Active Learning for Reward Modelling

Sam Clarke

September 2019

Abstract

Contents

Appendices

Chapter 1

Introduction

Chapter 2

Background

2.1 Reinforcement Learning

2.1.1 Deep Reinforcement Learning

2.2 Reinforcement Learning without a reward function

For many domains in which we might want to use RL, states of the environment are not inherently associated with rewards. Thus, we have the additional task of specifying a reward function, which maps states to rewards. Argue the case that directly specifying a reward function is hard, in order to motivate the next section: Reward Learning

2.2.1 Reward Learning

In Standard RL

In the Deep Case

2.3 Uncertainty in Deep Learning

Standard deep learning models output point estimates. For example, a model trained to classify pictures of dogs according to their breed takes a picture of a dog and outputs its predicted breed. However, what will the model do if it is given a picture

of a cat? [2].

We probably want the model to be able to recognise that this is an out of distribution example, and request more training data, or simply say that it doesn't know the answer. However, since standard deep learning models output only point estimates, the model will just go ahead and classify the cat as some breed of dog, just as confidently as any other input.

Thus, the field of Bayesian Deep Learning aims to equip neural networks with the ability to output a point estimate along with its uncertainty in that estimate. Historically, many of the attempts to do so were not very practical. For example, one algorithm, Bayes by Backprop, requires doubling the number of model parameters, making training more computationally expensive, and is very sensitive to hyperparameter tuning. However, recent techniques allow almost any network trained with a stochastic regularisation technique, such as dropout, to, given an input, obtain a predictive mean and variance (uncertainty), without any complicated augmentation to the network [2, p. 15].

2.4 Active Learning

2.4.1 Applying Active Learning to RL without a reward function

2.5 Relation to Material Studied on the MSc Course

Chapter 3

Method

Chapter 4

Experiments

Chapter 5

Results

Chapter 6

Conclusions

6.1 Summary

6.2 Evaluation

6.3 Future Work

References

[1] Paul Christiano, Jan Leike, Tom B. Brown, Miljan Martic, Shane Legg, and Dario Amodei. Deep reinforcement learning from human preferences. 2017.

[2] Yarin Gal. Uncertainty in Deep Learning. *Phd Thesis*, 1(1):1–11, 2017.

Appendices

Appendix A

Some Appendix Material