Toward Scalable and Robust Agent Based Methods.

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).

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

3 April 2023

Contents

0.1 Introduction

0.1.1 Motivation and Background

Many real-world problems are characterised 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 ammount of uncertainty about the dynamics one wishes to predict and/or control.

An effective means by which an artificial agent miht 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 sucess in recent years: Silver et al. [22] and Hafner et al. [12], owing to the advantages outlined above (among other things). Indeed, greater sample-efficency 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 canonical 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. These issues are related, often an attempt to “scale up” a working model will increase the model's complexity and this very often negatively impinges upon the model's robustness.

Indeed an exact Bayesian approach to optimal inference is almost always intractible, due to the necessity for marginalisation over all states - in what is very often, an exponential search space. This is computationally intractible for any but the simplest search spaces and most real-world environments are continuous, high-dimensional and non-stationary.

Active inference is a recent ambitious theory, proporting to explain how it is that cognitive agents perform optimal actions under uncertainty. Before ellaborating upon the central research objectives of this proposed thesis, it will be necessary to provide some background on Active Inference, and to justify its interest as an apt agent-based method.

0.1.2 Active Inference: An Overview

Active Inference (AIF) is a corollary of the Free Energy Principle (FEP), as it pertains to the imperatives attendent upon embodied, cognitive agents. A few words about the FEP shall suffice. The FEP is itslef born out of the interface between Statistical Mechanics, Infomation Theory and Cognitive Science and can trace its roots back to the work of Gibbs, Hamilton and Helmholtz (among many others). Largely popularised 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 minimisation - over sensory data - so as to maximise its own model-evidence (thereby resisting a thermodynamic tendency to dissolution and ultimately, death). See: Friston [7] and Buckley et al. [3] for details.

The Variational Free Energy: $\mathcal{F}$ is a functional of beliefs over uncertain sensory observations. $\mathcal{F}$ is provably a lower bound, on the quantity called "surprisal". Surpirsal quantifies the "atypicallity" or "unexpectedness" of an observation. Since the Free Energy is a lower bound on sensory suprisal, if we can merely minimise this VFE, we shall have implemented a means of approximate Bayesian infernce. This method of approximating the true posterior is well known, indeed it is called "Variational Inference". See Blei, Kucukelbir, and McAuliffe [2] for more.

Now the agent can only minimise its surprisal vicariously. To this end, the agent can either change the structure of its generative model so as to better conform to its present observations (perception) or it can act on its environment, so as to change its 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 in order to resolve this discrepancy.

0.2 Research Objectives

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 entierly limited to small, discrete, stationary environments, see: Sajid, Ball, and Friston [21], Friston et al. [10]. 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: Friston, Daunizeau, and Kiebel [11] and Friston, Samothrakis, and Montague [9].

Hence, in the course of this proposed thesis, we shall implement various 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.

0.2.1 Central Research Questions

The principle aim of this research thesis may be regarded as twofold. The first aim is to investigate the robustness of AIF methods in noisy, uncertain or partially observable environments, relative to Reinforcement Learning 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 raised in this Thesis:

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

0.3 Previous Work

Work of this kind has already begun to appear in the literature, though it is still very much in its infancy. Markovic et al. [15] 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 robust 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: Laar and Vries [14], Cox, Laar, and Vries [5], Bagaev and Vries [1] and Vries and Friston [27]. 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.

Yet other approaches have atempted to leverage the ability for deep neural networks to parameterise the distributions of interest. See: Ueltzhöffer [24], Millidge [18] and Himst and Lanillos [13]. Of particular intetest with this aproach are Tschantz et al. [23] and Mazzaglia, Verbelen, and Dhoedt [17]. 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 Tschantz et al. [23] is promising, it was restricted in every case to fully observable environments.

Lastly, the approach of Mazzaglia, Verbelen, and Dhoedt [17], implemented a contrastive objectove 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, 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.

0.4 Methods and Objectives

In each case we shall aim to investigate our agent’s performance with respect to the following key performance metrics: robustenss 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.

0.4.1 Epoch 1: Benchmark AIF 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, 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 with a Forney-style factor graph, generative model, thereby testing performance differences between the “standard” and “graphical“ AIF methods.
2. Whether the environment is “uncertain” (characterised by either noisy observations or non-stationarity parameters).

The key distributions which encode the AIF 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 characterised by a degree of non-stationarity. An example of a potential kind of non-stationarity in the CartPole 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 AIF agent?

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

0.4.2 Epoch 2: Deep AIF - The Partially Observable Case

This portion of the study will take up the investigation offered in Tschantz et al. [23] 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 playing agent, from the Atari library in OpenAI gym, and indeed with several other environments (time permitting). 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 in this "epoch" will repeat those of the last.

0.4.3 Epoch 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 Atari Asteroids. The Deep AIF agent will be extended to use the contrastive learning approach developed in Mazzaglia, Verbelen, and Dhoedt [17]. 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.

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

0.5 Software and Hardware Requirements

Most of the simulations will be instantiated in Python and associated libraries - such as OpenAI gym and Pytorch and Pyro. RxInfer.jl will be used to implement some of 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.

0.6 Rough Schedule

- April-May: Epoch 1
- June-July: Epoch 2
- August-September: Epoch 3
- October: Miscellaneuos 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.

Bibliography

[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] David M. Blei, Alp Kucukelbir, and Jon D. McAuliffe. "Variational Inference: A Review for Statisticians". In: *Journal of the American Statistical Association* 112.518 (Apr. 2017), pp. 859–877. DOI: 10.1080/01621459.2017.1285773.

[3] Christopher L. Buckley et al. "The free energy principle for action and perception: A mathematical review". In: *Journal of Mathematical Psychology* 81 (2017), pp. 55–79. ISSN: 0022-2496. DOI: https://doi.org/10.1016/j.jmp.2017.09.004. URL: https://www.sciencedirect.com/science/article/pii/S0022249617300962.

[4] Christopher Cherniak. *Minimal Rationality.* MIT Press, 1986.

[5] 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.

[6] K Friston et al. "Action and behavior: a free-energy formulation". In: *Biological cybernetics* 102.3 (2010), pp. 227–260. DOI: 10.1007/s00422-010-0364-z. URL: https://doi.org/10.1007/s00422-010-0364-z.

[7] Karl Friston. *The free-energy principle: a rough guide to the brain?* 2009. URL: https://doi.org/10.1016/j.tics.2009.04.005.

[8] Karl Friston. "What Is Optimal about Motor Control?" In: *Neuron* 72.3 (2011), pp. 488–498. ISSN: 0896-6273. DOI: https://doi.org/10.1016/j.neuron.2011.10.018. URL: https://www.sciencedirect.com/science/article/pii/S0896627311009305.

[9] Karl Friston, Spyridon Samothrakis, and Read Montague. "Active inference and agency: optimal control without cost functions". In: *Biological Cybernetics* 106.8 (Oct. 2012), pp. 523–541. ISSN: 1432-0770. DOI: 10.1007/s00422-012-0512-8. URL: https://doi.org/10.1007/s00422-012-0512-8.

[10] Karl Friston et al. "Active Inference, Curiosity and Insight". In: *Neural Computation* 29 (Aug. 2017), pp. 1–51. DOI: 10.1162/neco_a_00999.

[11] 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.

[12] Danijar Hafner et al. *Dream to Control: Learning Behaviors by Latent Imagination.* 2020. arXiv: 1912.01603 [cs.LG].

[13] Otto van der Himst and Pablo Lanillos. “Deep Active Inference for Partially Observable MDPs”. In: *CoRR* abs/2009.03622 (2020). arXiv: 2009.03622. URL: https://arxiv.org/abs/2009.03622.

[14] Thijs W. van de Laar and Bert de Vries. “Simulating Active Inference Processes by Message Passing”. In: *Frontiers in Robotics and AI* 6 (2019). ISSN: 2296-9144. DOI: 10.3389/frobt.2019.00020. URL: https://www.frontiersin.org/articles/10.3389/frobt.2019.00020.

[15] Dimitrije Markovic et al. “An empirical evaluation of active inference in multi-armed bandits”. In: *CoRR* abs/2101.08699 (2021). arXiv: 2101.08699. URL: https://arxiv.org/abs/2101.08699.

[16] 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.

[17] Pietro Mazzaglia, Tim Verbelen, and Bart Dhoedt. “Contrastive Active Inference”. In: *CoRR* abs/2110.10083 (2021). arXiv: 2110.10083. URL: https://arxiv.org/abs/2110.10083.

[18] Beren Millidge. “Deep Active Inference as Variational Policy Gradients”. In: (July 2019). URL: https://arxiv.org/pdf/1907.03876.pdf.

[19] Allan Newell and Herbert Simon. *Human problem solving.* Prentice-Hall, 1972.

[20] “Predictive processing and relevance realization: exploring convergent solutions to the frame problem”. In: *Phenom Cogn Sci* (2022). URL: https://doi.org/%2010.1007/s11097-022-09850-6.

[21] Noor Sajid, Philip J. Ball, and Karl J. Friston. “Demystifying active inference”. In: *CoRR* abs/1909.10863 (2019). arXiv: 1909.10863. URL: http://arxiv.org/abs/1909.10863.

[22] David Silver et al. “Mastering the game of Go without human knowledge”. In: *Nature* 550.7676 (Oct. 2017), pp. 354–359. ISSN: 1476-4687. DOI: 10.1038/nature24270. URL: https://doi.org/10.1038/nature24270.

[23] Alexander Tschantz et al. “Scaling active inference”. In: *CoRR* abs/1911.10601 (2019). arXiv: 1911.10601. URL: http://arxiv.org/abs/1911.10601.

[24] Kai Ueltzhöffer. “Deep active inference”. In: *Biological Cybernetics* 112.6 (Oct. 2018), pp. 547–573. DOI: 10.1007/s00422-018-0785-7.

[25] John Vervaeke. *Ep. 27 - Awakening from the Meaning Crisis - Problem Formulation.* Youtube. 2019.

[26] John Vervaeke, Timothy P. Lillicrap, and Blake A. Richards. “Relevance Realization and the Emerging Framework in Cognitive Science”. In: *Journal of Logic and Computation* 22.1 (Oct. 2009), pp. 79–99. ISSN: 0955-792X. DOI: 10.1093/logcom/exp067. eprint: https://academic.oup.com/logcom/article-pdf/22/1/79/3262477/exp067.pdf. URL: https://doi.org/10.1093/logcom/exp067.

[27] Bert de Vries and Karl J. Friston. “A Factor Graph Description of Deep Temporal Active Inference”. In: *Frontiers in Computational Neuroscience* 11 (2017). ISSN: 1662-5188. DOI: `10.3389/fncom.2017.00095`. URL: `https://www.frontiersin.org/articles/10.3389/fncom.2017.00095`.