READING ASSIGNMENT 2
16785-Integrated intelligence in robotics

[redacted]
Carnegie Mellon University

1. Maximum Entropy Deep Inverse Reinforcement Learning

In this paper, Wulfmeier et.a l proposes learning a non-linear reward function using maximum entropy deep inverse reinforcement learning with a deep convolutional neural network architecture. He claims that neural networks are the obvious choice for the task. They have higher computational efficiency, versatility of architecture and eliminate the need to manually design features. This being said however in the experimental section, fairly simple CNNs have been deployed. Performance could be improved by adding more layers or changing the architecture. Also for the spatial feature learning section, capsule networks can be used. They use neuron layers in a capsule within layers to attain better spacial recognition.

The paper suggests that it is well-suited for life-long learning scenarios in the context of robotics, which inherently provide sufficient amounts of training data. The primary limitation is that it is extremely data intensive. Also the estimation process will be negatively affected in scenarios where dataset trajectories are not the same length. The second drawback is that it computes the forward algorithm for estimating rewards in every iteration. It is time consuming and computationally expensive. Lastly, as stated explicitly in the paper, a model based approach requiring prior knowledge of the environment is assumed.

This framework must be extended to unknown environments and problems with less data. A sampling distribution such as Gaussian can be used or one can use a distribution in the neighbourhood of the demonstration to predict a state transition probability matrix. Also to negate the problem of different trajectory lengths, ,compute an estimator of the expectation of a function under the given distribution. Calculate important weights using estimator and select samples accordingly.

Also as in the case of guided cost learning, its policy could be updated simultaneously with the cost function so that the algorithm better distinguished successful trials from failures even in unknown environments.

2. Guided Cost Learning: Deep Inverse Optimal Control via Policy Optimization

This framework uses a maximum entropy inverse optimal control algorithm that can learn complex, nonlinear cost representations, such as neural networks, and can be applied to high-dimensional systems with unknown dynamics. It works when expert intuition fails to design good features to represent costneural networks. It couples learning the cost with learning the policy for that cost. Consider certain fairly complex tasks where the target is moved. An optimum policy is learned. However, the algorithm

fails to find a global cost function. In such scenarios, the cost function must be discarded and only the learned policy should be retained.

Also it is data inefficient because it requires a set of trajectories to estimate reward and expert demonstrations. Extend the approach to using a vision based technique on image pixels from a series of images or a video of the demonstration. The algorithm should be made to learn a cost function from the images and avoid overfitting. Convolutional neural networks or capsule neural networks can be used. They detect spacial correlations better especially in semi-supervised settings. Also to negate the afore mentioned drawbacks as proposed by Tzeng et. al, develop regularization methods for domain adaptation in computer vision. Develop a network architecture that optimizes domain invariance, facilitates domain transfer and effectively transfers task information between domains via soft labels.

3. Feature-Based Prediction of Trajectories for Socially Compliant Navigation

This paper provides a novel framework for predicting movements of pedestrians using maximum entropy learning based on features. These features supposedly capture the relevant aspects of trajectories and help determine a probability distribution for human behaviour. This enables mobile robots to navigate a shared environment with humans. The features captured in this scenario are fairly limited-time, acceleration, velocity, collision avoidance etc. To improve the approach, more feature that define relationships between pedestrians such as aggressiveness or passiveness of approach can be used. It will the ratio of relative distance to relative speed between pedestrians. The approach used in this paper is fairly outdated.

Also since this is a model based approach, it relies on handcrafted features such as preferred walking speed of pedestrian and it fails to capture human thought in progress. A valid suggestion would be to introduce a variant of recurrent neural networks-social LSTM (long short term memory). Deploy one LSTM for each pedestrian and share information between pedestrians through a social pooling layer.

Another aspect that must not be neglected is predicting behaviours of social interactions such as groups of people moving together or people pausing to acknowledge each other. Social LSTM would be capable of capturing these tendencies. Bidirectional LSTM can also be deployed to store potential destination regions so as to improve prediction accuracy.