Responses to QMDP-net [1]

Matthew Feng

October 19, 2018

1 Important contributions

In your opinion what is an important contribution of this paper? What is a limitation?

This paper is one of the first applications of value-iteration networks to POMDPs, which are considerably more complex than fully observable MDPs. The idea is to view neural networks as differentiable programs rather than function approximators.

The network, once trained, is able to generalize, compared to other networks.

Finally, one limitation is that it uses a very simple approximation algorithm for planning; using other algorithms may be more effective? (AMDP)?

2 Exploration vs. Exploitation

POMDPs model partially observable problems. A POMDP policy has to trade off exploration vs. exploitation. In the case of partially observable grid navigation, what is exploration and what is exploitation?

Exploration would encompass finding more information about its position in the grid-world, as well as figuring out associations between its observations and the floormap by changing views and positions without explicitly moving towards the goal.

Exploitation would be using its belief in order to navigate towards to goal.

3 Optimality

Q_{MDP} is an algorithm for solving POMDPs. Does it give an optimal solution? Why?

Q_{MDP} uses the belief in order to average over the MDP optimal value function. It is an approximate algorithm, which is robust but not optimal. The Q_{MDP} algorithm would be optimal if after following its policy for one time-step, the problem suddenly became fully observable.

4 QMDP-net vs. LSTM

What is the important difference between QMDP-net and an LSTM network? Why does QMDP-net work better for the grid navigation task?

The QMDP-net differs from the LSTM network in its structure, designed to act as an algorithmic prior, rather than a general LSTM network. QMDP-net works better than LSTM because QMDP-net implicitly forces the network to develop a model of the environment and plan for the model, rather than just generate a sequence of moves (as an LSTM does).

5 Backpropagation through time

Recurrent neural networks are often trained with backpropagation through time. What is backpropagation through time? Should we train QMDP-net with backpropagation through time?

References

[1] P. Karkus, D. Hsu, and W. S. Lee. Qmdp-net: Deep learning for planning under partial observability. *CoRR*, abs/1703.06692, 2017.