THESIS PLAN: “SCALABLE CONTINUOUS RL WITH KULLBACK-LEIBLER POLICY CHAIN”

UNDERGRADUATE FINAL YEAR PROJECT

CHRIS DOYLE

DEADLINES:

* 25th January – Introduction/Background
* 25th March – [Entire Thesis] & Demo to be completed
* 25th March – Demo Presentation
* 12th April – Final Deadline for Thesis

BY 25th JANUARY:

ABSTRACT (OUTLINE PROBLEM & SOLUTION):

Define lifelong learning.

Define lifelong reinforcement learning.

Existing issues being solved (scalability, catastrophic forgetting).

Introduce idea of Kullback-Leibler Policy Chain (i.e. 2 parts, KL & RL).

INTRODUCTION

Reinforcement Learning -> Lifelong Learning.

Define catastrophic forgetting.

What the worst case scenario etc. is.

BACKGROUND

Define Reinforcement Learning.

Define Policy Gradient Methods.

Define Kullback-Leibler Divergence & how it is being used in this project.

* Compare policies in adjacent figures to show effect

BY 25th MARCH

THEORY OF APPROACH

Explain the concept of Kullback-Leibler policy chain in a mathematical sense.

Derive & explain conceptualised cost function.

PRACTICAL EXPERIMENT

Create GridWorld environment & ability to change reward etc.

Create RL agent to learn GW.

Add functionality to record and turn to GIF.

Add functionality to change the reward.

Add KLPC functionality.

RESULTS

DEMO: GIF of standard RL agent vs GIF of KLPC agent

CONCLUSIONS

How generalizable is this approach to other environments?

Compare to other approaches.

Shortcomings of this approach.

* Outside boundary smaller % of grid as grid grows
* Technique works best when goals of tasks align similarly
* Worst case scenario with infinite tasks: uniform inside, perfect outside