

Toward Scalable and Robust Agent Based Methods.

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

Fraser Paterson

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



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

Contents

1	Introduction	2
1.1	Motivation and Background	2
1.2	Active Inference: An Overview	2
1.3	Research Questions	3
2	Previous Work	4
3	Methods and Objectives	5
3.1	Phase 1: Benchmark Active Inference and RL Agents	5
3.2	Phase 2: Deep Active Inference - The Partially Observable Case . . .	6
3.3	Phase 3: Deep, Contrastive Active Inference With Partial Observability	6
4	Software and Hardware Requirements	6
5	Schedule	6

1 Introduction

1.1 Motivation and Background

Many real-world problems are characterized by high degrees of noise, ill-definedness and uncertainty. This uncertainty assumes myriad forms, whether in the clarity of the observations one can solicit from the system of interest, or in the confidence of an inference as to the system-parameter values, that best account for the solicited observation. These tasks are only further complicated by a very common constraint on any candidate solution technique: partial-observability. Indeed, partial-observability is overwhelmingly characteristic of many “difficult” real-world systems.

Any cognitive agent faces a perennial problem, in attempting to ameliorate the above kinds of uncertainty. This consists in obtaining accurate *enough* observations and performing an optimal *enough* action - at the optimal *enough* location and time - for resolving the maximal amount of uncertainty about the dynamics one wishes to predict and/or control.

An effective means by which an artificial agent can managed to ameliorate these constraints, is by maintaining a *model* of its environment. A simplified model of the system at issue, affords an agent the ability to bias its attention to those parts of the system that are likely to be relevant to the task at hand.

Model based artificial agents have enjoyed great success in recent years: Hafner et al. [5], owing to the advantages outlined above (among other things). Indeed, greater sample-efficiency over their model-free cousins is a direct consequence of the model’s ability to bias attention toward the most relevant trajectories through the search space.

Any model-based method is beset by at least two problems. The first concerns the degree to which the model can continue to provide robust predictions, in the face of increasing degrees of uncertainty. This uncertainty may be in the form of noisy state-observations, partial observability or the existence of non-stationary environmental parameters. The second problem concerns the degree to which the model can be used to provide apt predictions in higher-dimensional environments, that is, how well it can “scale up” to a larger problem instance.

Active inference is a recent ambitious theory, purporting to explain how it is that cognitive agents perform optimal actions under uncertainty. This thesis will aim to compare the performance of Active Inference as against various Reinforcement Learning baselines, with respect to their ability to adapt to uncertainty, and their ability to scale up to higher-dimensional problem instances. Active Inference purports to enjoy several theoretical benefits over Reinforcement Learning, such as the absence of a scalar reward signal and the presence of a generative model (among many other examples). These examples, in particular, offer Active Inference a plausible advantage over Reinforcement Learning, in dealing with the above issues of uncertainty and scaling.

1.2 Active Inference: An Overview

Active Inference is a corollary of the Free Energy Principle (FEP) in Physics and Biology: Friston [3]. The FEP is born out of the interface between Statistical Mechanics, Information Theory and Cognitive Science and can trace its roots back to the work of Gibbs, Hamilton and Helmholtz (among many others). Largely

popularized by the work of Karl Friston, the FEP is a plausible, unifying account of brain function in which the brain is supposed as engaging in a scale-invariant process of Variational Free Energy minimization - over sensory data - thereby resisting a thermodynamic tendency to dissolution and ultimately, death.

An Active Inference agent maintains a probabilistic, generative model of its environment which encodes the agent’s beliefs about environmental states (hidden or otherwise), in addition to the agent’s beliefs about how its actions will effect the environment. This generative model can then be used to generate *predictions*, which are compared to the observations solicited by an action. This *prediction error* can then be used to update the agent’s *beliefs* about hidden states, via Bayesian Inference. The prediction error can be expressed as the sensory “surprisal” of an observation. This is simply the negative log-probability of the observation. Now the surprisal itself is often computationally intractable to determine exactly and hence Active Inference makes use of a Variational Free Energy functional of beliefs, to approximate this surprisal.

The Variational Free Energy is provably a lower bound on surprisal. Surprisal quantifies the “atypicality” or “unexpectedness” of an observation. Since the Free Energy is a lower bound on sensory surprisal, if we can merely minimize this VFE, we shall have implemented a means of approximate Bayesian Inference. This method of approximating the true posterior is well known, indeed it is called “Variational Inference”.

Now the agent can only minimize its surprisal vicariously. It can do this by either changing the structure of its generative model, so as to better conform to the present observations (perception) or it can manipulate its environment, so as to change these observations (action). Hence “Active” Inference. In a nutshell, if there is a discrepancy between your model of the world, and your observations about the world, you can either try and change your model or you can try and change the world. Either approach can resolve this discrepancy.

Reinforcement Learning has enjoyed a great deal of success in the attempt to scale to higher dimensional, continuous, noisy environments. While there is still much to be done on this front, Active Inference has thus far been almost entirely limited to small, discrete, stationary environments. 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 Inference might be better able to deal with sources of uncertainty, than the more “traditional” methods in Reinforcement Learning and Optimal Control: Friston, Daunizeau, and Kiebel [4].

1.3 Research Questions

In the course of this proposed study, we shall implement various Active Inference and Reinforcement Learning agents. These implementations will contend with fully, and partially-observable environments, with and without various sources of uncertainty and in both “low-dimensional” non-trivial “higher-dimensional” environments. The aim will be to assess the relative advantages and disadvantages of both “approach families” as these pertain to their ability to deal with various kinds of uncertainty. Additionally, we aim to compare their ability to scale up into higher-dimensional environments.

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

Thus are the central questions raised in this Thesis:

- Are Active Inference agents more robust to noisy observations and non-stationarity than a comparable RL baseline?
- Can Active Inference be more efficiently scaled up to higher dimensional environments than a comparable RL baseline?

2 Previous Work

Work of this kind has already begun to appear in the literature, though it is still very much in its infancy. Marković et al. [6] implemented an Active Inference agent for the multi-armed bandit problem, in the stationary and non-stationary case. The Active Inference 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 Active Inference agent outperformed the UCB agent. While this conducted over a small, discrete space of environmental states, the results plausibly suggest that Active Inference would be an effective means of robust inference and control in a higher-dimensional or continuous problem.

An approach that has enjoyed some success as of late involves the minimization of Free Energy as a process of message-passing on a Forney-style factor graph: Cox, Laar, and Vries [2] and Bagaev and Vries [1]. In this framework, the agent’s generative model is factored in such a way as to instantiate a Forney or “Normal” factor graph. Free Energy minimization 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; affording much more efficient inference and a great potential for scaling up to higher-dimensional settings. As it stands, the main difficulty with this approach is in deciding upon the exact nature of the optimal factorization for the problem at hand. It is scarcely ever obvious or straightforward to determine the optimal structure of the graph, and it can be computationally expensive to achieve this.

Yet other approaches leverage the ability for deep neural networks to approximate the distributions of interest: Ueltzhöffer [9]. Of particular interest are: Tschantz et al. [8] and Mazzaglia, Verbelen, and Dhoedt [7]. The former makes use of amortized inference, in the form of neural network function approximators to parameterize the relevant distributions. Free Energy minimization is then performed with respect 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 stationary 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 efficiency than the baseline agents. Although the approach offered in Tschantz et al. [8] is promising, it was restricted in every case to fully observable environments.

Lastly, the approach of Mazzaglia, Verbelen, and Dhoedt [7], implemented a contrastive method for their Active Inference agent, which significantly reduced the computational burden of learning the parameters for the generative model and planning future actions. This method performed substantially better than the usual, likelihood-based “reconstructive” means of implementing Active Inference 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.

3 Methods and Objectives

In each case we shall aim to investigate our agent’s performance with respect to the following key performance metrics: robustness to noise and non-stationarity, learning rate and sample-efficiency. The proposed structure of our investigation shall be constituted by three major “epochs”, as follows.

3.1 Phase 1: Benchmark Active Inference and RL Agents

Initially, our analysis will be limited to relatively simple, low-dimensional and overwhelmingly fully observable environments. The environments of interest will be the well-known Thermostat, Cart-Pole and Mountain-Car baseline environments. The aim will be to assess the performance - robustness, learning rate and sample efficiency - of the Active Inference and Reinforcement Learning agents across two dimensions. These two dimensions are as follows:

1. Whether the Active Inference agent is implemented with a Forney-style factor graph, generative model, thereby testing performance differences between the “standard” and “graphical” Active Inference methods.
2. Whether the environment is “uncertain” (characterized by either noisy observations or non-stationarity parameters).

The key distributions which encode the Active Inference agents will be instantiated by means of “explicit”, “factor graphical” and “deep” implementations. The RL baseline agent will be simple 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 characterized by a degree of non-stationarity. An example of a potential kind of non-stationarity in the 1 Cart-Pole environment might include randomly changing the length of the pole at iteration n , for all remaining episodes - for instance.

Research questions at this level shall include:

- What is the relative performance of the two approach families, with respect to the key performance metrics? Is one substantively better than the other?
- Does the factor-graphical method appreciably effect the performance metrics of the Active Inference agent?

- Is one approach family - Active Inference or RL - ultimately more robust to either observation noise or parameter non-stationarity?

3.2 Phase 2: Deep Active Inference - The Partially Observable Case

This portion of the study will take up the investigation offered in Tschantz et al. [8] whereby we shall attempt to implement their Deep Active Inference agent, only now in the partially observable case. We shall attempt to implement a Deep Active Inference Pong playing agent, from the Atari library in Open-AI gym, and indeed with several other environments (time permitting). Pong is technically a partially-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 Active Inference as against a DQN baseline, in the stationary and non-stationary POMDP.

Research questions in this “epoch” will repeat those of the last.

3.3 Phase 3: Deep, Contrastive Active Inference With Partial Observability

Time permitting, and assuming that the latter two tasks are achieved with some degree of satisfaction, the analysis will be extended finally to a much higher-dimensional POMDP such as Atari Asteroids. The Deep Active Inference agent will be extended to use the contrastive learning approach developed in Mazzaglia, Verbelen, and Dhoedt [7]. This approach significantly reduces the computational complexity of the model and hence, appears to be a promising avenue of investigation in the extension of Deep Active Inference methods to high-dimensional, partially observable models.

Again, the same suite of questions as in the last two epochs will be recapitulated here.

4 Software and Hardware Requirements

Most of the simulations will be instantiated in Python and associated libraries - such as Open-AI gym, PyTorch and Pyro. RxInfer.jl will be used to implement some of the factor-graph based Active Inference 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.

5 Schedule

- April-May: Phase 1, literature review, revised proposal and code repository.
- June-July: Phase 2, draft Thesis by end of July.
- August-September: Phase 3, Second draft thesis by early September.

- October: Various tasks furthering the analysis of any given implementation, time permitting. This time can also act as a buffer to afford some adaptability in the schedule.

References

- [1] Dmitry Bagaev and Bert de Vries. “Reactive Message Passing for Scalable Bayesian Inference”. In: *CoRR* abs/2112.13251 (2021). arXiv: 2112.13251. URL: <https://arxiv.org/abs/2112.13251>.
- [2] Marco Cox, Thijs van de Laar, and Bert de Vries. “A factor graph approach to automated design of Bayesian signal processing algorithms”. In: *International Journal of Approximate Reasoning* 104 (Jan. 2019), pp. 185–204. DOI: 10.1016/j.ijar.2018.11.002.
- [3] Karl Friston. *The free-energy principle: a rough guide to the brain?* 2009. URL: <https://doi.org/10.1016/j.tics.2009.04.005>.
- [4] Karl J. Friston, Jean Daunizeau, and Stefan J. Kiebel. “Reinforcement Learning or Active Inference?” In: *PLOS ONE* 4.7 (July 2009), pp. 1–13. DOI: 10.1371/journal.pone.0006421. URL: <https://doi.org/10.1371/journal.pone.0006421>.
- [5] Danijar Hafner et al. *Dream to Control: Learning Behaviors by Latent Imagination*. 2020. arXiv: 1912.01603 [cs.LG].
- [6] Dimitrije Marković et al. “An empirical evaluation of active inference in multi-armed bandits”. In: *Neural Networks* 144 (2021), pp. 229–246. ISSN: 0893-6080. DOI: <https://doi.org/10.1016/j.neunet.2021.08.018>. URL: <https://www.sciencedirect.com/science/article/pii/S0893608021003233>.
- [7] Pietro Mazzaglia, Tim Verbelen, and Bart Dhoedt. “Contrastive Active Inference”. In: *Advances in Neural Information Processing Systems*. Ed. by M. Ranzato et al. Vol. 34. Curran Associates, Inc., 2021, pp. 13870–13882. URL: https://proceedings.neurips.cc/paper_files/paper/2021/file/73c730319cf839f143bf40954448ce39-Paper.pdf.
- [8] Alexander Tschantz et al. “Scaling active inference”. In: *CoRR* abs/1911.10601 (2019). arXiv: 1911.10601. URL: <http://arxiv.org/abs/1911.10601>.
- [9] Kai Ueltzhöffer. “Deep active inference”. In: *Biological Cybernetics* 112.6 (Oct. 2018), pp. 547–573. DOI: 10.1007/s00422-018-0785-7.