Review-20

Model-Based Inverse Reinforcement Learning from Visual Demonstrations

Lack of scalability to real time robot manipulation tasks hinders Inverse Reinforcement Learning (IRL) under unknown dynamics. To this end, the work proposes a gradient-based IRL framework which learns cost functions from a pre-trained visual dynamics model trained on visual demonstrations. The learned cost functions are then employed to reproduce manipulation behavior using visual Model Prediction Control (MPC). Experiments carried out on a real robot learning from human demonstrations validates the suitabilty of the proposed framework.

The framework takes a step towards IRL by learning cost functions which are tied to the inner and outer loops of the optimization problem. Unlike modern-day meta-learning methods which simply optimize the outer loop consisting of an inner loop as learning problem, the framework connects the two objectives together using optimization over action sequences. Firstly, a keypoint detector identifies keypoint locations with maximum variability in the visual input. The detector then predicts these points in latent space using self-supervised learning. Following identification of key points, the dynamics model is trained to predict next state corresponding to latent state and joint state. The dynamics model is then executed during planning stage which comprises of gradient-based visual MPC towards a goal keypoint state. Actions sequences are rolled out in order to minimize the distance between the object and goal configuration. Corresponding to planned trajectories, costs are computed which are minimized using gradient descent. The outer loop optimizes cost parameters as a function of inner loop optimization based on actions. Gradients do not require tracking the inner loop but directly depend on imagined sequences and hence, facilitate stability in bi-level optimization. Cost functions utilized in the optimization process consist of weighting schemes which are varied based on time dependency and choice of kernel.

In comparison to weighted appreticeship learning, the proposed framework demonstrated improved convergence in cost function on placing and reaching tasks. Additionally, the time-dependent weighting scheme is found to be more efficient but utilizes more number of parameters. In the real robot setting, RBF weighted objective depicts improved convergence along X-axis keypoints in comparison to baselines. Finally, the learned cost function yields suitable behaviors on the robot which allow it to place the object successfully and the distance between object and goal location. While, the framework depicts strong convergence properties and is a suitable step in IRL, the training setup highlights some caveats. Time-dependence and RBF weighted schemes do not present significant difference in parameters indicating that the proposed RBF scheme is empirically equivalent to the computationally expensive time-dependent scheme. Secondly, the connection of gradients between loops leads to stability. However, the work does not throw light on when this may fail. For instance, if action trajectories of the model are inaccurate or the goal state is significantly far from the object it may lead to very large or small gradients which is troublesome for the outer loop.

The proposed framework demonstrates suitability of inner and outer loops connected through gradients which aid in learning cost functions. The work presents several novel directions for future research. A more robust keypoint detector could be modeled which would demonstrate invariance to viewpoints. Similarly, mechanism of demonstrations could be improved from different contexts.

With the objective of bridging gaps between practical IRL in model-based setting, the work proposeds a framework for learning objective functions. In contrast to recent literature, the framework links gradients between inner and outer loops using action trajectories. This leads to improved stability and convergence in weighted cost functions which enable learning from visual demonstrations.