Deriving Intrinsic Motivation from Uncertainty about Future Goals

https://github.com/jnegrea/csc2541project1

Jeffrey Negrea

Department of Statistical Sciences
University of Toronto
negrea@utstat.toronto.edu

November 17, 2016

Table of Contents

Table of Contents

Heuristics of Reinforcement Learning

- Recall that *reinforcement learning* addresses the problem of how machines can learn to perform a task by trial and error.

Heuristics of Reinforcement Learning

- Recall that *reinforcement learning* addresses the problem of how machines can learn to perform a task by trial and error.
  - A reward mechanism is pre-prescribed for the learner
  - The learner does not know the topology of the state space
  - The learner does not know the reward of each state
  - The learner does not know the transition mechanics

Heuristics of Reinforcement Learning

- Recall that *reinforcement learning* addresses the problem of how machines can learn to perform a task by trial and error.
  - A reward mechanism is pre-prescribed for the learner
  - The learner does not know the topology of the state space
  - The learner does not know the reward of each state
  - The learner does not know the transition mechanics
- The machine learns the topology of the state space, the transition mechanics, the topography of the reward function, and how to use the mechanics to optimise the specified reward.

Questions Leading to Intrinsic Motivation

- What if the reward function has yet to be specified?

Questions Leading to Intrinsic Motivation

- What if the reward function has yet to be specified?
- For animal learners, reward signals are recieved by engaging in exploratory, ‘fun’ behaviours. Can machines be incited to play and explore? To have fun?

Questions Leading to Intrinsic Motivation

- What if the reward function has yet to be specified?
- For animal learners, reward signals are recieved by engaging in exploratory, 'fun' behaviours. Can machines be incited to play and explore? To have fun?
- Exploratory behaviour is important for animal development – human children learn about the world by playing. Can a machine which learns by playing outperform one that does not when performance is integrated over a wide variety of tasks?

Table of Contents

Qualitative Aspects of Intrinsic Motivation I

- In order to teach a machine to 'play' we need measures which quantify what it means to 'play' effectively.
- In order to develop such measures we need heuristics which defines what it means to 'play effectively.'

Qualitative Aspects of Intrinsic Motivation I

- In order to teach a machine to 'play' we need measures which quantify what it means to 'play' effectively.
- In order to develop such measures we need heuristics which defines what it means to 'play effectively.'
- Schmidhuber [12] posits that an intrinsically motivated learner should have:
  - An adaptive world model
  - A learning algorithm to update the model
  - An intrinsic reward system measuring model improvement
  - A behavioural policy optimising intrinsic reward

Qualitative Aspects of Intrinsic Motivation I

- In order to teach a machine to ‘play’ we need measures which quantify what it means to ‘play’ effectively.
- In order to develop such measures we need heuristics which defines what it means to ‘play effectively.’
- Schmidhuber [12] posits that an intrinsically motivated learner should have:
  - An adaptive world model
  - A learning algorithm to update the model
  - An intrinsic reward system measuring model improvement
  - A behavioural policy optimising intrinsic reward
- The kernel of this thesis is that intrinsic motivation is derived from the pursuit of a better model of the dynamics of the world.

Qualitative Aspects of Intrinsic Motivation II

- Salge et al. [11] posit that an intrinsically motivated learner should seek states of high (heuristic) *empowerment*.
- Empowerment is an intrinsic utility measure which is *Local, Universal, and Task-Independent*:

Qualitative Aspects of Intrinsic Motivation II

- Salge et al. [11] posit that an intrinsically motivated learner should seek states of high (heuristic) *empowerment*.
- Empowerment is an intrinsic utility measure which is *Local, Universal, and Task-Independent*:
  - *Local*: An agent can determine intrinsic utility of a state from the local environment

Qualitative Aspects of Intrinsic Motivation II

- Salge et al. [11] posit that an intrinsically motivated learner should seek states of high (heuristic) *empowerment*.
- Empowerment is an intrinsic utility measure which is *Local, Universal, and Task-Independent*:
  - *Local*: An agent can determine intrinsic utility of a state from the local environment
  - *Universal*: The utility scale should be independent of the structure of the actor and environment – utility should be comparable across distinct problem classes

Qualitative Aspects of Intrinsic Motivation II

- Salge et al. [11] posit that an intrinsically motivated learner should seek states of high (heuristic) *empowerment*.
- Empowerment is an intrinsic utility measure which is *Local, Universal, and Task-Independent*:
    - *Local*: An agent can determine intrinsic utility of a state from the local environment
    - *Universal*: The utility scale should be independent of the structure of the actor and environment – utility should be comparable across distinct problem classes
    - *Task-Independent*: The utility is independent of any particular goal or reward function

Table of Contents

Quantitative Measures for Intrinsic Motivation I: Optimal Reward

- Singh et al. [13] introduces the concept of *optimal reward functions*.
- Requires a fixed fitness function to be pre assigned, as in traditional RL

Quantitative Measures for Intrinsic Motivation I: Optimal Reward

- Singh et al. [13] introduces the concept of *optimal reward functions*.
- Requires a fixed fitness function to be pre assigned, as in traditional RL
- The learning problem is posed as a two-stage optimisation problem;

Quantitative Measures for Intrinsic Motivation I: Optimal Reward

- Singh et al. [13] introduces the concept of *optimal reward functions*.
- Requires a fixed fitness function to be pre assigned, as in traditional RL
- The learning problem is posed as a two-stage optimisation problem;
  - Find the reward function such that an RL agent maximising the reward has highest expected fitness

Quantitative Measures for Intrinsic Motivation I: Optimal Reward

- Singh et al. [13] introduces the concept of *optimal reward functions*.
- Requires a fixed fitness function to be pre assigned, as in traditional RL
- The learning problem is posed as a two-stage optimisation problem;
  - Find the reward function such that an RL agent maximising the reward has highest expected fitness
  - Find the RL strategy which maximises the optimal reward function

Quantitative Measures for Intrinsic Motivation I: Optimal Reward

- Singh et al. [13] introduces the concept of *optimal reward functions*.
- Requires a fixed fitness function to be pre assigned, as in traditional RL
- The learning problem is posed as a two-stage optimisation problem;
  - Find the reward function such that an RL agent maximising the reward has highest expected fitness
  - Find the RL strategy which maximises the optimal reward function
- By construction, performs no worse on average than RL using the the natural reward corresponding to fitness.
- Avoids greedy behaviour in favour of exploratory behaviour (when beneficial).

Quantitative Measures for Intrinsic Motivation II: Empowerment

- *Empowerment* (metric) aims to achieve the heuristic it is named for
- Quantified as the channel capacity of a state:

$$\mathcal{E}(s) = \max_{\omega \in \Omega_s} \mathcal{I}(A, S_1 | s) = \max_{\omega \in \Omega_s} \mathbb{E}\left[\log \frac{p(S_1, A|s)}{\omega(A|s)p(S_1|s)}\right] = \max_{\omega \in \Omega_s} \mathbb{E}\left[\log \frac{p(S_1|s, A)}{p(S_1|s)}\right]$$

Where the expectation is taken with respect to the joint distribution of the action and the resultant state, (A, S_1), conditional on the starting state, s, for a fixed choice of ω — the distribution of the action given the starting state.

Quantitative Measures for Intrinsic Motivation II: Empowerment

- *Empowerment* (metric) aims to achieve the heuristic it is named for
- Quantified as the channel capacity of a state:

$$\mathcal{E}(s) = \max_{\omega \in \Omega_s} \mathcal{I}(A, S_1 | s) = \max_{\omega \in \Omega_s} \mathbb{E}\left[\log \frac{p(S_1, A|s)}{\omega(A|s)p(S_1|s)}\right] = \max_{\omega \in \Omega_s} \mathbb{E}\left[\log \frac{p(S_1|s, A)}{p(S_1|s)}\right]$$

Where the expectation is taken with respect to the joint distribution of the action and the resultant state, (A, S_1), conditional on the starting state, s, for a fixed choice of ω — the distribution of the action given the starting state.

- Discussed thoroughly in Singh et al. [13], applied in Mohamed and Rezende [9]
- Aims to find the state in which an actor is most able to travel to any arbitrary state

Quantitative Measures for Intrinsic Motivation II: Empowerment

- *Empowerment* (metric) aims to achieve the heuristic it is named for
- Quantified as the channel capacity of a state:

$$\mathcal{E}(s) = \max_{\omega \in \Omega_s} \mathcal{I}(A, S_1 | s) = \max_{\omega \in \Omega_s} \mathbb{E}\left[\log \frac{p(S_1, A|s)}{\omega(A|s)p(S_1|s)}\right] = \max_{\omega \in \Omega_s} \mathbb{E}\left[\log \frac{p(S_1|s, A)}{p(S_1|s)}\right]$$

Where the expectation is taken with respect to the joint distribution of the action and the resultant state, (A, S_1), conditional on the starting state, s, for a fixed choice of ω — the distribution of the action given the starting state.

- Discussed thoroughly in Singh et al. [13], applied in Mohamed and Rezende [9]
- Aims to find the state in which an actor is most able to travel to any arbitrary state

- Not obvious if the information theoretic definition is suitable
- Does not consider the how rewards may be assigned in the future

Quantitative Measures for Intrinsic Motivation III: Expected Reward

- Purpose of my project: Define a Bayesian framework for intrinsically motivation.
- Assume the agent has an internal prior on the class of possible reward functions

Quantitative Measures for Intrinsic Motivation III: Expected Reward

- Purpose of my project: Define a Bayesian framework for intrinsically motivation.
- Assume the agent has an internal prior on the class of possible reward functions
- Determine the policy which maximises the *intrinsic expected reward* (iER)
  - Expectation is taken with respect to the random reward function, possibly random choice of action, and the results of actions (where the latter two may form a feedback loop)

Quantitative Measures for Intrinsic Motivation III: Expected Reward

- Purpose of my project: Define a Bayesian framework for intrinsically motivation.
- Assume the agent has an internal prior on the class of possible reward functions
- Determine the policy which maximises the *intrinsic expected reward* (iER)
  - Expectation is taken with respect to the random reward function, possibly random choice of action, and the results of actions (where the latter two may form a feedback loop)
  - *Intrinsic* since the prior on potential rewards is internal to the agent

Quantitative Measures for Intrinsic Motivation III: Expected Reward

- Purpose of my project: Define a Bayesian framework for intrinsically motivation.
- Assume the agent has an internal prior on the class of possible reward functions
- Determine the policy which maximises the *intrinsic expected reward* (iER)
  - Expectation is taken with respect to the random reward function, possibly random choice of action, and the results of actions (where the latter two may form a feedback loop)
  - *Intrinsic* since the prior on potential rewards is internal to the agent
  - Agent learns about the environment and transition mechanics as in traditional RL *and*
  - Agent learns about distribution of rewards and how to improve heuristic empowerment by computing the posterior distribution of rewards and updating the iER.

Quantitative Measures for Intrinsic Motivation III: Expected Reward

- Purpose of my project: Define a Bayesian framework for intrinsically motivation.
- Assume the agent has an internal prior on the class of possible reward functions
- Determine the policy which maximises the *intrinsic expected reward* (iER)
  - Expectation is taken with respect to the random reward function, possibly random choice of action, and the results of actions (where the latter two may form a feedback loop)
  - *Intrinsic* since the prior on potential rewards is internal to the agent
  - Agent learns about the environment and transition mechanics as in traditional RL *and*
  - Agent learns about distribution of rewards and how to improve heuristic empowerment by computing the posterior distribution of rewards and updating the iER.
- Originally motivated by goal to show that the empowerment metric was Bayes-optimal the class of problems in Mohamed and Rezende [9]
  - It turns out this is false

References I

[1] Braverman, M. and Bhowmick, A. (2011). Lecture notes in information theory in computer science. https://www.cs.princeton.edu/courses/archive/fall11/cos597D/L04.pdf.

[2] Chentanez, N., Barto, A. G., and Singh, S. P. (2004). Intrinsically motivated reinforcement learning. In *Advances in neural information processing systems*, pages 1281–1288.

[3] Christiano, P., Shah, Z., Mordatch, I., Schneider, J., Blackwell, T., Tobin, J., Abbeel, P., and Zaremba, W. (2016). Transfer from simulation to real world through learning deep inverse dynamics model. *arXiv preprint arXiv:1610.03518*.

[4] Depeweg, S., Hernández-Lobato, J. M., Doshi-Velez, F., and Udluft, S. (2016). Learning and policy search in stochastic dynamical systems with bayesian neural networks. *arXiv preprint arXiv:1605.07127*.

[5] Finn, C. and Levine, S. (2016). Deep visual foresight for planning robot motion. *arXiv preprint arXiv:1610.00696*.

References II

[6] Krishnan, R. G., Shalit, U., and Sontag, D. (2016). Structured inference networks for nonlinear state space models. *arXiv preprint arXiv:1609.09869*.

[7] McAllister, R. and Rasmussen, C. E. (2016). Data-efficient reinforcement learning in continuous-state pomdps. *arXiv preprint arXiv:1602.02523*.

[8] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., et al. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533.

[9] Mohamed, S. and Rezende, D. J. (2015). Variational information maximisation for intrinsically motivated reinforcement learning. In *Advances in Neural Information Processing Systems*, pages 2125–2133.

[10] Oudeyer, P.-Y., Kaplan, F., et al. (2008). How can we define intrinsic motivation. In *Proc. 8th Int. Conf. Epigenetic Robot.: Modeling Cogn. Develop. Robot. Syst.*

[11] Salge, C., Glackin, C., and Polani, D. (2014). Empowerment–an introduction. In *Guided Self-Organization: Inception*, pages 67–114. Springer.

References III

[12] Schmidhuber, J. (2010). Formal theory of creativity, fun, and intrinsic motivation (1990–2010). *IEEE Transactions on Autonomous Mental Development*, 2(3):230–247.

[13] Singh, S., Lewis, R. L., Barto, A. G., and Sorg, J. (2010). Intrinsically motivated reinforcement learning: An evolutionary perspective. *IEEE Transactions on Autonomous Mental Development*, 2(2):70–82.

[14] Sutton, R. S. and Barto, A. G. (1998). *Reinforcement learning: An introduction*, volume 1. MIT press Cambridge.