Several features reflect those often used in the driving literature, whereas extensive use of traffic features is scarce [1], [2], [4]. Prior work has shown there to be a significant improvement in the performance of the models when traffic features are used [8]. Our use of simple rule features to provide a characterization of the road is also novel. The most similar prior work involved creating and using lane-level maps for that purpose [2].

Missing feature values are set to zero, such as headway distance when no lead vehicle is present. Potentially missing features include accompanying indicator features to indicate whether or not they are available.

A. Base Features

The base features used in all experiments are the magnitude of velocity and acceleration, the lane-relative heading, the number of lanes to the curb and to the median, the headway distance to the preceding vehicle (and associated

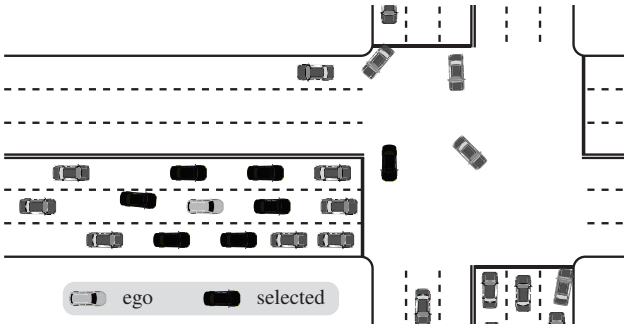

Fig. 2: The traffic features include information about the neighbors of the ego vehicle and the closest vehicle in the intersection.

indicator), and the distance to the intersection. These features are based on ego vehicle odometry and its immediate surroundings. Distance to the intersection is used instead of the time to intersection as the distance is invariant to cars stopping before passing through the intersection.

B. History Features

The base features only capture a single moment in time, so we add history features to provide several frames, thereby allowing models to identify short and long term maneuvers. The history features consist of base features extracted from the frames 0.5 s, 1 s, 2 s and 3 s in the past. If the frame does not exist, then the feature vector is a zero vector and the accompanying indicator feature is set to false.

C. Traffic Features

Other traffic participants strongly influence driving behavior. The traffic features include information about up to six direct neighbors of the ego vehicle and the closest vehicle, by Euclidean distance, currently in the intersection, as shown in Fig. 2. The magnitude of velocity and acceleration, the lane-relative heading, the headway distance (and indicator), and the distance to the ego vehicle are extracted for each of these vehicles. The lane features and distance to intersection are not included as they can be inferred from the ego base features. The traffic features are an original contribution, as the reviewed literature uses simple features such as headway to represent traffic, but the inclusion of more comprehensive features aims to illustrate how beneficial they can be.

D. Rule Features

The rule features are indicators for the legal actions in the current lane. They encode whether it is legal to go left, go right, go straight, and if the lane is a turning bay. This information would be available to vehicles with access to detailed road maps, but is difficult to extract from raw images. We hypothesize that these features will have a significant impact on the actions a car takes in almost every situation, which is why they are included despite the relative expense of adding them to an actual driving system.

TABLE I: Discretization bin edges for continuous variables.

Variable	Bin Edges	Unit
distance	$0, 5, 25, 100, 500$	ft
speed	$0, 0.5, 20, 40, 60$	ft/s
acceleration	$-20, -5, -0.5, 0.5, 5, 20$	ft/s^2

V. Models

This section details the intent classification models whose performance is compared in Section VI. Models were trained via maximum likelihood estimation unless otherwise noted. All neural networks were trained in Tensorflow[15]. Both support vector machines and naive Bayes classifiers were tested but omitted from the final results due to their poor performance.

A. Marginal Baseline (MA)

A marginal distribution over the predicted actions provides a baseline for subsequent approaches. It is defined purely by the frequency of each action taken in the training dataset.

B. Conditional Probability Table (CPT)

A conditional probability table represents a conditional probability distribution between discrete features and a discrete target variable. They are commonly used in Bayesian networks and their theory is well-established [16].

The parameters for a CPT with a given set of features can be obtained from frequency statistics extracted from a training dataset. Training a CPT model requires choosing the incorporated features. Greedy hill climbing was used along with the K2 parameter prior, which assigns a baseline uniform distribution over the CPT statistics [16].

Use of the CPT model required continuous features to be discretized. The features were discretized by hand, using bin edges designed to balance a low number of bins, consistent distribution, and bin importance in order to ensure different bins represent distinct situations. Discretization bin edges for each variable are shown in Table I. Orientation was discretized into 7 uniform width bins.

C. Multilayer Perceptron (MLP)

Two models based on neural networks are included. The first, the multilayer perceptron, is a feedforward neural network which passes the input features through a series of affine transformations separated by rectified linear units [17]. The final layer has one entry for each action, and a softmax layer produces the unnormalized parameters to a categorical distribution over the intersection action.

The MLP has two hidden layers, 128 units each. Models were trained with ADAGRAD [18] using a batch size of 1024. The remaining parameters in the TensorFlow DNN classifier were left at default.

D. Recurrent Neural Network (RNN)

Recurrent neural networks maintain a hidden state that changes over time, allowing them to learn to identify and remember important events that affect future predictions. Driving is a sequential task, and a recurrent model may benefit from information gathered over multiple timesteps. Long short term memory units (LSTMs) are one popular type of recurrent neural network that can retain information over long periods [19], and have been successfully used in driver behavior modeling [20].

Three LSTMs of varying complexity are evaluated: the *R*128×2, *R*128×3, and *R*256×2 models, where $Rn \times m$ indicates a model with m hidden layers of n units each. Model size is varied to show the effect of layer and unit count.

All three LSTM models were trained with the same hyperparameters. The forget bias was 1, the weight initialization scale was 0.05, the learning rate and decay were 1 and 0.8, and the maximum gradient norm was 5. All LSTMs were trained with a batch size of 10 and with traces of length 20.

VI. Experiments

The following experiments evaluate both the models and the features used to train them. The models are evaluated over all intersections and all features in Section VI-A, generalization to new intersections is investigated in Section VI-B, and prediction performance as a function of the distance from the intersection is evaluated in Section VI-C. Feature selection is used to determine feature importance in Section VI-D.

A. Overall Model Performance

All models were trained on the full feature set using 9-fold cross validation with folds split across intersections. The mean cross-validated accuracy and log likelihood was extracted for each model. The accuracy measures the fraction of correct predictions on withheld data, and is an indicator of absolute predictive performance. The likelihood measures the probability assigned to the correct prediction, and thus provides a measure of confidence. The mean for each metric over all folds is reported in Fig. 3. The results presented are micro-averages: each prediction is weighted equally. Micro-averages contrast with macro-averages, where each cross-validation fold would be weighted equally.

The deep learning models exhibit the highest overall mean accuracy and the lowest overall variation in accuracy. The recurrent neural networks outperform the MLP with higher overall mean accuracy. The remaining models have much lower minimum accuracies, suggesting less ability to generalize to certain intersections. The best model appears to be *R*128×2, which has the highest mean values and lowest variances among RNNs.

The recurrent neural networks tend to exhibit the best overall mean likelihood, while having typical amounts of variation in likelihood. This performance stands in stark contrast to the MLP, which has the worst likelihood in both average and variation, despite outperforming the baselines in terms of overall accuracy. The poor performance suggests that the MLP assigns very low likelihoods to a few correct actions. The remaining models have similar likelihoods as the RNNs, although the best model remains the *R*128×2, which

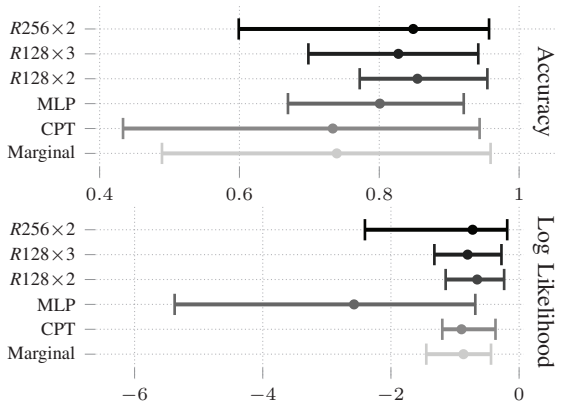

Fig. 3: The mean cross-validated accuracy and log-likelihood for each candidate model trained on all features. Error bars indicate maximum and minimum observed scores among all 9 folds.

		Actual			
		straight	left	right	
Predicted	straight	28 153	2653	2024	
	left	312	5702	0	
	right	621	0	82	
					overall
	accuracy	0.858	0.925	0.933	0.905
	precision	0.968	0.682	0.039	0.563
	F_1 score	0.909	0.794	0.058	0.587

Fig. 4: The confusion matrix for $R128\times2$ model over all intersections for predictions within 6.1 m (20 ft) of the intersection. Large imbalances among the classes are an inherent difficulty.

has the highest mean values and lowest variances among all the models for likelihood.

A confusion matrix for the $R128\times2$ model for samples close to the intersection is given in Fig. 4. The class imbalances is clearly shown, with passing straight through the intersection occurring 74% of the time, left 21%, and right only about 5% of the time. This imbalance leads to relatively high accuracy scores despite poor precision in predicting right turns, which is present in all models.

B. Generalization between Intersections

The results from the previous section can be inspected on a per-intersection basis, and are shown in Table II and Table III. The primary trend is the variability in performance according to the withheld intersection. In particular, intersection 3 has the worst performance. It is the busiest intersection, accounting for about a quarter of all of the data points. The simplest RNN, at 128 units and two layers, almost doubles the performance accuracy on this intersection. Furthermore, fewer than half of the cars continue straight through intersection 3, whereas other intersections are much more dominated by the action of continuing straight. The next most complicated intersection, intersection 6, shows a similar increase in accuracy for the recurrent models over the baselines. This pattern demonstrates the modeling performance of LSTMs.

Secondly, the results are relatively consistent across three and four-way intersections. Intersections 1 and 8 are three-way, but the performance on them is very similar to the other intersections, strengthening the claim that this approach is generalizable over intersection layouts.

Finally, the complicated recurrent models tended to be best more often in particular folds, but also had some of the lowest results in certain folds. The increased model complexity combined with the variable performance suggests some degree of overfitting. By using fewer units and layers, the $R128\times2$ model maintained the advantages of learning long-term dependencies while limiting the model complexity and keeping it simple. This hypothesis is further supported by the outcomes described in Section VI-D, where $R128\times2$ utilized fewer features than the more complicated models, allowing it to learn more general trends.

C. Distance Experiments

A third experiment investigates the relationship between model performance and distance to the intersection. The results show that the prediction accuracy is relatively constant over the distance, with no major changes in prediction accuracy. This finding is consistent with major trends being largely determined by the lane one is in. Nevertheless, the prediction variance increases with distance from the intersection.

Figure 5 shows the general trends for model accuracy, with variance included. It shows how the baselines have much higher variance than the neural networks, and how the accuracy decreases slightly further away from the intersection.

Figure 6 shows the performance versus distance for intersection 3. Intersection 3 is the busiest intersection and yielded the worst performance when withheld. It reinforces the trends seen in Fig. 5 and highlights the performance discrepancy between the baselines and the neural networks.

D. Feature Selection

The features selected by the models provide insight into what information is important to making good predictions. Models were trained using all available features, but in this post-test analysis forward feature search determines which features contributed to increases in accuracy for the model. In forward feature search, a feature set is built up by incrementally adding the feature which leads to the greatest improvement in accuracy. There are some caveats to this method: increases in accuracy may not lead to better overall performance, and the greedy method may miss the best set of features.

The most important features were the rule-based features, based on the number of times they were selected by all models. The ego vehicle headway also appeared in the top five features, along with the distance to the intersection. A variety of other features were also significant, notably the historical headway, and a handful of traffic features.

The best performing model, $R128\times2$, only used three features when tested on the most complicated intersection:

TABLE II: Accuracies for over all folds with specified held-out intersections.

Model	1	2	3	4	5	6	7	8	9
Marginal	0.914	0.956	0.480	**0.946**	0.868	0.738	**0.929**	0.804	0.837
CPT	0.923	0.509	0.438	0.945	0.868	**0.824**	0.819	0.856	0.837
MLP	0.920	0.910	0.669	0.871	0.848	0.793	0.743	0.915	0.714
$R128\times2$	0.925	0.955	0.808	0.897	0.868	0.772	0.826	0.888	0.847
$R128\times3$	0.925	0.924	0.699	0.940	**0.869**	0.802	0.849	**0.942**	**0.865**
$R256\times2$	**0.926**	**0.957**	**0.834**	0.940	0.833	0.778	0.599	0.853	0.861

TABLE III: Log Likelihood over all folds with specified held-out intersections.

Model	1	2	3	4	5	6	7	8	9
Marginal	−0.563	−0.439	−1.446	−0.408	−0.501	−0.815	−0.487	−0.611	−1.787
CPT	−0.722	−0.668	−1.192	−1.101	−0.833	**−0.441**	−0.622	−0.367	**−0.483**
MLP	−0.906	−0.942	−5.375	−0.698	−1.922	−1.620	−2.114	−0.684	−1.754
$R128\times2$	**−0.346**	−0.234	**−0.904**	−0.381	−0.538	−0.902	−1.146	−0.335	−0.631
$R128\times3$	−0.429	−0.307	−1.322	**−0.275**	−0.746	−0.810	−1.105	**−0.331**	−0.624
$R256\times2$	−0.380	**−0.186**	−0.965	−0.340	−0.576	−0.750	−2.406	−0.363	−0.686

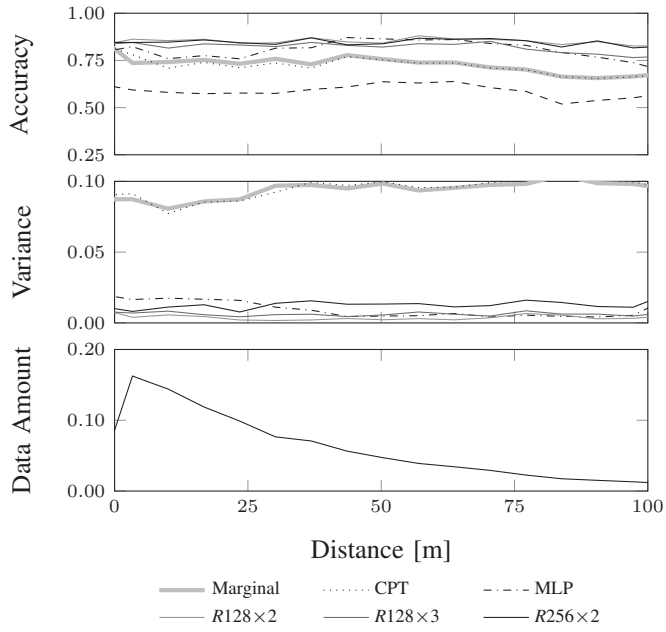

Fig. 5: Mean prediction accuracy and variance versus distance to the intersection. Performance is based on distance averaged over all intersections.

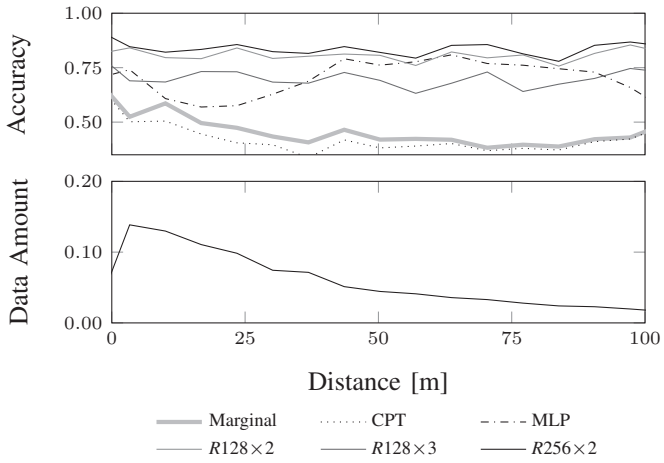

Fig. 6: Prediction accuracy versus distance to intersection 3, the busiest intersection yielding the worst performance when withheld.

intersection 3. It used the rules for continuing straight, turning right, and whether or not the lane was a turning bay. The prominence of these features indicates that they are extremely import for future work, supporting our initial hypothesis that the rules based features would be critical for achieving exceptional performance.

Another result that arose from examining the features selected by the models is that the MLP was able to capitalize on many more features than the other models. Most models utilized less than 10 features, but the MLP was using approximately 20. However, this did not lead to better scores, most likely due to overfitting on the training set. A likely cause for why the LSTMs did not use so many features is that their recurrent nature allowed for many of the historical features to be redundant and thus not contribute to better performance.

VII. Conclusion and Future Work

LSTMs are shown to outperform other models in the supervised classification task of predicting the action a human will take at an upcoming intersection. In the most interesting case, that of a high-throughput intersection, the LSTMs excel relative to the other models. Averaged across all intersections, the best model achieves over 85% accuracy, whereas prior work by Tang, Khokhar, and Gupta on a restricted class of intersections achieved approximately 90% accuracy [2].

Future work can extend this approach to include other features, including traffic signal information, road markings, and vehicle types. This work can be adopted to intention prediction in other contexts, including predicting lane changes or exiting on highways. A primary difficulty during model

training was the heavily unbalanced class labels. Future work can use simulated data to balance the classes or use weighting methods to synthetically balance the class ratios. As more data becomes available it will be possible to test these models on a more diverse range of intersections, including smaller roads where the road rules are less dominant features.

REFERENCES

[1] S. Lefévre, C. Laugier, and J. Ibañez-Guzmán, "Evaluating risk at road intersections by detecting conflicting intentions", in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2012.

[2] B. Tang, S. Khokhar, and R. Gupta, "Turn prediction at generalized intersections", in *IEEE Intelligent Vehicles Symposium (IV)*, 2015.

[3] H. Kretzschmar, M. Spies, C. Sprunk, and W. Burgard, "Socially compliant mobile robot navigation via inverse reinforcement learning", *The International Journal of Robotics Research*, 2016.

[4] J. Heine, M. Sylla, I. Langer, T. Schramm, B. Abendroth, and R. Bruder, "Algorithm for driver intention detection with fuzzy logic and edit distance", in *IEEE International Conference on Intelligent Transportation Systems (ITSC)*, 2015.

[5] Q. Tran and J. Firl, "Online maneuver recognition and multimodal trajectory prediction for intersection assistance using non-parametric regression", in *IEEE Intelligent Vehicles Symposium (IV)*, 2014.

[6] T. Streubel and K. H. Hoffmann, "Prediction of driver intended path at intersections", in *IEEE Intelligent Vehicles Symposium (IV)*, 2014.

[7] C. Laugier, I. E. Paromtchik, M. Perrollaz, M. Yong, J. D. Yoder, C. Tay, K. Mekhnacha, and A. Négre, "Probabilistic analysis of dynamic scenes and collision risks assessment to improve driving safety", *IEEE Intelligent Transportation Systems Magazine*, vol. 3, no. 4, 2011.

[8] T. Gindele, S. Brechtel, and R. Dillmann, "Learning context sensitive behavior models from observations for predicting traffic situations", in *IEEE International Conference on Intelligent Transportation Systems (ITSC)*, 2013.

[9] G. S. Aoude, B. D. Luders, K. K. H. Lee, D. S. Levine, and J. P. How, "Threat assessment design for driver assistance system at intersections", in *IEEE International Conference on Intelligent Transportation Systems (ITSC)*, 2010.

[10] A. F. Foka and P. E. Trahanias, "Probabilistic autonomous robot navigation in dynamic environments with human motion prediction", *International Jounral of Social Robotics*, vol. 2, no. 1, 2010.

[11] T. N. Hoang and K. H. Low, "Interactive POMDP lite: Towards practical planning to predict and exploit intentions for interacting with self-interested agents", in *International Joint Conference on Artificial Intelligence (IJCAI)*, ser. IJCAI '13, 2013.

[12] V. Sezer, T. Bandyopadhyay, D. Rus, E. Frazzoli, and D. Hsu, "Towards autonomous navigation of unsignalized intersections under uncertainty of human driver intent", in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2015.

[13] FHWA, *NGSIM program Lankershim boulevard data*, version 1, 2005.

[14] FHWA. (2006). NGSIM program peachtree street data. version 1.

[15] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, *TensorFlow: Large-scale machine learning on heterogeneous systems*, Software available from tensorflow.org, 2015.

[16] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*. 2009.

[17] R. H. Hahnloser, R. Sarpeshkar, M. A. Mahowald, R. J. Douglas, and H. S. Seung, "Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit", *Nature*, vol. 405, no. 6789, 2000.

[18] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization", *Machine Learning Research*, 2011.

[19] S. Hochreiter and J. Schmidhuber, "Long short-term memory", *Neural Computation*, vol. 9, no. 8, 1997.

[20] J. Morton, T. A. Wheeler, and M. J. Kochenderfer, "Analysis of recurrent neural networks for probabilistic modeling of driver behavior", *IEEE Transactions on Intelligent Transportation Systems*, vol. PP, no. 99, 2016.