Towards A Computational Model for Intrinsic Motivation in Imitation Learning

Manfred Diaz [1]

1. Preamble

In a common Imitation Learning (IL) setting, two decision making agents: a *learner* and an *expert* interact with the environment while the *learner* observes the *expert* trying to maximize a measurement of its efficiency (a reward signal $\mathcal{R}$) emitted by the environment. The computational models for IL proposed so far have been focused on either retrieving $\mathcal{R}$ or on the direct mapping observations-actions to learn a policy $\hat{\pi}$ from traces τ of the expert's policy π^* (Argall et al., 2009; Hussein et al., 2017). Both models have proven to be more efficient to find π^* than other learning-based methods like Reinforcement Learning (RL) (Abbeel, 2008). By using τ as a prior, IL reduces the learner's interaction with the environment and eliminates the requirement to design $\mathcal{R}$ (the behavior is demonstrated instead).

However, the construction of current IL models is not exempt from significant issues. For instance, the advantage of IL over RL comes from an extensive teacher (usually human) supervision. Furthermore, unless $\mathcal{R}$ can be entirely determined, the learner's performance is limited to the behavior described in τ, which may not be optimal (non-optimal expert). Also, the notion of *where* and *when* the teacher's interventions are no longer required is not entirely clear (there is no measure of learning progression (Oudeyer, 2018)). Another issue is that the expert's behavior is usually observed from a first-person perspective instead of a third-person perspective as it happens in nature (Stadie et al.; Liu et al.).

More fundamental questions arise from contrasting IL computational models with existing models of imitative behavior in neurosciences or psychology. For example, in psychology, the Self-Determination Theory (Deci & Ryan, 1985) distinguishes between *intrinsic motivation* (performing a task because it is interesting or enjoyable) and extrinsic motivation (doing something because it leads a required result) (Ryan & Deci, 2000). However, current IL computational models do not precisely model the *extrinsic-intrinsic duality* of an agent rewards system. Modelling this duality could, for instance, facilitate the application of IL to continual or lifelong learning settings where an IL learner would require the ability to discriminate the teacher's extrinsic and intrinsic rewards in the wild.

RL already exploits the discrimination between extrinsic $\mathcal{R}_{\mathcal{E}}$ and intrinsic rewards $\mathcal{R}_I$ (Singh et al., 2010). The incorporation of *intrinsic motivation* enables more efficient and robust policies to be discovered, or use it as a unique reward signal (Pathak et al., 2017; Pathak et al.). The success of intrinsic motivation in RL encourages the exploration of the potential impact it may have on IL.

2. Intrinsic Motivation in IL

Neurosciences and psychological theories have been a template for the development of intelligent systems (Oudeyer et al., 2016; Oudeyer, 2018; Schmidhuber, 2010). Rooted in seminal work of *observational learning* (Bandura, 2004), IL computational models are no exception. Hence, it is not hard to believe that the inclusion of intrinsic motivation into IL must be ingrained into neurosciences developments, a path also followed by *developmental robotics*. In this context, one can hypothesize about several ways intrinsic motivation can be integrated into IL and how it can immediately (and in a longer term) improve existing computational models of IL.

The literature on intrinsic motivation provides plausible theories that may serve as templates on how this integration could happen. For instance, Triesch (Triesch, 2013) studied the phenomena of Intrinsically-Motivated Imitation Learning (IMIL) with efficient encoding. In IMIL, the learner observes the teacher performance to learn a model $\mathcal{M}$ that encodes behaviours' sensory consequences. With an efficient encoding, the learner can subsequently act using as $\mathcal{R}_I$ a measure of how well the sensory consequences of its acts are encoded by $\mathcal{M}$. In computational terms, $\mathcal{M}$ is forward model of the environment dynamics $p(s_{t+1}|s_t, a_t)$ derived from τ. The actions (a_t) or states (s_t) that are more similar to those of the teacher could lead to higher reinforcement signals.

One could also hypothesize that an immediate consequence

[1] Mila, Department of Computer Science and Operations Research, University of Montréal, Montréal, Canada.. Correspondence to: Manfred Diaz <diazcabm@iro.umontreal.ca>.

of learning $\mathcal{M}$ and using the encoding error as $\mathcal{R}_I$ could be that the knowledge of $\mathcal{M}$ reduces the burden IL models impose over the teacher. The idea that part of the learning process would occur unsupervised is particularly crucial for IL models that directly estimate observation-actions mappings. These models currently require higher levels of teacher-learner interactivity (Ross, 2013; Ross & Bagnell, 2014) to reduce *distributional shifts* (exposure bias). It seems plausible then that with an efficient encoding $\mathcal{M}$, the teacher's involvement in the process should not extend beyond offering sufficient samples to approximate $p(s_{t+1}|s_t, a_t)$ for the task at hand.

Another interesting aspect of learning $\mathcal{M}$ from τ is related to efference copy mechanisms (Crapse & Sommer, 2008). If $\mathcal{M}$ is modelled as a powerful generative model (e.g., VAE (Kingma & Welling, 2013), GAN (Goodfellow et al., 2014) or others) and it is possible to estimate controllable features of the environment from τ, it could also be feasible to devise a sampling mechanism by which the learner could continuously improve its performance by using the sensory discrepancy between $\mathcal{M}$ predictions and its observations of the environment. This continual improvement could potentially help the learner to overcome an imperfect world model that could be a result of either an imperfect learning process or non-optimal demonstrations (common in, for instance, robotics tasks).

An even more interesting and relevant question is if it is possible to retrieve other agents' intrinsic reward in a framework that would closely resemble what Inverse Reinforcement Learning (IRL) (Abbeel, 2008) is for traditional RL. From the Self-Determination Theory, a complete model of an agent motivations (or rewards system) is not complete unless we are capable of discriminating $\mathcal{R}_E$ from $\mathcal{R}_I$ by observing its behaviour. We acknowledge that this separation is not a trivial task, but we believe, as mentioned before, that IL in a continual and lifelong setting in the wild requires this ability.

In summary, it is not difficult to hypothesize that the incorporation of intrinsic motivation into IL. A computational model for incorporating intrinsic motivation in IL, if grounded in neurosciences and psychology theories, could bring exciting advances to IL learning paradigm that is already giving essential results in multiple areas, including robotics.

References

Abbeel, Pieter. *Apprenticeship Learning and Reinforcement Learning with Application to Control*. PhD thesis, Stanford University, 2008.

Argall, Brenna D., Chernova, Sonia, Veloso, Manuela, and Browning, Brett. A survey of robot learning from demonstration. *Robotics and Autonomous Systems*, 57(5):469–483, may 2009. ISSN 0921-8890.

Bandura, Albert. Observational Learning. *Learning and Memory*, pp. 482–484, nov 2004.

Crapse, Trinity B and Sommer, Marc A. Corollary discharge across the animal kingdom. *Nature Reviews Neuroscience*, 9:587, aug 2008.

Deci, Edward and Ryan, Richard. *Intrinsic Motivation and Self-Determination in Human Behavior*, volume 3. 1985.

Goodfellow, Ian J, Pouget-Abadie, Jean, Mirza, Mehdi, Xu, Bing, Warde-Farley, David, Ozair, Sherjil, Courville, Aaron, and Bengio, Yoshua. Generative adversarial nets. In *Advances in neural information processing systems*, pp. 2672–2680, 2014.

Hussein, Ahmed, Gaber, Mohamed Medhat, Elyan, Eyad, and Jayne, Chrisina. Imitation Learning. *ACM Computing Surveys*, 50(2):1–35, apr 2017. ISSN 03600300. doi: 10.1145/3054912.

Kingma, Diederik P. and Welling, Max. Auto-encoding variational bayes. *International Conference on Learning Representations*, 2013.

Liu, Yuxuan, Gupta, Abhishek, Abbeel, Pieter, and Levine, Sergey. Imitation from Observation: Learning to Imitate Behaviors from Raw Video via Context Translation. Technical report.

Oudeyer, P.-Y, Gottlieb, J, and Lopes, M. Intrinsic motivation, curiosity, and learning: Theory and applications in educational technologies. *Progress in brain research*, 229:257–284, 2016. ISSN 1875-7855. doi: 10.1016/bs.pbr.2016.05.005.

Oudeyer, Pierre-Yves. Computational Theories of Curiosity-Driven Learning. 2018.

Pathak, Deepak, Mahmoudieh, Parsa, Luo, Guanghao, Agrawal, Pulkit, Chen, Dian, Shentu, Yide, Shelhamer, Evan, Malik, Jitendra, Efros, Alexei A, and Darrell, Trevor. Zero-shot Visual Imitation. Technical report.

Pathak, Deepak, Agrawal, Pulkit, Efros, Alexei A, and Darrell, Trevor. Curiosity-driven exploration by self-supervised prediction. *arXiv preprint arXiv:1705.05363*, 2017. ISSN 1938-7228. doi: 10.1109/CVPRW.2017.70.

Ross, Stéphane. *Interactive Learning for Sequential Decisions and Predictions*. Ph.d., Carnegie Mellon University, 2013.

Ross, Stéphane and Bagnell, J. Andrew. Reinforcement and Imitation Learning via Interactive No-Regret Learning. 2014.

Ryan, Richard M and Deci, Edward L. Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary Educational Psychology*, 25(1):54–67, 2000. ISSN 0361476X. doi: 10.1006/ceps.1999.1020.

Schmidhuber, Jrgen. Formal theory of creativity, fun, and intrinsic motivation (1990-2010). *IEEE Transactions on Autonomous Mental Development*, 2(3):230–247, 2010. ISSN 19430604. doi: 10.1109/TAMD.2010.2056368.

Singh, Satinder P, Lewis, Richard L, Barto, Andrew G, and Sorg, Jonathan. Intrinsically Motivated Reinforcement Learning: An Evolutionary Perspective. *IEEE Transactions on Autonomous Mental Development*, 2(2):70–82, 2010. ISSN 1943-0604. doi: 10.1109/TAMD.2010.2051031.

Stadie, Bradly C, Abbeel, Pieter, and Sutskever, Ilya. Third-Person Imitation Learning. Technical report.

Triesch, Jochen. Imitation learning based on an intrinsic motivation mechanism for efficient coding. *Frontiers in Psychology*, 4(NOV):1–8, 2013. ISSN 16641078. doi: 10.3389/fpsyg.2013.00800.