BATMAN: BAYESIAN TARGET MODELLING FOR ACTIVE INFERENCE

Magnus T. Koudahl *

Eindhoven University of Technology
Groene Loper 19, 5612 AP Eindhoven (NL)
m.t.koudahl@tue.nl

Bert de Vries

Eindhoven Univ. of Technology & GN Hearing
Groene Loper 19, 5612 AP Eindhoven (NL)
bert.de.vries@tue.nl

ABSTRACT

Active Inference is an emerging framework for designing intelligent agents. In an Active Inference setting, any task is formulated as a variational free energy minimisation problem on a generative probabilistic model. Goal-directed behaviour relies on a clear specification of desired future observations. *Learning* desired observations would open up the Active Inference approach to problems where these are difficult to specify a priori. This paper introduces the BAyesian Target Modelling for Active iNference (BATMAN) approach, which augments an Active Inference agent with an additional, separate model that learns desired future observations from a separate data source. The main contribution of this paper is the design of a coupled generative model structure that facilitates learning desired future observations for Active Inference agents and supports integration of Active Inference and classical methods in a joint framework. We provide proof-of-concept validation for BATMAN through simulations.

Index Terms— Active Inference, variational inference, Graphical Models, Adaptive Agents, Bayesian

1. INTRODUCTION

Active Inference is emerging as a biologically grounded framework for designing intelligent agents. Originating in the field of computational neuroscience, it was conceived as a Bayesian model of how brains of biological agents perceive and act [1].

A hallmark feature of Active Inference is that we can cast any task as a variational free energy minimisation problem on a probabilistic generative model[1]. Recent work has started to apply these principles to the design of synthetic agents [2, 3, 4].

To elicit goal-directed behaviour from minimising free energy, Active Inference relies on encoding desired future observations. The usual approach to this encoding step [5, 6, 7] is to employ *goal priors* specifying desired future observations. However this may become a bottleneck for applying Active Inference in situations where desired future observations are either unknown or prohibitively difficult to specify. In this paper we introduce an alternative approach to encoding goals by extending the notion of what can constitute a goal prior.

The main contribution of this paper is the Bayesian Target Modelling for Active iNference (BATMAN) architectural framework for designing Active Inference agents. We show that the notion of a goal prior can be extended to include a full probabilistic model. This move allows alternate encodings of attracting states, providing an opportunity to elicit goal directed behaviour without having to specify future observations explicitly. An extended view of goal priors may prove relevant to the design of synthetic Active Inference agents. As an illustrative example we consider the problem of teaching an Active Inference agent where to park a cart from performance appraisals.

2. ACTIVE INFERENCE

Active Inference is a unifying framework for perception and action [1, 5]. At its core, Active Inference assumes that an agent entails a probabilistic model of its environment and is continually engaged in the task of improving accuracy and minimising complexity of this model [1]. Formally this is accomplished through minimising variational free energy between observed and predicted sensory inputs [1]. The agent can accomplish this in two ways: Either by updating its generative model (perception/learning) or performing actions on the environment to elicit inputs consistent with its model.

At the heart of an Active Inference agent is a generative model $p(x, y, u)$. This encompasses sequences of observations y and two sequences of hidden states x and u. Internal states x model the evolving state of the agent while control states u affect environmental states through the agents actuators. The generative model takes the form of a dynamical system since the agent must consider the future in order to satisfy its goals.

Additionally the agent is equipped with a posterior or recognition model $q(x, y_{t+1:T}, u)$ that encodes approximate Bayesian posterior beliefs about the agent's control states u, internal states x after having observed data and observations

*Corresponding Author. This work is part of the research programme Efficient Deep Learning with project number P16-25 project 5, which is (partly) financed by the Netherlands Organisation for Scientific Research (NWO).

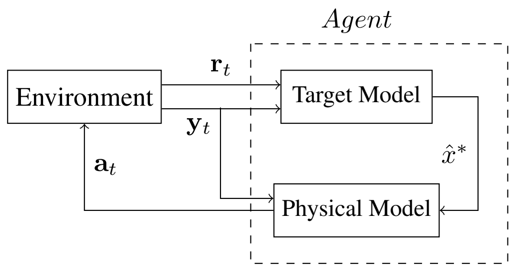

Fig. 1. Block diagram showing communication channels between elements of the experimental setup.

$y_{t+1:T}$ after the current time step t until some horizon T. To realise actions and receive observations the agent is equipped with a set of actuators and sensory channels.

To elicit goal directed behaviour the agent is usually endowed with *goal priors* [1, 5] specifying observations that the agent expects to encounter in the future. This constrains the inference problem ensuring the agent plans a sequence of actions consistent with the goal. The function of goal priors is thus to constrain the inference problem by introducing attractors in the agents state space [8]. This formulation is consistent across both discrete and continous time versions of Active Inference. For clarity we focus on the discrete time version going forward. Applying BATMAN in continous time constitutes a potential avenue for future research.

3. A CART PARKING TASK

Consider an agent tasked with parking a cart on a rail. The cart is equipped with an engine allowing it to move back and forth. In addition we attach a sensory channel that provides noisy binary appraisals as to whether the cart is moving away from or towards the target. Finally, the cart is equipped with a sensor that provides information about its current position. The simulation consists of running separate models for the environment (Sec. 3) and the agent (Sec. 4). A schematic representation of the entities in our simulation and the quantities passed between them within a time step is shown in Fig. 1.

Our simulated environmental process defines (1) the equations of motion governing movement of the cart as well as the processes for generating (2) the position and (3) appraisal data. This requires three functions - one for each sub-process.

Using $\mathbf{bold}$ font to denote environmental variables, we define the equations of motion in terms of the current state $\mathbf{x}_t \in \mathbb{R}$ (the position) of the cart at time t and action $\mathbf{a}_t \in \mathbb{R}$ (the force) administered by the agent at time t. We define the update equation

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \tanh(\mathbf{a}_t)\,, \tag{1}$$

where the nonlinearity $\tanh(\cdot)$ limits the velocity of the cart to the interval $(-1, 1)$. This constraint captures the upper limit on velocity the engine can generate.

To produce sensory input for the agent on its position, we assume fixed measurement noise and draw measurements $\mathbf{y}_t$ by

$$\mathbf{y}_t \sim \mathcal{N}(\mathbf{x}_t, 1)\,. \tag{2}$$

Throughout the paper we will use the mean-variance parameterisation for all Gaussian distributions. Finally, the environment provides appraisals based on evaluations of the utility function

$$U(\mathbf{x}_t; \mathbf{x}^*, \lambda) = -\sqrt{\lambda}|\mathbf{x}_t - \mathbf{x}^*| \tag{3}$$

where $\mathbf{x}^*$ denotes the true target and λ is a precision parameter controlling the width of the utility function. In our experiments we work in $1D$ but scaling to higher dimensions is possible [9].

To compare two positions $\mathbf{x}_t, \mathbf{x}_{t-1}$ and generate a performance appraisal signal $\mathbf{r}_t$, we rely on the difference in respective utilities as calculated by Eq. 3. This is passed through a logistic function and the output is used to parameterise a Bernoulli distribution, from which we sample $\mathbf{r}_t$ by

$$\mathbf{r}_t \sim \mathcal{B}er\left(\sigma(U(\mathbf{x}_t; \mathbf{x}^*, \lambda) - U(\mathbf{x}_{t-1}; \mathbf{x}^*, \lambda))\right)\,. \tag{4}$$

The environmental process governing our experiment is thus fully described by Eqs. 1- 4.

4. THE AGENT

Our proposed agent architecture relies on coupling two models. The first model, denoted the Physical Model, performs inference tasks related to trajectories and policies. This part mirrors an Active Inference setup similar to [10]. The second model, denoted the Target Model, replaces the notion of a goal prior. The task of this model is to infer the correct goal state from performance appraisals $\mathbf{r}_t$.

In order to visualise the two models, we employ the Forney-style Factor Graph (FFG) formalism. An FFG is a representation of a factorized probability distribution where edges represent variables and nodes represent factors. The edge of an observed variable is terminated by a small black square. If a section of the FFG repeats over time steps, we will indicate this by an ellipsis. An edge is connected to a node iff the variable is an argument of that factor. For a thorough introduction to FFGs, see [11].

4.1. The Physical Model

The Physical Model follows a common state space factorization [12] where we assume that successive states obey the Markov property. A single time slice of the corresponding FFG is shown in Fig. 2. Let x_t denote the hidden internal state of the agent, y_t an observation and u_t a (hidden) control

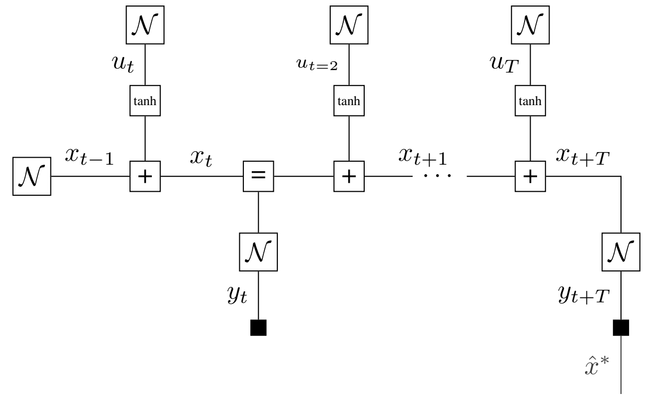

Fig. 2. FFG of the Physical Model at time t.

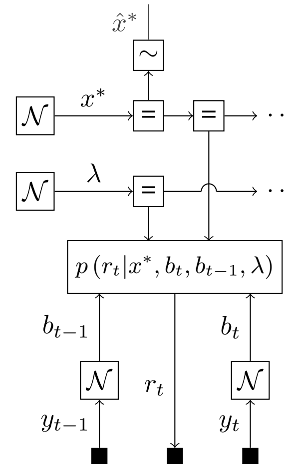

Fig. 3. FFG of a single time slice of the Target Model.

state at time t. At each time step the generative model for the next T time steps is given by

$$p(x, y, u|\hat{x}^*) \propto \tag{5}$$

$$p(x_{t-1}) \underbrace{p'(y_{t+T}|\hat{x}^*)}_{\text{Goal Prior}} \prod_{k=t}^{t+T} \underbrace{p(y_k|x_k)p(x_k|x_{k-1}, u_k)p(u_k)}_{\text{State Space Model}} .$$

We use the notation of [10] to write goal priors. $\delta(\cdot)$ denotes the Dirac delta. $\hat{x}^*$ is set by the Target Model and detailed in Sec. 4.2. We specify the factors in Eq. 5 as follows:

$$p(x_k|x_{k-1}, u_k) = \delta(x_{k-1} + \tanh(u_k) - x_k) \tag{6a}$$
$$p(x_0) = \mathcal{N}(x_0|\mathbf{x}_0, 1) \tag{6b}$$
$$p(y_k|x_k) = \mathcal{N}(y_k|x_k, 1) \tag{6c}$$
$$p(u_k) = \mathcal{N}(u_k|0, 10) \tag{6d}$$
$$p'(y_k|\hat{x}^*) = \delta(y_k - \hat{x}^*) . \tag{6e}$$

In Eq. 6a, we assume the Physical Model has access to an accurate and deterministic transition model. For the mathematical details of this operation we refer to [11].

4.2. The Target Model

The Target Model induces attracting states for the agent by parameterising the goal prior. The task of this model is to infer the true target position $\mathbf{x}^*$ and communicate this to the Physical Model by fixing the value of $\hat{x}^*$. To elicit goal directed behaviour we assume the agent has accurate knowledge of the functional form of the environmental utility function, Eq. 3. x^* estimates the position of the peak of the utility function, corresponding to $\mathbf{x}^*$. We consider an adaptation of [9] for the Target Model:

$$p(r_t, b_t, b_{t-1}, x^*, \lambda|y_t, y_{t-1}) =$$
$$p(\lambda)p(x^*)p(b_t|y_t)p(b_{t-1}|y_{t-1})p(r_t|x^*, b_t, b_{t-1}, \lambda) \tag{7}$$

where

$$p(b_t|y_t) = \mathcal{N}(b_t|y_t, 1) \tag{8a}$$
$$p(b_{t-1}|y_{t-1}) = \mathcal{N}(b_{t-1}|y_{t-1}, 1) \tag{8b}$$
$$\log p(\lambda) = \mathcal{N}(\lambda|0, 1) \tag{8c}$$
$$p(x^*) = \mathcal{N}(x^*|\mathbf{x}_0, 10) \tag{8d}$$
$$p(r_t|x^*, b_t, b_{t-1}, \lambda) =$$
$$\mathcal{B}er(r_t|\sigma(U(b_t;x^*, \lambda) - U(b_{t-1}; x^*, \lambda))) \tag{8e}$$
$$\hat{x}^* \sim q(x^*) . \tag{8f}$$

An FFG of a single time slice of the model is shown in Fig 3. To link the Target and Physical Models we sample the current goal state $\hat{x}^* \sim q(x^*)$. The sample is then used to parameterise the goal prior $p'(y_k|\hat{x}^*)$ of the Physical Model. Experiments utilizing the full posterior distribution showed decreased performance due to compound variance from $q(x^*)$ and $p(y_{t+T}|x_{t+T})$. For an exposition on the role of precision in controlling action, we refer to [13, 14]

Simulation Loop:

for $t = 1, 2, \ldots$ **do**
- *Agent*
- Observe new data (y_t, r_t)
- Infer new Target Model beliefs $q(x^*)$,$q(\lambda)$
- Sample new goal prior $\hat{x}^* \sim q(x^*)$
- Infer new Physical Model beliefs $q(u_t)$
- Sample new action $\mathbf{a}_t \sim q(u_{t+1})$
- *Environment*
- Update environmental states $\mathbf{x}_{t+1}$
- Generate new observations $\mathbf{y}_{t+1}$, $\mathbf{r}_{t+1}$
- **End**

Fig. 4. Experimental Protocol

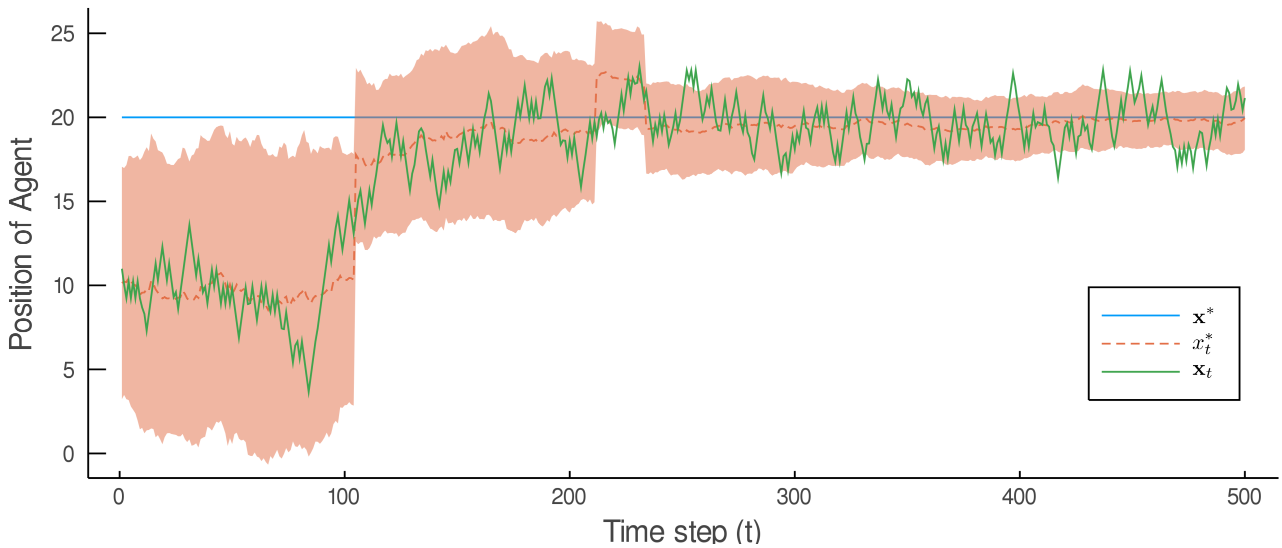

Fig. 5. Trajectory of true state $\mathbf{x}_t$ and mean posterior belief over x^* during an experimental run. x^* augmented by error bars of 1 standard deviation.

5. EXPERIMENTS

We validated the BATMAN approach on the problem of learning to park a cart on rails from binary performance appraisals. The agent's goal is to reach a particular point $\mathbf{x}^*$ on the rails, within a margin of uncertainty. We simulated the environmental processes described by Eqs. 1- 4. We initialise the environment with $\mathbf{x}_0 = 10$, $\mathbf{x}^* = 20$, $\lambda = 7$ and the Physical Model with parameter $T = 2$. The Physical Model was implemented in `ForneyLab` [15] and the Target Model in `Turing` [16]. The steps taken within a single time step is presented in Fig. 4. The experiment ran for 500 iterations. Initially we perform inference in the Target Model by Sequential Monte Carlo [17], drawing 1000 samples at each time step. The goal is to obtain posterior estimates of $q(x^*)$ from which to sample an informative goal prior and $q(\lambda)$ to refine parameter estimates for the next iteration. Subsequently a sample from the posterior $\hat{x}^* \sim q(x^*)$ is passed to the goal prior of the physical model. Inference in the Physical Model is performed by sum-product message passing [18] allowing us to obtain exact marginals and infer posterior beliefs over control states $q(u)$. As seen in Fig. 5 the agent successfully learns the true goal $\mathbf{x}^*$ (blue line) and positions the cart (green line) close. This provides a proof-of-concept validation for our approach.

6. RELATED WORK

Utilising coupled models in Active Inference has mostly been studied in the neuroscience community. In [19] the authors investigated the human visual system using hierarchical coupled models to imitate brain structure. Notably [19] interfaced two models at the level of observations by attaching them to the same inputs. This results in an architecture that is close to BATMAN but differs in two important ways: one, BATMAN attaches only at the goal prior and two, BATMAN utilizes a separate feedback channel. These differences define BATMAN as specifically a mechanism for inducing goal-directed behaviour. In [20] the coupling of models was likened to connections between cortical areas, leading to a formulation of cortical connectivity as message passing algorithms. While both of these works are impressive, their focus is mainly the theoretical and neuroscientific aspects of Active Inference. Exploring the relation between the BATMAN framework and the biological plausibility of [19, 20] is a promising direction for future research.

7. DISCUSSION AND CONCLUSIONS

In this paper we introduced the BATMAN approach to Active Inference-based agent design. Specifically we showed that by extending the goal prior to a full probabilistic model, it is possible to induce attractors in the agents state space that are consistent with goal directed behaviour without having to manually specify future observations. We provided proof-of-concept validation of our approach through simulation. This approach provides a novel way of eliciting goal directed behaviour from synthetic Active Inference agents that may allow scaling to domains where goal priors are prohibitively difficult to specify a priori.

8. ACKNOWLEDGEMENTS

We wish to thank Thijs van de Laar and Wouter Kouw for their invaluable insights during the writing of this manuscript.

9. REFERENCES

[1] Karl Friston, “The free-energy principle: a rough guide to the brain?,” *Trends in Cognitive Sciences*, vol. 13, no. 7, pp. 293–301, July 2009.

[2] Kai Ueltzhöffer, “Deep Active Inference,” *Biological Cybernetics*, Oct. 2018.

[3] Maell Cullen, Ben Davey, Karl J. Friston, and Rosalyn J. Moran, “Active Inference in OpenAI Gym: A Paradigm for Computational Investigations Into Psychiatric Illness,” *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, vol. 3, no. 9, pp. 809–818, Sept. 2018.

[4] Simon McGregor, Manuel Baltieri, and Christopher L. Buckley, “A Minimal Active Inference Agent,” *arXiv:1503.04187 [cs]*, Mar. 2015, arXiv: 1503.04187.

[5] Thomas Parr and Karl J. Friston, “Generalised free energy and active inference: can the future cause the past?,” *bioRxiv*, p. 304782, 2018.

[6] Raphael Kaplan and Karl J. Friston, “Planning and navigation as active inference,” *Biological Cybernetics*, vol. 112, no. 4, pp. 323–343, Aug. 2018.

[7] Karl Friston, Thomas FitzGerald, Francesco Rigoli, Philipp Schwartenbeck, John O’Doherty, and Giovanni Pezzulo, “Active inference and learning,” *Neuroscience & Biobehavioral Reviews*, 2016.

[8] Karl Friston and Ping Ao, “Free energy, value, and attractors,” *Computational and mathematical methods in medicine*, vol. 2012, 2012.

[9] Marco Cox and Bert de Vries, “A parametric approach to Bayesian optimization with pairwise comparisons,” in *NIPS Workshop on Bayesian Optimization (BayesOpt 2017)*, Long Beach, USA, Dec. 2017.

[10] Thijs van de Laar and Bert de Vries, “Simulating Active Inference Processes by Message Passing,” *Frontiers in Robotics and AI*, vol. 6, pp. 20, 2019.

[11] Sascha Korl, *A factor graph approach to signal modelling, system identification and filtering*, Ph.D. thesis, Swiss Federal Institute of Technology, Zurich, 2005.

[12] Daphne Koller and Nir Friedman, *Probabilistic graphical models: principles and techniques*, MIT press, 2009.

[13] Giovanni Pezzulo, Francesco Rigoli, and Karl Friston, “Active inference, homeostatic regulation and adaptive behavioural control,” *Progress in Neurobiology*, vol. 134, pp. 17 – 35, 2015.

[14] Karl Friston, Thomas FitzGerald, Francesco Rigoli, Philipp Schwartenbeck, and Giovanni Pezzulo, “Active inference: a process theory,” *Neural computation*, vol. 29, no. 1, pp. 1–49, 2017.

[15] Marco Cox, Thijs van de Laar, and Bert de Vries, “A factor graph approach to automated design of bayesian signal processing algorithms,” *International Journal of Approximate Reasoning*, vol. 104, pp. 185–204, 2019.

[16] Hong Ge, Kai Xu, and Zoubin Ghahramani, “Turing: Composable inference for probabilistic programming,” in *International Conference on Artificial Intelligence and Statistics, AISTATS 2018, 9-11 April 2018, Playa Blanca, Lanzarote, Canary Islands, Spain*, Amos J. Storkey and Fernando Pérez-Cruz, Eds. 2018, vol. 84 of *Proceedings of Machine Learning Research*, pp. 1682–1690, PMLR.

[17] Jun S. Liu and Rong Chen, “Sequential monte carlo methods for dynamic systems,” *Journal of the American Statistical Association*, vol. 93, no. 443, pp. 1032–1044, 1998.

[18] Hans-Andrea Loeliger, Justin Dauwels, Junli Hu, Sascha Korl, Li Ping, and Frank R. Kschischang, “The Factor Graph Approach to Model-Based Signal Processing,” *Proceedings of the IEEE*, vol. 95, no. 6, pp. 1295–1322, June 2007.

[19] Thomas Parr and Karl J. Friston, “The Discrete and Continuous Brain: From Decisions to Movement—and Back Again,” *Neural Computation*, pp. 1–29, June 2018.

[20] Thomas Parr and Karl J Friston, “The anatomy of inference: generative models and brain structure,” *Frontiers in computational neuroscience*, vol. 12, pp. 90, 2018.