Understanding the Past to Predict the Future: Multipolicy Decision-Making using Changepoints for Autonomous Driving

Enric Galceran, Alexander G. Cunningham, Ryan M. Eustice, and Edwin Olson

***Abstract*— We present an integrated behavioral inference and decision-making approach that models vehicle behavior for both our vehicle and nearby vehicles as a discrete set of closed-loop policies that react to the actions of other agents. Each policy captures a distinct high-level behavior and intention, such as driving along a lane or turning left or right at an intersection. We first employ Bayesian changepoint detection on the history of observed states of nearby cars to estimate the distribution over policy assignments for each car. We then sample from these distributions to obtain high-likelihood actions for each participating vehicle. Through closed-loop forward simulation of these samples, we evaluate the outcomes of the interaction of our vehicle with other participants. Our vehicle executes the policy that maximizes the expected reward over these samples. Thus, our system makes decisions incorporating the coupled interactions between cars in a tractable, online manner. We evaluate our approach using real-world traffic tracking data collected on our autonomous vehicle platform, and present decision-making results in simulation.**

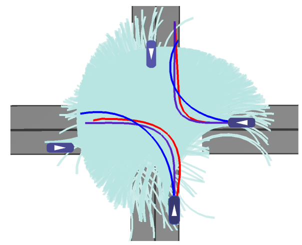


Fig. 1. Our multipolicy approach samples from the likely coupled interactions between vehicles. In this simulation instance we evaluate the outcomes of the possible intentions of other cars, where the bottom and right cars proceed through the intersection, while the other two cars yield. This experiment shows that our multipolicy sampling strategy generates high-likelihood samples over the coupled interactions of vehicles, and that is orders of magnitude faster than uninformed sampling strategies used in other algorithms. Legend: human-driven trajectories (red); rollouts obtained via our multipolicy sampling strategy (purple); high-likelihood trajectories obtained by an uninformed sampling strategy (dark blue); samples by the uninformed strategy before finding a high-likelihood sample (light blue).

I. Introduction

Decision-making within autonomous driving is hard due to uncertainty on the continuous state of participating vehicles and, especially, over their potential discrete intentions (such as turning at an intersection or changing lanes).

Previous approaches have employed hand-tuned heuristics [1–3] and numerical optimization [4–6], but these methods fail to capture the coupled dynamic effects of interacting traffic agents. Partially observable Markov decision process (POMDP) solvers [7–9] offer a theoretically-grounded framework to capture these interactions, but have difficulty scaling up to real-world scenarios. In addition, current approaches for anticipating the future intentions of other traffic participants [10–13] either consider only the current state of the target vehicle, ignoring the history of its past actions, or rather require expensive collection of training data.

We present an integrated behavioral anticipation and decision-making system that models vehicle behavior for both our vehicle and nearby vehicles as a discrete set of policies $\pi \in \Pi$. Each policy captures a different high-level behavior, such as following a lane or changing lanes, turning left or right at an intersection. In addition, we parameterize each policy $\pi \in \Pi$ by a parameter vector θ. For example, for a lane following policy, θ can capture the "driving style" of the policy by regulating its acceleration profile to be more or less aggressive.

We leverage Bayesian changepoint detection to estimate which policy each vehicle was executing at each point in its observed history, and infer the probability distribution over potential intentions of each vehicle. Furthermore, we propose a statistical test based on changepoint detection to identify anomalous behavior of other vehicles, such as swerving off of lanes or driving in prohibited directions.

Using this distribution, we sample over permutations of other vehicle policies and the policies available for our car, and use forward-simulation of these sampled intentions to evaluate their outcomes via a user-defined reward function. Our vehicle finally selects the policy to execute that maximize the expected reward function of the samples. Thus, our system makes decisions based on closed-loop interactions between cars in a principled manner, while providing the convenience and tractability of a discrete set of engineered policies that capture likely driving behavior.

We evaluate our behavioral prediction system using data from a real-world autonomous vehicle, and present decision-making results in simulation involving challenging traffic situations in highway scenarios.

This work extends our previous work in [14], where we evaluated a only single sample for each policy available to our vehicle, through the addition of behavior prediction via changepoint detection, enabling sampling over a broader set of outcomes.

II. Approach

Our approach first estimates the probability distribution over policy assignments of other vehicles given sensor observations. Next, we sample from these distributions to obtain

This work was supported in part by Ford Motor Co. via the Ford-UM Alliance (award N015392) and in part by DARPA (award D13AP00059).

The authors are with the University of Michigan, Ann Arbor, MI. {egalcera,alexgc,eustice,ebolson}@umich.edu

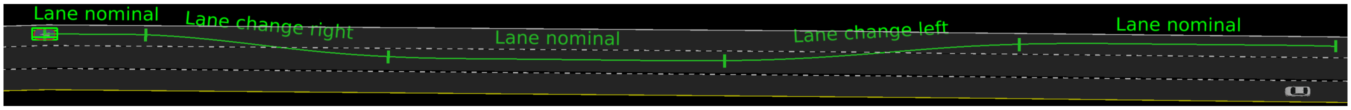


Fig. 2. Policy changepoint detection on a simulated passing maneuver on a highway. Our vehicle (far right) tracks the behavior of another traffic agent (far left) as it navigates from right to left. Using the tracked vehicle's history of past observations (green curve), we are able to infer which policies are most likely to have generated the maneuvers of the tracked vehicle.

joint policy assignments for our car and other cars, which we simulate forward to find the optimal policy for our vehicle to execute. The process iterates in a receding horizon fashion.

Our behavioral anticipation method segments the history of observed states of each vehicle, where each segment is associated with the policy most likely to have generated the observations in the segment. We obtain this segmentation using Bayesian changepoint detection, which infers the points in the history of observations where the underlying policy generating the observations changes. Thereby, we can compute the likelihood of each available policy for the target car given the observations in the most recent segment, capturing the distribution $p(\pi_t^v | x_t, \mathbf{z_{0:t}})$ over potential policies for the target car given the current state of all vehicles x_t and a time-series of sensor observations $\mathbf{z_{0:t}}$. Further, full history segmentation allows us to detect anomalous behavior that is not explained by the set of policies in our system. The changepoint detection procedure is illustrated by the simulation shown in Fig. 2.

We adopt the recently proposed CHAMP algorithm [15] to segment the target car's history of observed states. CHAMP has been previously applied to detection of changes in articulation models for manipulation tasks. Here, we show that policies can be used as models to explain the observed behavior of other agents in a changepoint detection framework. In contrast to prior work [6, 13, 16], this allows us to leverage past observations for behavior anticipation online.

Given the set of available policies Π and a time series of the observed states of a given vehicle $\mathbf{z_{1:n}} = (z_1, z_2, \ldots, z_n)$, we apply CHAMP to infer the maximum *a posteriori* (MAP) set of times $\tau_1, \tau_2, \ldots, \tau_m$, at which changepoints between policies have occurred, yielding $m+1$ segments. Thus, the i^{th} segment consists of observations $\mathbf{z}_{\tau_\mathbf{i}+\mathbf{1}:\tau_{\mathbf{i}+\mathbf{1}}}$ and has an associated policy $\pi_i \in \Pi$ with parameters θ_i. Further details on this changepoint detection method are given in [15].

Considering the $(m+1)^{\text{th}}$ segment (the most recent), consisting of observations $\mathbf{z}_{\tau_\mathbf{m}+\mathbf{1}:\mathbf{n}}$, the likelihood and parameters of each latent policy $\pi \in \Pi$ for the target vehicle given the present segment can be computed by solving the following maximum likelihood estimation (MLE) problem:

$$\forall \pi \in \Pi, \qquad \mathcal{L}(\pi) = \operatorname*{argmax}_{\theta} \log p(\mathbf{z}_{\tau_\mathbf{m}+\mathbf{1}:\mathbf{n}} | \pi, \theta). \tag{1}$$

The policy likelihoods obtained via (1) capture the probability distribution over the possible policies that the observed vehicle might be executing. As outlined in Alg. 1, we can therefore compute the approximated posterior over likely future vehicle interactions by sampling from this distribution for each vehicle. We can then simulate forward in time throughout the decision horizon, being able to reason about the potential actions of other vehicles.

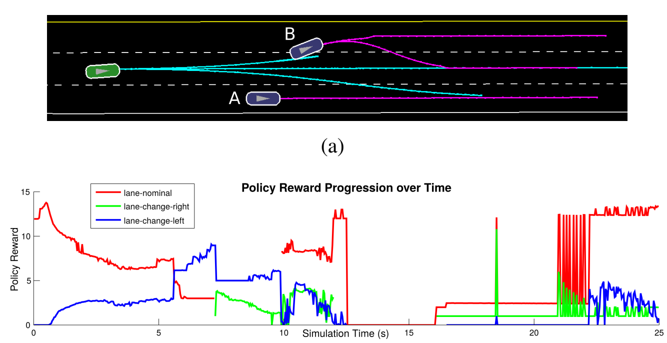
Algorithm 1: Policy selection procedure.

Input:
- Current MAP estimate of the state, x_0.
- Set of available policies Π.
- Prob. distributions over other vehicles' policies.

```
Get a set of samples S via (1), where each sample
  s ∈ S assigns a policy to each nearby vehicle.
C ← ∅ // Rewards for each rollout
foreach π ∈ Π do // Our car's policies
    foreach s ∈ S do // Other cars' policies
        Ψ ← SimulateForward(x0, π, s)
        C ← C ∪ {(π, ComputeReward(Ψ))}
return π* ← SelectBest(C)
```

III. Results

Our results, involving 67 dynamic object trajectories recorded in an urban area using our autonomous vehicle platform, show that our approach identifies other agents' policies with over 85% accuracy and precision after only 50% of the trajectory has been completed. The closed-loop nature of our policies allows us to maintain safety at all times regardless of anticipation performance, while correct anticipation provides smoother driving behavior than without a model of vehicle interactions. Sampling experiments (Fig. 1) show that our multipolicy strategy achieves high-likelihood samples orders of magnitude faster than uninformed strategies, allowing us to perform decision-making online (Fig. 3).



(a)

(b)

Fig. 3. (a) Results of a simulated multi-car interaction scenario, where our car (green) approaches the slower vehicles A and B from behind. Vehicle B starts by executing a lane change from the center to left lane, which it is just completing by the time shown, while A remains in the right lane. Cyan lines show the simulated rollouts for our car; magenta lines show the simulated rollouts for the other vehicles. (b) Evaluation of the reward functions for each of the three policies involved over the course of the simulated scenario.

Acknowledgments

The authors are sincerely grateful to Patrick Carmody from Ford Motor Company for his help in collecting the traffic tracking data used in this work and to Ryan Wolcott for his helpful comments.

References

[1] M. Montemerlo *et al.*, "Junior: The Stanford entry in the Urban Challenge," *J. Field Robot.*, vol. 25, no. 9, pp. 569–597, 2008.

[2] I. Miller *et al.*, "Team Cornell's Skynet: Robust perception and planning in an urban environment," *J. Field Robot.*, vol. 25, no. 8, pp. 493–527, 2008.

[3] C. Urmson *et al.*, "Autonomous driving in urban environments: Boss and the Urban Challenge," *J. Field Robot.*, vol. 25, no. 8, pp. 425–466, 2008.

[4] D. Ferguson, T. M. Howard, and M. Likhachev, "Motion planning in urban environments," *J. Field Robot.*, vol. 25, no. 11-12, pp. 939–960, 2008.

[5] W. Xu, J. Wei, J. Dolan, H. Zhao, and H. Zha, "A real-time motion planner with trajectory optimization for autonomous vehicles," in *Proc. IEEE Int. Conf. Robot. and Automation*, May 2012, pp. 2061–2067.

[6] J. Hardy and M. Campbell, "Contingency planning over probabilistic obstacle predictions for autonomous road vehicles," *IEEE Trans. Robot.*, vol. 29, no. 4, pp. 913–929, 2013.

[7] H. Kurniawati, D. Hsu, and W. Lee, "SARSOP: Efficient point-based POMDP planning by approximating optimally reachable belief spaces," in *Proc. Robot.: Sci. & Syst. Conf.*, 2008.

[8] D. Silver and J. Veness, "Monte-carlo planning in large POMDPs," in *Advances in Neural Information Processing Systems 23*, J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel, and A. Culotta, Eds. Curran Associates, Inc., 2010, pp. 2164–2172.

[9] H. Bai, D. Hsu, and W. S. Lee, "Integrated perception and planning in the continuous space: A POMDP approach," *Int. J. Robot. Res.*, vol. 33, no. 9, pp. 1288–1302, 2014.

[10] K. Kim, D. Lee, and I. Essa, "Gaussian process regression flow for analysis of motion trajectories," in *Proc. IEEE Int. Conf. Comput. Vis.*, Nov. 2011, pp. 1164–1171.

[11] J. Joseph, F. Doshi-Velez, A. S. Huang, and N. Roy, "A Bayesian nonparametric approach to modeling motion patterns," *Auton. Robot.*, vol. 31, no. 4, pp. 383–400, 2011.

[12] G. S. Aoude, B. D. Luders, J. M. Joseph, N. Roy, and J. P. How, "Probabilistically safe motion planning to avoid dynamic obstacles with uncertain motion patterns," *Auton. Robot.*, vol. 35, no. 1, pp. 51–76, 2013.

[13] F. Havlak and M. Campbell, "Discrete and continuous, probabilistic anticipation for autonomous robots in urban environments," *IEEE Trans. Robot.*, vol. 30, no. 2, pp. 461–474, 2014.

[14] A. G. Cunningham, E. Galceran, R. M. Eustice, and E. Olson, "MPDM: Multipolicy decision-making in dynamic, uncertain environments for autonomous driving," in *Proc. IEEE Int. Conf. Robot. and Automation*, Seattle, WA, USA, 2015, To Appear.

[15] S. Niekum, S. Osentoski, C. G. Atkeson, and A. G. Barto, "CHAMP: Changepoint detection using approximate model parameters," in *Technical report CMU-RI-TR-14-10, Robotics Institute, Carnegie Mellon University*, 2014.

[16] D. Ferguson, M. Darms, C. Urmson, and S. Kolski, "Detection, prediction, and avoidance of dynamic obstacles in urban environments," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2008, pp. 1149–1154.