Paper: *Maximum Entropy Inverse Reinforcement Learning*

Nirajan Koirala

CSC 9010_001

Summary

Ziebart et al. develop a probabilistic method based on the principle of maximum entropy which provides a well-defined, globally normalized distribution over decision sequences, while providing similar performances as existing methods. This method is particularly developed for solving the problem of modeling the real-world routing preferences of drivers. In comparison to many imitation learning methods, their probabilistic model of purposeful behavior integrates seamlessly with other probabilistic methods including hidden variable techniques. The underpinnings of the imitation learning is are provided in the background section. Contrary to the previous work that reasons about policies, they consider a distribution over the entire class of possible behaviors and it corresponds to paths of variable length for deterministic MDPs (Markov Decision Processes).

For non-deterministic path distributions, they use the maximum entropy distribution of paths conditioned on the transition distribution T and constrained to match feature expectations. They measure the empirical, sample-based expectations of the feature values and not the true values of the agent to be imitated. For the experiments, online exponentiated gradient descent algorithm is used which is both very efficient and induces an l_1-type regularizing effect on the coefficients. The algorithm approximates the state frequencies for the infinite time horizon using large fixed time horizon and recursively "backs up" from each possible terminal state and computes the probability mass associated with each branch. Road networks are chosen for modeling and destination is represented within the MDP. After modeling, they evaluate their model on different criterias like most likely path estimate, predicted path etc. and it is able to perform better than other models included in the evaluation.

In the applications section they describe how their approach can be applied to a wide range of scenarios involving driver prediction. Hence, their novel approach to inverse reinforcement and imitation learning cleanly resolves ambiguities in previous approaches and provides a convex and computationally efficient procedure for optimization while ensuring good performance.

Strengths

- The presented method can have many practical applications in Autonomous Driving.
- The IRL section discussed in the paper is very detailed.

Weaknesses

- I think the application section could have been made longer by presenting more practical applications of the method they described.

Points of Confusion

- How does Inverse Reinforcement Learning differ from Reinforcement Learning?
- Section dealing with Non-Deterministic Path Distributions was confusing.

Discussion Questions

- Is there any study done on the applicability of these methods in the field of Autonomous Driving?
- What are some other fields in which principles of maximum entropy used?
- A discussion on the algorithm computing the expected state occupancy frequency.