Thesis Plan and Roadmap.

An Analysis of Active Inference and Reinforcement Learning Paradigms in Large, Partially Observable and Non-Statioanry Environments.

Fraser Paterson

A Thesis proposal, pursuant to the requirements of the Degree: Bachelor of Science (Honours).

THE UNIVERSITY OF WESTERN AUSTRALIA

Supervisor: Dr Tim French
Department of Computer Science and Software Engineering
The University of Western Australia
27 March 2023

Contents

0.1 Research Objectives

Real-world artificial agents face two serious issues. The first is a high degree of uncertinty, whether this be a consequence of noisy observations, partial observability or non-stationary environmental dynamics. Uncertainties of this kind can reduce an otherwise effective agent to an utter failure. Thus, is the issue of affording **robust** agents. The second issue concerns the nature of difficult, real-world environments, in that they are very often high-dimensional and/or continuous. Simple exploration/planning algorithms can quickly be rendered ineffective when situated in a higher-dimesnioanl instantiation. This issue is known as **scaling**.

Reinforcement Learning has enjoyed a great deal of success in the attempt to scale to higher dimensional, continuous, noisy environments - examples. While there is still much to be done on this front, Active Inference has thus far been almost entierly limited to small, discrete environments, see: **AIF-D**, **AIF-Cur-Insight**. Given that Active Inference represents a potentially unifying paradigm - owing to its generality - and that it has no dependence on any ad-hoc scalar reward signal, it is plausible to suppose that Active Infernce might enjoy several theoretical advantages over more "traditional" methods in Reinforcement Learning and Optimal Control, see: **RL-or-AIF**, **Friston2012** and **Optim-Motor**.

Hence, in the course of this proposed thesis, we shall implement various agents - AIF and RL agents - in both the the fully observable and partially observable cases, with and without various sources of uncertainty and in both benchmark "low-diensional" environments and non-trivial "higher-dimensional" environments. The aim will be to asses the relative advantages and disadvantages of both "approach families" as these pertain to their ability to effectively deal with the various kinds of uncertrainty mentioned above, in addition to their ability to scale up into higher-dimensional environments.

0.1.1 Central Research Questions

The principle aim of this research thesis is twofold. The first aim is to investigate the robustness of AIF methods in noisy, uncertain or partially observable environments, relative to Reinforcement Leraning baselines. The second aim concerns the potental for "scaling up" AIF methods to continuous and/or higher-dimensional state-spaces.

Thus are the central questions asked in this Thesis:

- Are AIF agents more robust to noisy observations and non-stationarity than a comparable RL baseline?
- Can AIF be scaled up to higher dimensional Partially Observable envronments?

0.2 Previous Work

Work of this kind has already appeared in the literature, though it is still very much in its infancy. **Markovi-2021** implemented an Active Inference agent for the multi-armed bandit problem, in the stationary and non-stationary case. The AIF agent did not perform as well as a state-of-the-art bayesian UCB algorithm, in the stationary

case. However in the non-stationary case, the AIF agent outperformed the UCB agent. While this implementaion was over a small, discrete space of environmental states, the results plausibly suggest that AIF would be an effective means of optimal inference and control in a higher-dimensional or continuous problem.

An approach that has enjoyed some success as of late involves the implementaion of Free Energy minimisation as a process of message-passing on a Forney-style factor graph. See: **Sim-AIF-Message**, **Cox-2019**, **Reactive-MP** and **Deep-Temp-AIF**. In this framework, the agent's generative model is factorised in such a was as to be a Forney or "Normal" factor graph. Free Energy minimisation is then cast as a process of message passing over this factor graph. Various message passing algorithms exist, such as Belief Propagation and Variational Message Passing. This message passing scheme greatly reduces the number of terms over which it is necessary to sum, when computing the approximate marginal and posterior distributions thereby affording tractible Active Inference algorithms in relatively high-dimensional settings. The natural inclusion of a generative model and the built-in epistemic imperatives - toward the aim of uncertainty reduction - in AIF, make it highly plausable that this method will be better able to deal with non-stationary environments, dynamic constraint changes, noise and other such sources of uncertainty. Indeed, just as much has been shown in **Bandits**, where AIF performd better than a strong Bayesian UCB algorithm in a non-stationary multi-armed bandit problem.

Yet other approaches have atempted to leverage the ability for deep neural networks to parameterise the distributions of interest. See: **Deep-AIF-Ueltzh-2018**, **Deep-Var-Policy-Grad** and **DEEP-AIF-POMDPs**. Of particular intetest with this aproach are **Scaling-AIF** and **Contrastive-AIF**. The former makes use of amortized inference, in the form of neural network function approximators to parameterize the relevant distributions. Free Energy minimisation is then performed with repect to the function approximators. The use of amortized inference affords several advantages. For example, the number of parameters remains constant with respect to the size of the data and inference can be achieved via a single forward pass through the network. The resulting algorithm was able to explore a much greater proportion of the state space in a simple MountainCar environment, as opposed to two Reinforcement Learning, baseline agents. In addition, the agent was able to learn to control the continuous inverted pendulum task with a far greater sample efficency than the baseline agents. Although the approach offered in **Scaling-AIF** is promising, it was restricted in every case to fully observable environments.

Lastly, the approach of **Contrastive-AIF**, implemented a contrastive objectove for their Active Inference agent, which signifigantly reduced the computational burden of learning the generative model and planning future actions. This method performed signifigantly better than the usual, liklihood-based "reconstructive" means of implementing AIF and it was also computationally cheaper to train. Importantly, this method offered a unique way to afford increased model-robustness in the face of environmental distractors.

0.3 Methods and Objectives

The proposed structure of the investigation shall obey the follwoing three-part itinerary.

0.3.1 Benchmark AIF and RL Agents

Initially, our analysis will be limited to the fully observable case. The environments of interest will be the well-known MountainCar and CartPole baseline environments. The aim will be to assess the performance - robustness, learning rate and sample eficency - of the AIF and Reinforcement Learning agents across two dimensions. These two dimensions are as follows:

1. Whether the AIF agent is implemented in OpenAI gym or RxInfer.jl (testing performance diferences in the factor-graphical method)

2. Whether the environment is "uncertain" (characterised by either noisy observations or non-stationarity)

The AIF agent will be instantiated by means of a deep generative model, and the RL baseline agent will be DQN. The stochastic version of each environment will either furnish the agent with observations that have been subject to additive Gaussian white noise, or will be characterised by a degree of non-stationarity. An example of a potential kind of non-statioarity in the CartPole environment might include randomly changing the length of the pole at iteration n, for all remaining episodes - for instance.

Tasks:

1. Implement DQN for Cartpole - Image, based DQN with Convolutional networks.

2. Implement the Bayesian thermostat AIF Agent from the "Minimal AIF Agent" paper.

3. Implement the Bayesian thermosatat in RxInfer.jl

4. add noise to the observations for each agent.

5. Implement Deep AIF for CartPole in OpenAI gym.

Q-Learning is an off-policy, model free, bootstrapped method. We use one policy to explore the state-action space and another policy to update the action-value estimates.

Research questions at this level shall include:

- What is the sample efficency of the two approach families?

- What is the complexity of the two approach families?

- Does the Factorised generative model appreciably effect the performance of the AIF agent as against the baseline Rl agent?

- Is one approach family - AIF or RL - ultimately more robust to noise or non-stationarity?

0.3.2 Deep AIF - The Partially Observable Case

This portion of the study will take up the investigation offered in **Scaling-AIF** whereby we shall attempt to implement their Deep AIF agent, only now in the partially observable case. We shall attempt to implement a Deep AIF Pong plaing agent, from the Atari library in OpenAI gym. Pong is technically a partialy observable environment, since a single frame does not afford any information about the velocity of the ball. Again, we shall assess the performance of the Deep AIF as agianst a DQN baseline, in the stationary and non-stationary POMDP.

Research questions at this latter level will echo those of the latter:

- What is the sample efficency of the two approach families? has it appreciably changed?
- What is the complexity of the two approach families?
- Is one approach family - AIF or RL - ultimately more robust to noise or non-stationarity?

0.3.3 Deep Contrastive AIF With Partial Observability

Time permitting, and assuming that the latter two tasks are achieved with dome degree of satisfacation, the analysis will be extended finally to a much higher-dimensional POMDP such as Asteroids. The Deep AIF agent will be extended to use the contrastive learning approach developed in **Contrastive-AIF**. This approach signfigantly reduces the computational complexity of the model and hence, appears to be a promising avenue of investigation in the extension of Deep AIF methods to high-dimensional, partially observable models.

0.4 Software and Hardware Requirements

Most of the simulations will be instantiated in Python and associated libraries - such as OpenAI gym. RxInfer.jl will be used to implement the factor-graph based AIF agents in the first portion of the study.

The Deep Versions of each agent can be more effectively trained by means of an Nvidia 3060 GPU, available to the author.